



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO**

DOCTORADO EN CIENCIAS BIOMÉDICAS

INSTITUTO DE QUÍMICA

**CARACTERIZACIÓN ESTRUCTURAL Y EVOLUTIVA DE UNA  
GALECTINA TETRADOMINIO DE *Haliotis rufescens* Y SU  
INTERACCIÓN CON AZÚCARES SULFATADOS**

TESIS

QUE PARA OPTAR POR EL GRADO DE:  
DOCTOR EN CIENCIAS

PRESENTA:

ANGÉLICA LIZETH TOIBER ESTRELLA

DIRECTOR DE TESIS  
DRA. ALEJANDRA HERNÁNDEZ SANTOYO  
INSTITUTO DE QUÍMICA

COMITÉ TUTOR  
DRA. ADELA RODRÍGUEZ ROMERO.  
INSTITUTO DE QUÍMICA  
DR. HÉCTOR RIVEROS ROSAS.  
FACULTAD DE MEDICINA

CIUDAD DE MÉXICO,

MAYO 2023



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## MIEMBROS DEL JURADO

Presidente: Dr. José Francisco Torres Quiroz

Secretario: Dra. Adela Rodríguez Romero

Vocal: Dr. José Arcadio Farías Rico

Vocal: Dr. Gerardo Alfonso Corzo Burguete

Vocal: Dra. Nuria Victoria Sánchez Puig

Este trabajo se llevó a cabo en los laboratorios de Química de Biomacromoléculas 3 y Biología Molecular del Departamento de Química de Biomacromoléculas del Instituto de Química de la UNAM.

Del proyecto de investigación se publicó el siguiente artículo:

Toiber-Estrella AL, Quintero-Martínez A, Rodríguez-Romero A, Riveros-Rosas H, Hernández-Santoyo A. Structural and evolutionary insights into the multidomain galectin from the red abalone *Haliotis rufescens* with specificity for sulfated glycans. *Fish Shellfish Immunol.* 2022, 131, 1264-1274. doi: 10.1016/j.fsi.2022.11.015.

## **AGRADECIMIENTOS**

A la Universidad Nacional Autónoma de México, por permitir enriquecerme académicamente y por abrir mi panorama profesional.

Al programa de Doctorado en Ciencias Biomédicas por el apoyo fundamental que recibí en cada etapa de mi proyecto de investigación.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico otorgado a través del programa de becas nacionales para estudios de doctorado (No. de beca 492615).

Este proyecto se realizó con apoyo económico brindado por la Dirección General de Asuntos del Personal Académico (DGAPA) mediante el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT) con el No. de Apoyo IN208318 e IN213022.

A todos mis con profesores del Instituto de Química que compartieron su vasto conocimiento en cada clase.

Un agradecimiento especial a mi tutora principal, la Dra. Alejandra Hernández Santoyo, por brindarme toda su confianza, siempre animarme a seguir adelante y todo su invaluable tiempo brindado.

Al Dr. Héctor Riveros Rosas (Facultad de Medicina), le agradezco su asesoría, sus propuestas y toda la ayuda que me ofreció para enriquecer mi proyecto.

A la Dra. Adela Rodríguez Romero, por ser un ejemplo de perseverancia, y estar en cada avance apoyando y fortaleciendo el proyecto.

Al Dr. Adrián Quintero Martínez por su apoyo en los ensayos de desulfatación y termofluorescencia.

A la Dra. Patricia Cano Sánchez del Laboratorio de Biología Molecular por su asesoría en la sobreexpresión de rHrGal.

A los técnicos académicos del instituto de Química: Dr. Homero Gómez Velasco, y M. en C. Georgina E. Espinosa Pérez del Laboratorio Nacional de Estructura de Macromoléculas (LANEM), a la Dra. Adriana Romo Pérez del Laboratorio de Espectroscopía y Polarimetría, a la M en C. Lucero Mayra Ríos Ruiz, M. en C. Lucía del Carmen Márquez Alonso y Quím. Eréndira García Ríos del Laboratorio de Cromatografía, Al Ing. Yan Alejandro García Hernández, M.I.A. María Magdalena Aguilar Araiza, y M. en Admón. Gladys Cortés Romero de la Unidad de Cómputo UCTIC, a todos ellos muchas gracias por su invaluable apoyo técnico.

A la Q.F.B. Ma Concepción José Núñez, Técnico académico del Laboratorio de Bioquímica y Biología Estructural del Instituto de Fisiología Celular, UNAM, por su apoyo técnico.

A todos mis compañeros del laboratorio, Efrén, Laura, Jesica, Benjamín, Annia, Enrique, Arely, Jesiquita, Janet, Gaby, Miguel, Joshua y a todos los demás, les agradezco cada momento de risa y cooperación.

## RESUMEN

Las galectinas son una familia evolutivamente antigua de lectinas caracterizadas por su afinidad por  $\beta$ -galactósidos y aminoácidos altamente conservados en el dominio de reconocimiento de carbohidratos (DRC). Estas lectinas están involucradas en múltiples funciones fisiológicas, incluyendo el reconocimiento de glicanos en la superficie de virus y bacterias. Esta característica respalda su papel en la respuesta inmune innata de los moluscos marinos. En el presente estudio, purificamos y caracterizamos una galectina de 64.1 kDa, del molusco *Haliotis rufescens* a la que llamamos HrGal. Esta presenta cuatro repeticiones en tándem y cada una corresponde con un dominio de reconocimiento de carbohidratos (DRC) que no son idénticos. Las variaciones que se encuentran en cada dominio pueden ser responsables del reconocimiento de diferentes glicósidos como se determinó en los ensayos de especificidad donde se observó que HrGal reconoce no sólo galactósidos sino también glucósidos. Sin embargo, sólo los azúcares sulfatados heparina y el  $\iota$ -carragenano inhibieron su actividad hemaglutinante. Fue interesante notar que HrGal reconoció también a el  $\kappa$  y  $\lambda$ - carragenanos pero estos no inhibieron la actividad aglutinante por lo que no sólo es importante la presencia de los grupos sulfato, sino también la posición de estos. La heparina parcialmente desulfatada disminuyó notablemente su reconocimiento. La identidad de HrGal se verificó por secuenciación, confirmando que se trata de una proteína de 555 aminoácidos con una masa de 63.8 kDa. Esta proteína tiene un 100% de identidad con una galectina reportada en el GenBank con número de acceso AHX26603. Pudimos identificar que HrGal es estable en presencia de urea,  $\text{Cu}^{2+}$  y agentes reductores, esto último debido a la presencia de cuatro cisteínas libres que pueden inducir agregación por interacciones intermoleculares.

El modelo estructural de HrGal obtenido con el servidor AlphaFold mostró cuatro dominios no idénticos con un plegamiento tipo sándwich  $\beta$ . En el dominio 1 se observó un asa adicional, que también está presente en la galectina 8 de humano, la cual es considerada como la responsable del reconocimiento de los azúcares sulfatados. En este dominio, además de los aminoácidos canónicos ( $\text{H}^{57}$ ,  $\text{N}^{59}$ ,  $\text{R}^{61}$ ,

N<sup>70</sup>, W<sup>77</sup>, E<sup>80</sup>, K<sup>82</sup>), los aminoácidos que participan en el reconocimiento de ι-carragenano son: Glu<sup>11</sup>, His<sup>12</sup>, Asn<sup>41</sup>, Asp<sup>48</sup> y Gln<sup>128</sup> del DRC1 y sus equivalentes en los dominios 2, 3 y 4. La diferencia más notable entre los cuatro dominios está en el asa que une las hebras β3-β4, que es es la que se considera de gran importancia para el reconocimiento del azúcar en las galectinas.

Por otro lado, se logró obtener a la galectina recombinante a la que nombramos rHrGal. Esta se purificó a partir de los cuerpos de inclusión, solubilizándolos con TBS-Urea 6M, y observamos que la presencia de Cu<sup>+2</sup>, DTT y galactosa eran necesarios para un apropiado replegamiento.

Finalmente, los estudios de filogenia sugieren que en los diferentes taxones ocurrieron duplicaciones y fusiones de dominios que dieron lugar a proteínas con más de un dominio. Así, las distintas galectinas de 2, 3, 4 y más dominios, muy probablemente se originaron de manera independiente de las galectinas multidominio que se encuentran en vertebrados. Encontramos que las secuencias de galectinas de moluscos forman un grupo monofilético tetradominio no relacionadas con las galectinas de vertebrados.

# ÍNDICE

<b>1. INTRODUCCIÓN.....</b>	<b>13</b>
<b>2. MARCO TEÓRICO .....</b>	<b>14</b>
<b>2.1. Sistema Inmune.....</b>	<b>14</b>
<b>2.2. Sistema inmune en los invertebrados marinos .....</b>	<b>15</b>
<b>2.3. Lectinas .....</b>	<b>15</b>
<b>2.4 Clasificación de las lectinas de animales .....</b>	<b>16</b>
2.4.1. Calnexinas y Calreticulinas.....	20
2.4.2. Lectinas de tipo M .....	20
2.4.3. Lectinas tipo L.....	21
2.4.4. Lectinas tipo P.....	21
2.4.5. Lectinas tipo C.....	22
2.4.6. Galectinas .....	22
2.4.7. Lectinas tipo I (Siglecs).....	24
2.4.8. Lectinas tipo R .....	25
2.4.9. Lectinas tipo caja F.....	25
2.4.10. Ficolinas .....	26
2.4.11. Lectinas tipo Quitinasa .....	26
2.4.12. Lectinas tipo F.....	27
2.4.13. Intelectinas .....	27
<b>2.5 Nuevas familias de lectinas.....</b>	<b>28</b>
<b>2.6. Aplicaciones biomédicas de las lectinas de invertebrados marinos .....</b>	<b>29</b>
<b>3. ANTECEDENTES .....</b>	<b>32</b>
<b>3.1. Lectinas en invertebrados marinos .....</b>	<b>32</b>
<b>3.2 Galectinas de moluscos marinos.....</b>	<b>32</b>
<b>3.3 Proteínas con repeticiones en tándem.....</b>	<b>34</b>
3.3.1. Clasificación de las proteínas con repeticiones en tándem.....	35
3.3.2. Lectinas con repeticiones en tándem .....	36
3.3.3. Galectinas con repeticiones en tándem .....	40



3.4. Descripción del abulón rojo, <i>Haliotis rufescens</i> .....	41
4. JUSTIFICACIÓN .....	43
5. HIPÓTESIS .....	43
6. OBJETIVOS .....	44
6.1 Objetivo general.....	44
6.2 Objetivos específicos.....	44
7. MATERIAL Y MÉTODOS .....	45
7.1 Purificación de la galectina de <i>H. rufescens</i> (HrGal) .....	45
7.2. Determinación de la pureza de la galectina .....	45
7.2.1. Electroforesis SDS-PAGE .....	45
7.2.2. Espectrometría de masas .....	46
7.3. Formalinización de los eritrocitos de conejo .....	46
7.4 Ensayos de eritroaglutinación.....	47
7.5. Determinación de la especificidad .....	48
7.5.1. Ensayos de inhibición de la eritroaglutinación .....	48
7.5.2. Fluorometría diferencial de barrido .....	50
7.5.2.1. Ensayo de especificidad.....	51
7.5.2.2. Ensayo de estabilidad .....	51
7.5.2.3. Desulfatación de la heparina y del carragenano .....	52
7.6. Determinación de la secuencia de aminoácidos.....	53
7.7. Clonación y expresión de la galectina recombinante de <i>H. rufescens</i> (rHrGal).....	53
7.8. Purificación de la rHrGal.....	55
7.9. Dicroísmo circular.....	56
7.10. Análisis del modelo estructural y estudios de acoplamiento molecular.....	57
7.11. Análisis filogenético de HrGal .....	57
8. RESULTADOS Y DISCUSIÓN.....	59
8.1. Purificación de HrGal a partir de la fuente natural .....	59
8.2. Actividad hemoaglutinante de HrGal nativa .....	65

<b>8.3. Ensayos de especificidad de HrGal .....</b>	<b>65</b>
8.3.1. Evaluación del efecto de la presencia de los grupos sulfato en la actividad de HrGal.....	71
<b>8.4. Determinación de la estructura secundaria de HrGal por Dicroísmo circular .....</b>	<b>73</b>
<b>8.5. Ensayos de estabilidad utilizando Fluorimetría diferencial de barrido.....</b>	<b>74</b>
<b>8.6. Clonación y expresión de la galectina recombinante rHrGal.....</b>	<b>76</b>
<b>8.7. Purificación de la rHrGal .....</b>	<b>79</b>
<b>8.8. Aglutinación de eritrocitos HrGal.....</b>	<b>82</b>
<b>8.9. Análisis estructural de HrGal .....</b>	<b>85</b>
<b>8.10. Acoplamiento molecular con i-carragenano.....</b>	<b>89</b>
<b>8.11. Análisis filogenético de HrGal .....</b>	<b>92</b>
<b>9. CONCLUSIONES.....</b>	<b>99</b>
<b>10. REFERENCIAS .....</b>	<b>100</b>
<b>11. ANEXOS .....</b>	<b>122</b>

## ÍNDICE DE FIGURAS

Figura 1. Clasificación de las lectinas animales propuesta por Drickamer en 1998. ....	18
Figura 2. Organización de los dominios de las lectinas de vertebrados y moluscos acuáticos. .....	24
Figura 3. Estructuras de tres nuevas familias de lectinas identificadas en invertebrados marinos. ....	29
Figura 4. Estructuras de las diferentes familias de lectinas identificadas en invertebrados marinos .....	33
Figura 6. Clasificación estructural de proteínas basadas en la longitud de sus repeticiones. .....	36
Figura 7. Representación esquemática de diferentes tipos de simetría observados en lectinas oligoméricas, lectinas con repeticiones en tándem y lectinas híbridas.....	39
Figura 8. Galectina 8 de humano. ....	41
Figura 9. Abulón rojo <i>Haliotis rufescens</i> Swainson, 1822. ....	42
Figura 10. Esquema del ensayo de desnaturalización térmica utilizando la termofluorescencia.....	51
Figura 11. Insercion del gen de HrGal (KJ183034) en el vector de expresión pET-32a.....	54
Figura 12. Purificación de HrGal. A) Electroforesis SDS-PAGE.....	61
Figura 14. Alineamiento de secuencias de galectinas de diferentes organismos contra HrGal. ....	62
Figura 15. Secuencia de aminoácidos de HrGal.....	63
Figura 16. Alineamiento de las secuencias de las cuatro repeticiones en tándem de HrGal identificadas con el servidor Meta-Repeat-Finder .....	64
Figura 17. Interacción de HrGal con diferentes azúcares.....	67

Figura 18. Lectina de <i>Mytilus californianus</i> .....	68
Figura 19. Estructura de la heparina. ....	69
Figura 20. Estructura del carragenano k, i y l.. ....	70
Figura 21. Desulfatación de la heparina.....	72
Figura 22. Espectro de dicroísmo circular en la región del UV lejano de HrGal .....	74
Figura 23. Estudios de estabilidad de HrGal mediante ensayos de termofluorescencia.....	76
Figura 25. Ensayo de restricción. ....	77
Figura 26. Cinética de expresión de rHrGal en la cepa BL21 a 37 °C con 0.5 mM de IPTG. ..	78
Figura 27. Cinética de expresión de rHrGal.....	79
Figura 29. A) Purificación de rHrGal+Trx.....	81
Figura 30. Caracterización de la galectina recombinante. ....	82
Figura 31. Ensayos de estabilización de rHrGal.....	84
Figura 32. Modelo teórico de HrGal obtenido con el servidor AlphaFold.. ....	87
Figura 33. Dominio N-terminal de la galectina 8 de humano (rojo) sobrelapada con los diferentes dominios de HrGal. ....	88
Figura 34. Acoplamiento molecular de cada dominio de HrGal con el $\iota$ -carragenano. ....	90
Figura 35. Alineamiento de secuencias de los cuatro dominios de unión a carbohidratos (DRCs) no idénticos de HrGal. ....	91
Figura 36. Análisis filogenético de galectinas bidominio y tetradominio.. ....	93
Figura 37. Escala de tiempo para la divergencia de diferentes moluscos según la calibración de consenso recuperada principalmente de ( <a href="http://www.timetree.org">http://www.timetree.org</a> ). ....	94
Figura 38. Relación filogenética de dominios individuales de galectinas.....	97

## ÍNDICE DE TABLAS

Tabla 1. Clasificación de lectinas animales propuestas por Drickamer .....	19
Tabla 2. Azúcares empleados en los ensayos de especificidad de HrGal .....	49
Tabla 3. Ensayos de especificidad de HrGal utilizando diversos azúcares .....	66
Tabla 4. Frecuencias de infrarrojo de las bandas $\nu_{\text{as}}\text{S}=\text{O}$ de la heparina completa y de la químicamente desulfatada indicando sus baricentros.....	73
Tabla 5. Ensayo de eritroaglutinación de la HrGal con distintos aditivos.....	76
Tabla 6. Cisteínas libres y formando puentes disulfuro en cada DRC de HrGal .....	88

## 1. INTRODUCCIÓN

El sistema inmune constituye la primera línea de defensa de los organismos contra patógenos, si existe una falla en este sistema se pueden desencadenar infecciones recurrentes, autoinmunidad y otros desórdenes incluyendo en algunos casos la muerte [1]. Dicho sistema consta de los sistemas innato y adaptativo, únicamente los vertebrados pueden montar ambas respuestas. Cabe señalar que ambos sistemas tienen características distintas, el sistema adaptativo puede generar memoria inmunológica por medio de los anticuerpos, mientras que el sistema innato es el mecanismo de defensa contra infecciones más antiguo y universal. Este último se encuentra en plantas y en animales invertebrados, su línea de defensa recae en las estrategias genéricas de los receptores de reconocimiento de patrones (PRRs) que son altamente conservados y pueden limitar la invasión del patógeno [2]. Al menos se han descrito seis grupos de PRRs en invertebrados, entre los que se encuentran las lectinas incluyendo a las tipo C (CTL) y galectinas [3–5].

Las lectinas son proteínas de origen no inmune que poseen al menos un sitio capaz de unir específica y reversiblemente carbohidratos [6,7]. Estas pueden mediar interacciones célula-célula y célula-matriz extracelular. Además, están involucradas en una amplia variedad de funciones biológicas que incluyen reconocimiento, comunicación intercelular, aglutinación, adhesión, diferenciación y metástasis, apoptosis, actividad antibacteriana, antiviral y antiinflamatoria entre otras [8]. En el caso de los invertebrados marinos son una de las principales moléculas que componen el sistema inmune innato y están involucradas en la defensa contra una gran diversidad de patógenos bacterianos y virales [9–11].

La importancia de estudiar las lectinas de invertebrados marinos radica en la gran biodiversidad de estos organismos, y a que las lectinas de estas fuentes tienen múltiples aplicaciones biomédicas en distintos campos, como el desarrollo de fármacos por su actividad antibacteriana, antifúngica y antiviral o posibles terapias contra enfermedades como el cáncer [3], [12-15]. Además, en estos organismos se han encontrado lectinas con características únicas que han sido poco estudiadas, tal es el caso de la lectina de *Mytilus californianus* [55] que tiene actividad, bactericida, fungicida y anti cancerígena y que en presencia de sus ligandos forma oligómeros

con diferente arreglo cuaternario, la lectina de *Aplysia californica* que está constituida por tres dominios en tándem y que reconoce fuertemente al ácido galacturónico [16], y las galectinas con cuatro dominios recientemente descubiertas [17]. Estas últimas han sido muy poco estudiadas y sólo se han encontrado en moluscos, por lo que, con el fin de profundizar en su caracterización realizamos un análisis bioquímico, estructural y evolutivo de la galectina tetradominio del molusco univalvo *Haliothis rufescens*, también conocido como abulón rojo.

## **2. MARCO TEÓRICO**

### **2.1. Sistema Inmune**

El sistema inmune es el mecanismo de defensa que se activa contra agentes patógenos y ha evolucionado en los metazoos desde que aparecieron hace 500 millones de años, y se divide en inmunidad adaptativa e inmunidad innata. La inmunidad adaptativa que surgió a partir de los vertebrados con mandíbulas articuladas (Gnatóstomos) puede reconocer e iniciar una respuesta contra diversos patógenos potencialmente letales, es altamente específica y genera memoria inmunológica para repeler una segunda infección o eliminarla de forma más rápida. Los elementos celulares de la inmunidad adaptativa son los linfocitos T y los linfocitos B, mientras que, los elementos humorales principalmente son los anticuerpos y el sistema del complemento [18].

El sistema inmune innato ha protegido a vertebrados e invertebrados por millones de años de forma muy eficiente, entre sus componentes se incluye a las barreras físicas, como el tejido epitelial y membranas mucosas, y componentes celulares como las células polimorfonucleares que contienen gránulos citoplasmáticos que pueden producir grandes cantidades de especies reactivas de oxígeno y realizan fagocitosis [19]. Adicionalmente, las células reconocen a los patógenos por medio de elementos humorales, como los péptidos antimicrobianos, el sistema complemento y los PRRs que pueden ser expresados en la superficie celular, en compartimentos intracelulares o secretados en el torrente sanguíneo y fluidos de tejidos y que reconocen patrones conservados en distintas clases de

patógenos [2]. Estos patrones son comunes entre los grupos de patógenos y están presentes en los lipopolisacáridos (LPS) de las bacterias Gram negativas, los glicolípidos de micobacterias, ácido lipoteicoico de bacterias Gram positivas, la doble cadena de los virus de ARN, etc. [20].

## **2.2. Sistema inmune en los invertebrados marinos**

Los invertebrados marinos han tenido éxito para sobrevivir a condiciones únicas en su hábitat, que pueden incluir altas concentraciones de sal, alta presión, falta de oxígeno, falta de luz y temperaturas extremas [21], además, su naturaleza inherente los expone a múltiples patógenos que en el mar pueden alcanzar densidades muy elevadas de hasta  $10^6$  bacterias/mL,  $10^9$  virus/mL y  $10^7$  hongos/mL [22]. Los invertebrados marinos carecen de inmunidad adaptativa y su sistema de defensa es primitivo; sin embargo, es muy robusto y está mediada por factores de defensa celular y humoral que trabajan de manera coordinada para eliminar patógenos [22,23]. Las células de defensa son los hemocitos que circulan libres en la hemolinfa y los tejidos como centinelas. Son los responsables de llevar a cabo fagocitosis, producir especies reactivas de oxígeno, liberar enzimas hidrolíticas y péptidos antimicrobianos, como respuesta al reconocimiento de un antígeno [24-27]. Por otra parte, el factor humoral está constituido principalmente por los PRRs que intervienen en el reconocimiento de los patrones moleculares asociados a patógenos (PAMP's). En los moluscos marinos se han descrito, las proteínas con dominio C1q, las proteínas que reconocen peptidoglicanos, las proteínas relacionadas al fibrinógeno y las lectinas (tipo C, tipo F y galectinas) [25,28].

## **2.3. Lectinas**

Las lectinas son ubicuas en la naturaleza se encuentran en todos los seres vivos desde plantas, animales, algas, bacterias y virus [30]. Desde hace más de 100 años se observó que existían proteínas que podían aglutinar eritrocitos, a las que se les nombró como hemoaglutininas o fitoaglutininas, debido a que las primeras lectinas se habían obtenido en plantas. Posteriormente se descubrió que estas proteínas podían reconocer carbohidratos. Los siguientes estudios se enfocaron en determinar que



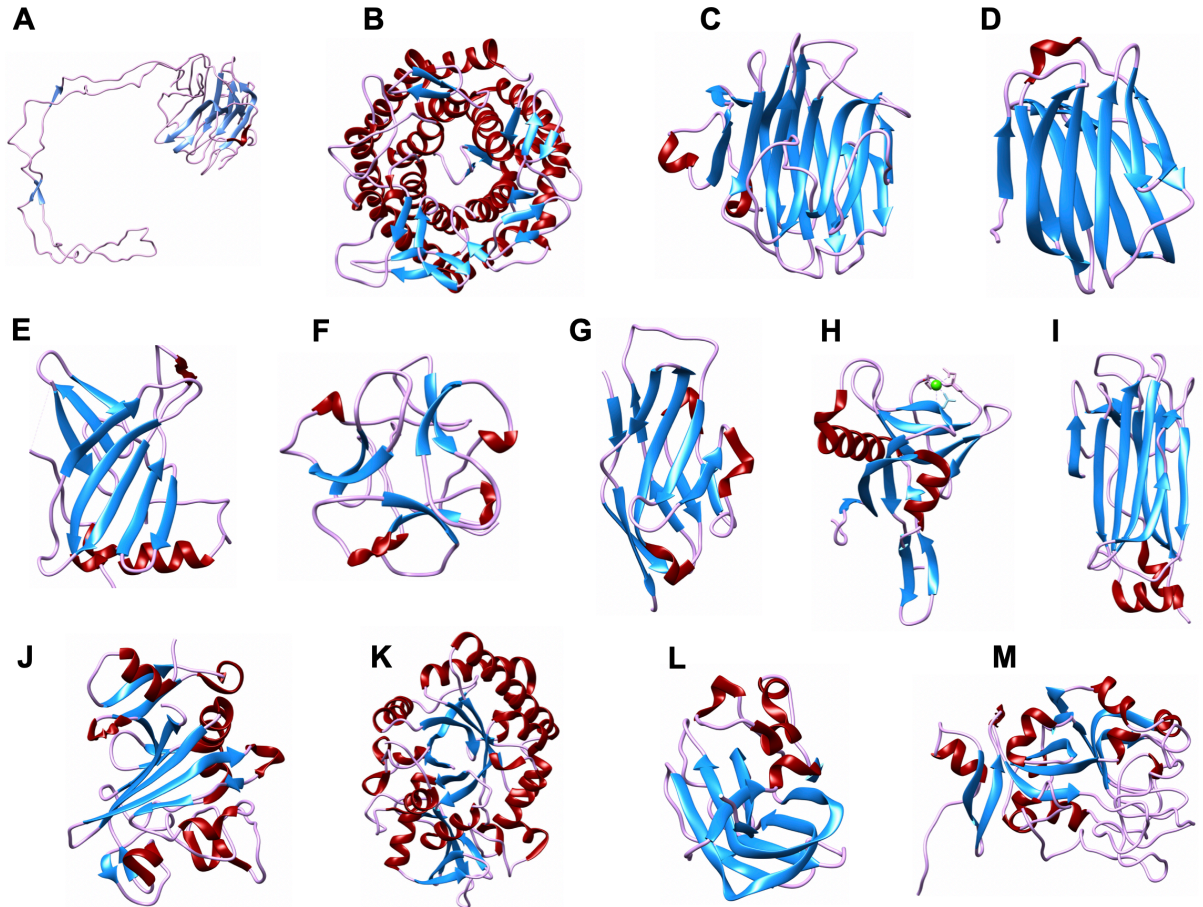
carbohidratos eran reconocidos por estas proteínas, y como resultado se distinguieron los grupos sanguíneos ABO. Este hallazgo evidenció la presencia de azúcares en las superficies celulares y su papel como marcadores de identidad. Debido a su capacidad de selección, a estas proteínas se les renombró como lectinas, del latín *legere* (elegir) [6]. La especificidad de las lectinas es muy elevada, pueden reconocer el azúcar completo, un sitio específico, una secuencia de glicósidos o sus enlaces glicosídicos, incluso una sola sustitución de un aminoácido puede alterar la especificidad por el ligando en la lectina [8]. Comparado con los anticuerpos, las lectinas son menos complementarias o afines, típicamente su  $K_a$  en las interacciones lectina-carbohidrato es de  $10^3$ - $10^7$   $M^{-1}$  [31].

#### **2.4 Clasificación de las lectinas de animales**

Kurt Drickamer, 1988 fue el primero en proponer motivos de secuencias conservadas en los dominios de unión a carbohidratos de las lectinas que requerían calcio o un entorno reductor para unir sus ligandos, clasificándolas como tipos C y S respectivamente [32]. Uno de los conceptos principales en la clasificación de lectinas son los dominios de reconocimiento a carbohidratos (DRC), en cada dominio se encuentra incorporada la actividad de unión a glicanos. Los DRCs son segmentos de proteínas que se pliegan independientemente; si la proteína tiene distintas funciones, a menudo se encuentran organizados de forma aislada en un dominio distinto para separar la actividad de unión a su ligando con las otras funciones [33].

Con la disponibilidad de un número cada vez mayor de secuencias de aminoácidos y estructuras cristalográficas de lectinas de animales, ahora es posible clasificarlas en varios grupos basados principalmente en motivos de secuencias conservadas de DRC, dependencia de iones, plegamiento y propiedades adicionales tales como el azúcar que reconocen y la dependencia de un ambiente reductor. Desde entonces, la identificación de motivos de secuencia conservada en los DRC's de numerosos grupos de lectinas ha resultado en el establecimiento de trece familias. A la fecha, la búsqueda de dichos motivos constituye la herramienta de diagnóstico más utilizada para la asignación de una familia en lectinas recién identificadas [34].

En las lectinas que se han aislado de vertebrados existen ocho grandes grupos de los cuales cuatro de ellos llevan a cabo sus funciones generalmente dentro de la célula, como son las que pertenecen a la familia de la calnexina y las tipo M, L y P, que se involucran en el correcto plegamiento de proteínas, así como su salida del retículo endoplásmico o en el reconocimiento de proteínas mal plegadas para su posterior degradación. Las familias restantes de lectinas son las tipo C, tipo R, siglecs y galectinas, que realizan sus funciones en el exterior celular, son excretadas dentro de la matriz extracelular o mediante fluidos, aunque también se pueden localizar en la membrana citoplasmática para mediar diversas funciones como la adhesión y comunicación celular, eventos de señalización, así como reconocimiento de glicanos y patógenos. Además de las familias de lectinas anteriormente mencionadas existen otras cinco que llevan a cabo funciones complementarias, y de las que aún se sigue estudiando la función biológica que desempeñan, estas son las tipo caja F, ficolinas, tipo quitinasa, tipo F e intelectinas [35] (Figura 1, Tabla 1).



**Figura 1.** Clasificación de las lectinas animales propuesta por Drickamer en 1998. A) Lectina tipo calnexina de *Canis lupus*, B) Lectina tipo M de *Homo sapiens*, C) Lectina tipo L de *Rattus norvegicus*, D) Lectina tipo Galectina de *Homo sapiens*, E) Lectina tipo P de *Bos taurus*, F) Lectina tipo R de *Marasmius oreades*, G) Lectina tipo I de *Mus musculus* H) Lectina tipo C de *Cucumaria echinata*, I) Lectina tipo caja F de *Mus musculus*, J) Lectina tipo ficolina de *Tachypleus tridentatus*, K) lectina tipo quitinasa de *Mus musculus*, L) Lectina tipo F de *Anguilla anguilla*, M) Lectina tipo Intellectina de *Homo sapiens*. Con códigos PDB 1JHN, 1X9D, 2A6V, 1KJL, 1M6P, 1DQ0, 2BVE, 3ALS, 1UMI, 1JC9, 1E9L, 1K12 y 4WMQ respectivamente.

**Tabla 1.** Clasificación de lectinas animales propuestas por Drickamer [31].

<b>Familia</b>	<b>Ligandos que reconocen</b>	<b>Función biológica</b>
Calnexina	Glc <sub>1</sub> Man <sub>9</sub>	Plegamiento de proteínas en RE
Tipo M	Man <sub>8</sub>	Reconocimiento de glicoproteínas desplegadas
Tipo L	Varios	Plegamiento de proteínas en RE
Tipo P	Man <sub>6</sub> -fosfato, otros	Tráfico de glicoproteínas
Tipo C	Varios	Inmunidad innata/Adhesión celular
Galectinas	β-Galactosidos	Señalización
Tipo I (siglecs)	Ácido siálico	Adhesión celular
Tipo R	Varios	Marcado de enzimas
Tipo caja F	NacGlc <sub>2</sub>	Reconocimiento de glicoproteínas desplegadas
Ficolinas	NacGlc, NacGal	Inmunidad innata
Tipo quitinasa	Quito oligosacáridos	Metabolismo del colágeno
Tipo F	Oligosacáridos con terminación en fucosa	Inmunidad innata
Intelectinas	Galactosa, galactofuranosa, pentosas	Inmunidad innata

#### 2.4.1. Calnexinas y Calreticulinas

La calnexina y la calreticulina son proteínas chaperonas homólogas que se encuentran presentes en todos los eucariotas. Son componentes del sistema de control de calidad de proteínas N-glicosiladas con un alto contenido de manosas ( $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2$ ) y un glicano precursor con tres residuos de glucosa (triglicosilada). La proteína N-glicosilada perderá dos residuos de glucosa y se genera la forma monoglicosilada que se puede unir a las chaperonas. Trabajan en conjunto con la glicosiltransferasa para promover el plegamiento correcto de las proteínas que entran en la vía secretora y se dirige a las proteínas mal plegadas para su degradación. La calreticulina es un componente luminal soluble del retículo endoplasmático y la calnexina está unida a la membrana. La estructura del dominio luminal de la calnexina consiste de un dominio membranal globular con un plegamiento de tipo sándwich  $\beta$  (sitio de reconocimiento a carbohidratos en ambas proteínas) similar al de las lectinas tipo L, además de un brazo que se proyecta fuera de la membrana que está formado por un giro conformado por segmentos polipeptídicos que se repiten [36].

#### 2.4.2. Lectinas de tipo M

El plegamiento que tienen estas lectinas es de glicosil hidrolasas y tienen homología con el receptor de manosa-6-fosfato. Al igual que las calnexinas su función está ligada con el tráfico y, control de calidad del plegamiento de proteínas N-glicosiladas. Aquellas glicoproteínas que no lograron plegarse adecuadamente se retienen en el retículo endoplásmico y se eliminan con el mecanismo de degradación asociada a este. Lo que resulta en la retro translocación al citoplasma para que estas proteínas sean degradadas por el proteosoma citoplasmático. El dominio de reconocimiento a carbohidratos de las lectinas tipo M es una estructura de tipo barril conformada por hélices alfa y hebras beta, lo que permite la interacción con su ligando a pesar de que tenga gran tamaño [37,38].

#### 2.4.3. Lectinas tipo L

Se les nombro tipo L debido a que las descubrieron en las semillas de leguminosas, posteriormente se encontraron también en hongos y animales. Las lectinas tipo L a pesar de encontrarse tanto en plantas como en animales, difieren en su secuencia y funciones biológicas. En animales se encuentran comúnmente como proteínas ancladas a la membrana, participando en la clasificación y tráfico de proteínas en las células. En humanos y otros mamíferos hay cuatro lectinas de tipo L: ERGIC-53, ERGL, VIP36 y VIPL. ERGL se encuentra solo en mamíferos y VIP36 está restringido a vertebrados, pero ERGIC-53 y VIPL también se encuentran en invertebrados [39].

Las lectinas de tipo L se distinguen de otras lectinas por su estructura terciaria y no por la secuencia primaria. Este tipo de proteínas poseen un plegamiento de tipo sándwich  $\beta$  formado de tres láminas  $\beta$  antiparalelas, una lámina posterior plana de seis hebras, una lámina frontal cóncava de siete hebras y una lámina superior corta que mantiene conectadas las dos hojas principales. Todas estas lectinas requieren  $\text{Ca}^{2+}$  y un ión de metal de transición (generalmente  $\text{Mn}^{2+}$ ) para su actividad de unión a carbohidratos. El sitio de unión a carbohidratos se compone de cuatro aminoácidos conservados que se encuentran distribuidos en las asas A, B, C y D [39].

#### 2.4.4. Lectinas tipo P

El dominio de reconocimiento a carbohidratos de tipo P fue identificado por primera vez en la enfermedad de las células I (mucopolipidosis-II) caracterizada por una falla en la degradación de los componentes celulares acumulados en los lisosomas. Las lectinas tipo-P están involucradas en el tráfico especializado de las enzimas lisosomales, su mecanismo se desencadena al reconocer los residuos de manosa 6 fosfato (M6P) que se hayan agregado a los N-glicanos presentes en las enzimas que se sintetizan en el retículo endoplásmico, y posteriormente estas enzimas llegan al aparato de Golgi, donde se fusionan con los lisosomas. Estructuralmente estas lectinas se pliegan en un barril  $\beta$  aplanado de nueve hebras [39,40].

#### 2.4.5. Lectinas tipo C

Es la familia más grande y diversa de proteínas de unión a glicanos y son dependientes de  $\text{Ca}^{2+}$ . Sin embargo, se ha propuesto la designación general de dominio similar a lectina de tipo C (CTLD), porque no todas las proteínas con este dominio se unen a glicanos o a  $\text{Ca}^{2+}$ . Con este término quedan incluidos todas aquellas proteínas que se relacionan estructuralmente, por lo que actualmente los CTL se han clasificado en 16 grupos que incluyen a colectinas, selectinas, receptores endocíticos y proteoglicanos entre otros [42]. Estos difieren significativamente en los tipos de glicanos que reconocen con alta afinidad. Estas proteínas funcionan como receptores de adhesión y señalización en muchas vías, incluidas la homeostasis y la inmunidad innata, y son cruciales en las respuestas inflamatorias y el tráfico de leucocitos y plaquetas [42].

El DRC es una región compacta de 110-130 aminoácidos con una hoja  $\beta$  antiparalela de dos hebras formada por los residuos amino y carboxilo terminal unidos por dos hélices  $\alpha$  y una hoja  $\beta$  antiparalela de tres hebras. El DRC tiene dos enlaces disulfuro conservados y hasta cuatro sitios para unirse a  $\text{Ca}^{2+}$ . La unión de sus respectivos ligandos es mediada por el ion  $\text{Ca}^{2+}$ , el cual estabiliza la conformación local de la proteína formando enlaces de coordinación de la superficie de la proteína a los hidroxilos de los anillos de los azúcares, también los residuos de la lectina que son capaces de formar un enlace de coordinación con el  $\text{Ca}^{2+}$  pueden formar puentes de hidrógeno con los hidroxilos del carbohidrato. Los residuos clave conservados que se unen a azúcares incluyen el motivo "EPN" (que promueve la unión a Man, GlcNAc, Fuc y Glc) y el motivo "WND" (que promueve la unión a Gal y GalNAc), como se observa en la L-selectina y en la proteína C de unión a manosa en ratón y rata, respectivamente [42].

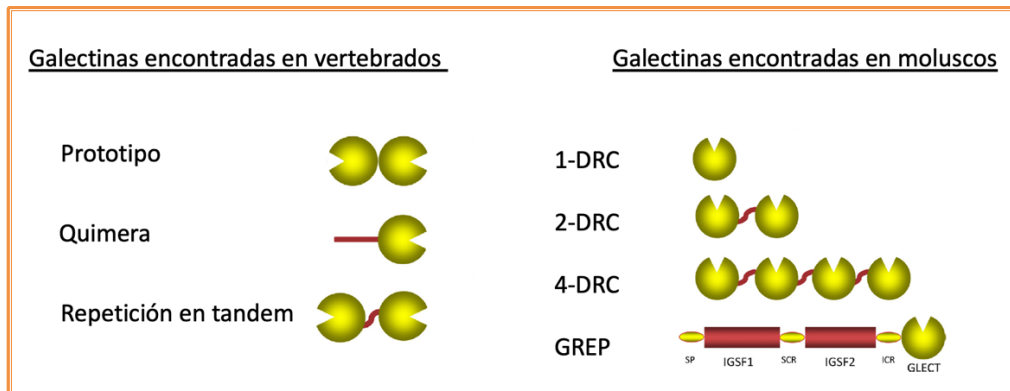
#### 2.4.6. Galectinas

Son probablemente la clase más antigua de proteínas de unión a glicanos, se encuentran en todos los organismos, desde esponjas y hongos, hasta invertebrados y vertebrados. Las galectinas son lectinas que se caracterizan por reconocer principalmente  $\beta$ -galactósidos mediante la interacción con 7 aminoácidos canónicos

(H,N,R,N,W,E,R). El DRC canónico de las galectinas tiene aproximadamente 130 aminoácidos que forman dos láminas  $\beta$  compuestas por 5 o 6 hebras antiparalelas, dichas láminas están dispuestas en una configuración de sándwich  $\beta$ . A la fecha se han descrito un total de 15 galectinas en mamíferos que se han clasificado en tres grupos principales: las galectinas prototipo, quimera y con repeticiones en tándem [43] (Figura 2). Las galectinas prototipo son aquellas que contienen un único DRC que puede asociarse de forma no covalente para formar homodímeros. Las galectinas de este tipo descritas a la fecha son la Gal-1, 2, 5, 7, 10, 11, 13, 14 y 15. Las galectinas quiméricas se caracterizan por tener un único DRC y un gran dominio amino terminal, que es rico en residuos de prolina, glicina y tirosina, que puede contribuir a su oligomerización y que es sensible a metaloproteinasas. Las galectinas quiméricas son más comunes en invertebrados. En el caso de los vertebrados, la única que se ha descrito es la galectina-3. Por último, se encuentran las galectinas con repeticiones en tándem (Gal-4, 6, 8, 9, 12), que se componen de dos CRD que provienen de una sola cadena polipeptídica. Estos dos DRC están unidos por un asa que puede tener una longitud de 5 a más de 50 aminoácidos [44].

Con respecto a su actividad fisiológica, las galectinas pueden contribuir con una amplia gama de funciones, como son el desarrollo, regeneración de tejidos, cáncer, así como sus roles en la regulación inmune y la inflamación. Cada tipo de galectina participa en funciones distintas que varían en diferentes especies, por ejemplo, las galectinas 1 y 3 se encuentran involucradas en el desarrollo muscular y del sistema nervioso central, por otra parte, galectinas como la 1, 2, 3, 4, 7, 9 y 12 se relacionan con la inducción de la apoptosis. También se relacionan con las distintas etapas del cáncer, ya sea para el desarrollo de nuevos vasos sanguíneos o para la progresión del tumor a metástasis. Adicionalmente, las galectinas promueven respuestas proinflamatorias y antiinflamatorias, ya que se encuentran expresadas por distintas células inmunes. Además, al reconocer los patrones moleculares asociados a patógenos (PAMPS) de virus, bacterias y parásitos, permiten que se inhiba la adhesión del patógeno o promueve la fagocitosis [44]. Resumiendo, las galectinas participan en un gran número de procesos biológicos y su función puede variar dependiendo de la naturaleza del ligando disponible [45].





**Figura 2.** Organización de los dominios de las lectinas de vertebrados y moluscos acuáticos. A la izquierda se muestra una representación esquemática de la organización de dominios de los tres tipos de galectina (prototipo, quimera y repetición en tándem) descritos en vertebrados. A la derecha se observa la representación esquemática de la organización de dominios descritos en moluscos acuáticos. La GREP es una proteína quimérica en la que un dominio C-terminal de galectina (GLECT) se une a través de una región intermedia corta (ICR) a dos dominios de la superfamilia de inmunoglobulinas (IgSF1 e IgSF2) separados por una pequeña región de conexión (SCR), y con el péptido señal (SP). Tomado de Vasta et al. 2015 [46].

#### 2.4.7. Lectinas tipo I (Siglecs)

Las lectinas tipo I se unen carbohidratos a través de DRCs que son homólogos a la superfamilia de las Inmunoglobulina (IgSF). Entre las lectinas tipo I, la familia Siglec son el subgrupo que reconoce al ácido siálico y es el que mejor se ha caracterizado funcional y estructuralmente. Entre sus principales funciones, destaca que participan en el sistema inmune y el mantenimiento del sistema nervioso. Estructuralmente, los Siglecs poseen el plegamiento de Ig que está formado por hebras  $\beta$  antiparalelas organizadas en un sándwich  $\beta$  que contiene 100-120 aminoácidos y, generalmente se encuentra estabilizado por un enlace disulfuro entre las láminas. De forma conservada se encuentra una arginina en la hebra  $\beta$ , que se ha descrito por ser esencial para establecer las interacciones electrostáticas con su ligando, que es el ácido siálico. Se ha reportado un número muy pequeño de proteínas de tipo Siglec

que carecen de esta arginina en el sitio de reconocimiento, tal es el caso de la Siglec 12 de humano, en este caso la unión está dirigida por dos residuos aromáticos que establecen interacciones de apilamiento de anillos [45,46].

#### 2.4.8. Lectinas tipo R

Fue la primera lectina en ser descubierta y se encontraba en los extractos de las semillas de ricino, de ahí su nombre. Es una toxina cuyo DRC es el único dominio de lectinas animales que también se encuentra en procariontes, este DRC une residuos de galactosa de las glicoproteínas que se encuentran en la superficie celular y se importa a los endosomas, luego migra al retículo endoplásmico, a través del aparato de Golgi, iniciando el transporte de la toxina a través de la vía secretora para finalmente llegar al citoplasma y matar células. Se sintetiza como una sola cadena polipeptídica que contiene a las cadenas A y B. Ambas cadenas unidas con un puente disulfuro. El dominio de tipo R tiene una estructura de trébol  $\beta$  conformado por tres lóbulos ( $\alpha$ ,  $\beta$  y  $\gamma$ ). El sitio de unión a ligando, mantiene un motivo conservado QxW, que se conforma por un aminoácido aromático que se apila contra el anillo de Gal/GalNac, así como por puentes de hidrógeno formados por los residuos del sitio y los grupos hidroxilos del carbohidrato [47,48].

#### 2.4.9. Lectinas tipo caja F

Estas proteínas juegan un papel muy importante en la vía proteolítica de la ubiquitinación de proteínas desplegadas, y su posterior reconocimiento por el proteosoma 26S para su degradación. La ubiquitinación regula el recambio, la ubicación y la actividad de las proteínas. El DRC en estas lectinas está conformado por un dominio tipo sándwich constituido de cinco hebras  $\beta$  antiparalelas, al final de este dominio se encuentran dos hélices  $\alpha$  que corresponden a la región N- y C-terminal. En el dominio C-terminal se lleva a cabo el reconocimiento selectivo de las proteínas diana. El sitio de reconocimiento a carbohidratos es específico para el disacárido quitobiosa formado por dos unidades de N-acetil glucosamina, la interacción con el ligando se da mediante apilamiento hidrofóbico con una de las unidades del disacárido, mientras que el grupo metilo del segundo interacciona con

una cavidad hidrofóbica del sitio de la lectina. Se ha señalado que en ocasiones las proteínas de caja F pueden desencadenar la ubiquitinación a través de interacciones proteína-carbohidrato [51,52]

#### 2.4.10. Ficolinas

Las ficolinas son un grupo de lectinas de naturaleza soluble que participan en eventos de la inmunidad innata, reconociendo los carbohidratos de la superficie de los patógenos, lo que desencadena la activación de proteasas. La ubicación de estas proteínas dentro de la clasificación de las lectinas es controversial, ya que su dominio de unión a ligandos tiene la capacidad de reconocer grupos acetilo en carbohidratos (N-acetilglucosamina, N-acetilmanosa, N-acetilgalactosamina) y en otros ligandos sin carbohidratos (N-acetilglicina, N-acetilcisteína, acetilcolina). La unión a los azúcares no es dependiente de los anillos del carbohidrato, ya que los azúcares que no se encuentran acetilados no son reconocidos por las ficolinas. La cadena polipeptídica de una ficolina consta de una pequeña región N-terminal, un dominio de tipo colágeno, una región que actúa como un cuello y una región de tipo fibrinógeno. La región de tipo colágeno es responsable de la asociación de los monómeros de ficolina en trímeros, a su vez el N-terminal contiene residuos de cisteína que permite la asociación covalente a oligómeros superiores con una apariencia de ramillete, lo que ha destacado por que llega a ser similar al componente del complemento C1q [51,52].

#### 2.4.11. Lectinas tipo Quitinasa

Las lectinas tipo quitinasa son proteínas de mamíferos que pertenecen a la familia 18 de las glicosil hidrolasas. Se pueden unir a la quitina, la heparina y el ácido hialurónico, y se regulan por cambios en la matriz extracelular. Se secretan en diferentes tipos celulares del sistema inmune, del músculo liso y células tumorales. Se encuentra involucrada en la protección contra patógenos, inflamación, crecimiento, remodelación de tejidos y cicatrización. Tienen una estructura de tipo barril  $\alpha+\beta$  como el presente en la triosa fosfato isomerasa. El ligando que reconocen

este tipo de lectinas son oligosacáridos formados por unidades de N-acetil glucosamina [53,54].

#### 2.4.12. Lectinas tipo F

Las lectinas tipo F se han estudiado principalmente en peces y bacterias, y tienen un papel importante en el sistema inmune en los animales y en el caso de las bacterias se relaciona con actividades enzimáticas. El dominio tipo F se compone de 140 residuos con un motivo de unión a ligando conservado hX(26)RXDX(4)R/K (X, se refiere a cualquier residuo), y un motivo de unión a calcio, h2DGX (donde h se refiere a cualquier residuo hidrofóbico y X se refiere a un residuo hidrofílico pequeño). Además, puede llegar a tener un dominio extra que varía según la especie, puede ser de pentraxina o Tipo C. Los dominios de reconocimiento de este tipo de lectinas tienen un alto contenido de hebras  $\beta$  como es el caso de la lectina de *Anguilla anguilla*, la cual es una fucolectina que se ha empleado principalmente para la tipificación de sangre, ya que reconoce de forma específica los residuos terminales de fucosa en grupos sanguíneos. Sin embargo, existen algunas lectinas atípicas que pueden reconocer derivados de galactosa, 3-O-metil-galactosa, 3-O-metil-D-fucosa y también N-acetil glucosamina y N-acetil galactosamina. La función fisiología de esta lectina es reconocer los patógenos séricos, por lo que participa en el sistema inmune innato. El DRC de esta lectina posee una estructura de tipo barril  $\beta$ , un lado del barril contiene cinco asas que forman una especie de anillo que encierra el sitio de unión a fucosa y está cargado positivamente (H, R, R), un ion  $\text{Ca}^{2+}$ , que se une en este subdominio de asas para estabilizar esta región tan flexible y permitir la formación de oligómeros estabilizando la estructura cuaternaria [55,56].

#### 2.4.13. Intelectinas

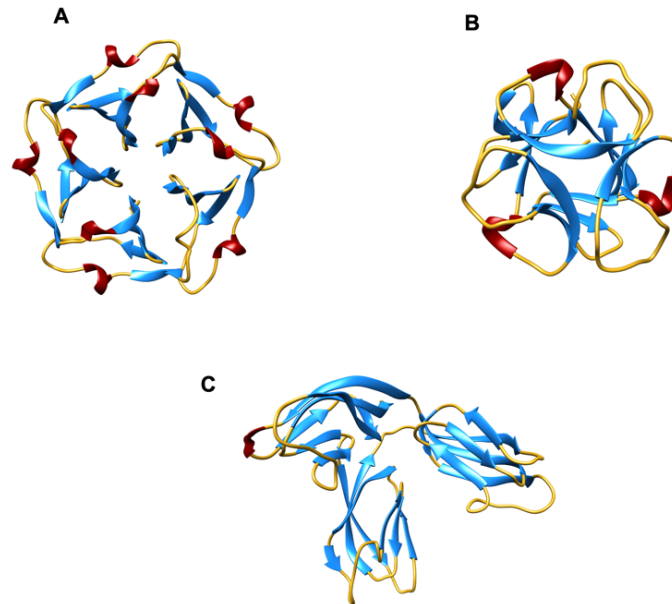
Las intelectinas fueron descritas por primera vez en la rana *Xenopus laevis* y posteriormente se han identificado también en cordados. Este tipo de lectinas posee por lo regular un dominio simple que consta de un solo DRC, con una extensión en el N-terminal no conservado de longitud variable que contiene un péptido señal para ser secretada fuera de la célula. El DRC de intelectina no está relacionado en secuencia

con alguna otra lectina con excepción de 45 residuos del N-terminal, los cuales presentan similitud con el dominio globular de fibrinógeno. Los miembros del grupo intelectina pueden identificarse por un motivo de secuencia único y altamente conservado (GGWTLVASVHEN o similar), que abarca las regiones nuevas y similares al dominio fibrinógeno. En estos dominios de reconocimiento a carbohidrato se ha reportado que presentan dependencia por  $\text{Ca}^{2+}$  y reconocimiento por diferentes monosacáridos como galactosa y algunas otras pentosas. Son muy pocas las Intelectinas que se han descrito en animales y están relacionadas con distintas funciones en la inmunidad innata, fertilización y embriogénesis [59].

### **2.5 Nuevas familias de lectinas**

Es interesante notar que en invertebrados marinos se han identificado lectinas que no se han logrado clasificar en las familias propuestas por Drickamer, debido a la falta de similitud en secuencia de los DRC's o en su estructura y reconocimiento de ligandos. Es por esto se ha propuesto incluir nuevas familias, como las lectinas de *Tachypleus tridentatus* (taquilectinas) y las lectinas descritas en la familia *Mytilidae*, las cuales generalmente se relacionan con actividades de defensa. Las taquilectinas presentan similitudes estructurales con la taquilectina 1 de *T. tridentatus* que se caracteriza por tener seis secuencias repetidas en tándem de 31 aminoácidos de longitud, formando una estructura de propela beta [60]. En la familia *Mytilidae*, se encuentran las lectinas de *Crenomytilus grayanus*, *Mytilus galloprovincialis*, *Mytilus trossulus* y *Mytilus californianus* que comparten características estructurales y una alta similitud en su secuencia [61]. Estas lectinas tiene un plegamiento tipo trébol  $\beta$  similar a la familia tipo R; sin embargo, no presentan un motivo de secuencia conservado y tampoco los puentes disulfuro altamente conservados que estabilizan la estructura terciaria de las lectinas tipo R [62,63]. Además, se han identificado lectinas tipo fibrinógeno en moluscos que se caracterizan por poseer un péptido señal, un dominio tipo fibrinógeno y en la región amino terminal contienen un dominio similar a las ficolinas, que puede favorecer su oligomerización [64]. Una familia más reciente es la de las lectinas que reconocen ácido galacturónico y que presentan 3

dominios en tándem que se ha encontrado en moluscos del género *Aplysia* (Figura 3).



**Figura 3.** Estructuras de tres nuevas familias de lectinas identificadas en invertebrados marinos. A) Tipo taquilectina (PDB: 1TL2), B) Tipo *Mytilidae* de *Mytilus californianus* (6BFM), C) Lectina de *Aplysia californica* (GalUL)

## **2.6. Aplicaciones biomédicas de las lectinas de invertebrados marinos**

El agua cubre el 71% de la superficie de nuestro planeta y alberga una gran biodiversidad de especies [65]. Cabe destacar algunas características de los filos de invertebrados marinos, por ejemplo, el filo Porífera, son los organismos más primitivos entre los animales multicelulares y sintetizan una gran variedad de moléculas para su defensa y adaptación al medio [66,67]. El filo Artrópoda tiene una gran importancia en la industria pesquera, las lectinas estudiadas en estos organismos pueden reconocer y unirse a una gran variedad de microorganismos al reconocer diferentes PAMPS [68,69]. El filo Mollusca está en el segundo lugar de diversidad con cerca de 85,000 especies descritas, ha ganado mayor importancia en la ciencia por el gran potencial para el desarrollo de fármacos encontrado en sus moléculas [70,71]. Debido a que las actividades de las lectinas de organismos

marinos son diversas, se han relacionado con el posible desarrollo de fármacos en aplicaciones biomédicas y biotecnológicas, por su actividad antibacteriana, antifúngica, antiparasitaria, antiviral, anticancerígena y de regeneración de tejidos [10,72–75].

De forma notable, las lectinas de los invertebrados marinos pueden actuar contra un amplio espectro de bacterias tanto Gram positivas como Gram negativas [76–81] y pueden facilitar el diagnóstico o nuevas estrategias terapéuticas contra bacterias de mayor importancia médica. Una ventaja del uso de estas lectinas respecto a la mayoría de los antibióticos es su mecanismo de unión, ya que las lectinas interfieren con la síntesis de la pared celular o la alteración en la permeabilidad de la membrana lipídica impidiendo la resistencia a antibióticos polipeptídicos [82]. Actualmente, han surgido estrategias novedosas en el desarrollo de nuevos antibióticos para evitar la resistencia de los microorganismos hacia el medicamento, como el uso de nanomateriales, en esta búsqueda se han combinado nanopartículas con lectinas y se han obtenido resultados exitosos contra bacterias y hongos [83]. También se ha observado que varias lectinas presentan actividad antifúngica, la cual es de amplio espectro contra diversos patógenos como *Pichia pastoris* [84], *Aspergillus*, *Alternaria* y en menor medida a *Trichoderma* [82,85]. Adicionalmente a sus propiedades de defensa contra patógenos, también se encuentra su potencial actividad antiparasitaria contra enfermedades de gran relevancia médica [62].

Por otro lado, algunas lectinas de invertebrados marinos tienen actividad antiviral, como PcLec6 de *Procambarus clarkii* que mostró actividad contra el virus de la mancha blanca que afecta a crustáceos [12,80]. Además, se estudió el efecto antiviral de distintas lectinas con especificidad a GlcNac o manosa, incluyendo a la de *Crenomytilus grayanus* (CGL), *Didemnum ternatanum* (DTL, DTL-A) y *Serpula vermicularis* (SVL-1 y SVL-2 ) y como resultado se planteó una posibilidad de control para el Virus de Inmunodeficiencia Humana, previniendo la interacción del virus con su receptor celular [87].

Cabe destacar que además de las aplicaciones de control microbiológico también se ha estudiado a las lectinas como estimulantes de la mitosis [88] y

adicionalmente para el diagnóstico de enfermedades [89–91]. Por su actividad antitumoral, recientemente se estudió una lectina que reconoce distintas líneas celulares de cáncer y que puede ser utilizada para identificar células transformadas de diferentes tejidos. Esta lectina reconoce a monosacáridos y oligosacáridos de las células transformadas GalNAc y GlcNAc en la superficie de estas células [92]. Adicionalmente ya sea con fines de diagnóstico o terapéuticos, las lectinas se destacan por reconocer patrones alterados en las glicosilaciones de células cancerígenas [93]. Algunas lectinas desencadenan señales intracelulares que provocaran la muerte celular, como la iNol de *Ibacus novemdentatus* y lectinas de *Halichondria crater*, *H. caerulea*, *Cliona varians* y *Cinachyrella apion* que han desplegado actividad citotóxica en distintas líneas cancerígenas [30,94]. Además, se han reportado lectinas de moluscos como la de *C. grayanus* y la de *Mytilis californianus* (MCL) con actividad anticancerígena [94]. La CGL presenta actividad citotóxica en células de cáncer mamario MCF7 y linfoma de Burkitt al reconocer globotriosa [95]. La MCL tiene actividad contra diferentes líneas celulares cancerígenas [96]. En nuevos estudios se utilizan vectores virales y se insertan genes terapéuticos para terapias oncológicas [61]. Se han utilizado como gen terapéutico el HddSBL de *H. discus discus* y el de la taquilectina 2 de *T. tridentatus* (TTL) y ambos mostraron actividad anticancerígena en modelos murinos [97].



### 3. ANTECEDENTES

#### 3.1. *Lectinas en invertebrados marinos*

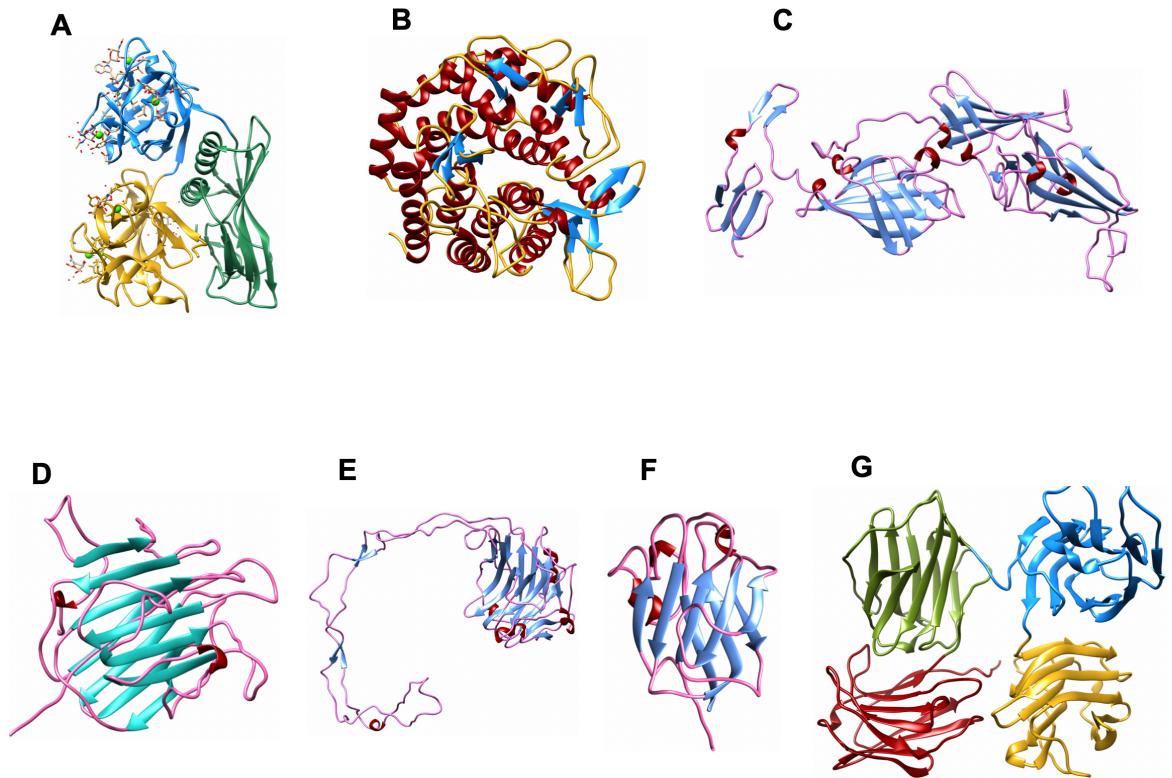
De acuerdo con la clasificación de Drickamer, las lectinas que se han identificado en invertebrados marinos pertenecen a las familias tipo C, L, P, M, F, calnexina/calreticulina y galectinas [97]. Las lectinas tipo L tienen su DRC con homología a lectinas de leguminosas, las tipo P incluyen a lectinas con homología al dominio del receptor de manosa-6-fosfato, las tipo-M son proteínas transmembranales tipo II que contienen el dominio estructural de la familia de las glicosil hidrolasas [66]. Las lectinas tipo F pertenecen a una familia identificada recientemente que se caracterizan por presentar un DRC corto de 140 residuos que contienen un motivo conservado de unión a fucosa y uno de unión a calcio. Pueden reconocer L-fucosa y oligosacáridos fucosilados y se han encontrado en moluscos, crustáceos y esponjas [98] (Figura 4).

En los invertebrados marinos, una de las principales funciones de las lectinas en el sistema inmune es distinguir a las moléculas propias de las invasoras [93,94], y se encuentran involucradas en distintos mecanismos como aglutinación [101], activación de la profenoloxidasa [102], fagocitosis [103], adhesión celular [104], factores citotóxicos [99,100], actividad antimicrobiana [107], encapsulación [87] o la eliminación de los patógenos [108].

#### 3.2 *Galectinas de moluscos marinos*

Las galectinas de tipo S, se caracterizan por reconocer  $\beta$ -galactósidos. Forman parte muy importante del sistema de defensa contra bacterias, por ejemplo, la aglutinina de *Macoma birmanica* (MBA) que reconoce bacterias Gram-negativas (*E. coli* y *V. parahaemolyticus*) y Gram-positivas (*B. subtilis* y *S. aureus*) [46] y las lectinas de *Crassostrea gigas* (CgGal2 y CgGal3) que promueven la encapsulación de patógenos [109].

Las galectinas de moluscos marinos pueden tener algunas diferencias estructurales en comparación a las que se han mencionado en mamíferos. Por ejemplo, las galectinas de los moluscos se han clasificado en cuatro grupos basados



**Figura 4.** Estructuras de las diferentes familias de lectinas identificadas en invertebrados marinos. A) Lectina tipo C de *C. echinata* (PDB 2Z48), B) Lectina tipo M de *M. japonicus* (modelo), C) Lectina tipo P de *M. japonicus* (modelo), D) Lectina tipo L de *P. trituberculatus* (modelo), E) Calnexina de *P. monodon* (modelo), F) Lectina tipo F de *P. martensi* (modelo) y G) Galectina de *H. rufescens* (modelo).

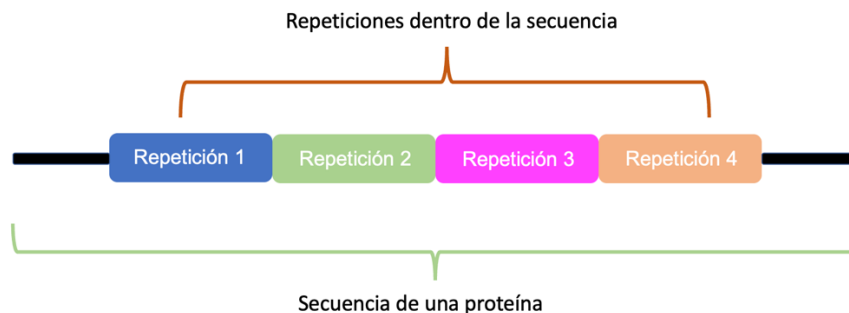
en el número de DRC: 1-DRC, 2-DRC, 4-DRC y la proteína relacionada a la galectina quimérica [9,77,110]. Recientemente, se han estudiado galectinas con cuatro dominios involucrados en la respuesta inmune innata con actividad antibacteriana debido a un aumento de ARNm después de retar contra *Vibrio anguillarum*, *Micrococcus luteus* y *V. Alginolyticus* [111,112]. A la fecha se han logrado expresar solo cuatro galectinas que estructuralmente cuentan con cuatro dominios en tándem, una de ellas es CvGal de *Crassostrea virginica* con actividad antiparasitaria contra *Perkinsus marinus* [113]. Así como, CgGal-3 de *C. gigas* que cuenta con un amplio espectro antimicrobiano contra *Escherichia coli*, *V. vulnificus*, *M. luteus* y actividad antifúngica contra *Saccharomyces cerevisiae* y *P. pastoris* [114]. La tercera es

AiGal2 de *Argopecten irradians* con actividad antibacteriana contra *E. coli*, *V. anguillarum*, *V. fluvialis*, *Edwardsiella tarda* y *M. luteus* [115]; y por último, ChGal de *C. hongkongensis* que presentó actividad antibacteriana contra *E. coli*, *V. alginolyticus* y *Bacillus thuringiensis* [116].

Adicionalmente, se ha descrito como mecanismo de defensa innata a la HrGal de *Haliotis rufescens*, ya que se obtuvo un incremento de ARNm después de retar con *V. anguillarum* [17]. Hasta la fecha la información respecto a las galectinas de cuatro dominios se ha limitado a definir si tienen actividad antimicrobiana. Cabe destacar que hacen falta más estudios con este tipo de lectinas para complementar la información y saber qué tipo de ligando pueden reconocer y como se está llevando a cabo el reconocimiento, profundizar en el origen de las galectinas de moluscos y conocer qué tipo de relación tienen con las galectinas de vertebrados.

### 3.3 Proteínas con repeticiones en tándem

Las proteínas pueden tener secuencias periódicas que representan conjuntos de repeticiones directamente adyacentes entre sí (Figura 5). Estas repeticiones en tándem son considerablemente diversas y pueden incluir dominios de 100 o más residuos y están presentes en al menos el 14% de todas las proteínas, lo que define su importancia en la duplicación para desarrollar nuevas funciones proteicas [117]. Este tipo de arreglos con repeticiones permiten obtener formas no globulares que pueden favorecer las interacciones proteína-proteína al permitir la formación de interfaces largas y diversas así como interacciones multivalentes cooperativas [118].



**Figura 5.** Representación esquemática de la secuencia de una proteína con repeticiones en tándem. Las repeticiones pueden ser continuas o estar separadas por asas que varían en longitud.

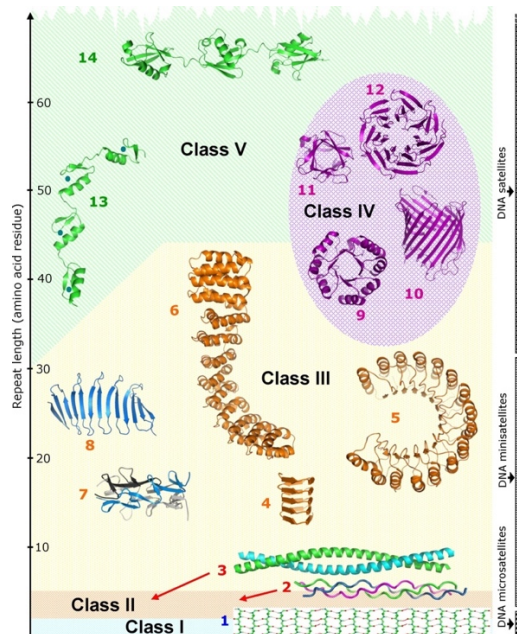
### 3.3.1. Clasificación de las proteínas con repeticiones en tándem

La clasificación de proteínas con repeticiones en tándem, que cuentan con estructura depositada en el PDB, se puede encontrar en la página RepeatsDB:Tandem repeat proteins detection and classification (<https://repeatsdb.bio.unipd.it/about>). Esta clasificación incluye cinco niveles [119]:

- 1) Clase: Depende de la longitud de la repetición y refleja una forma general, el modo de interacción entre los elementos repetitivos y el estado de oligomerización.
- 2) Topología: Se distingue por el tipo de estructura secundaria en la unidad repetitiva.
- 3) Plegamiento: Variantes de la topología que presentan diferencias estructurales en el número de elementos de estructura secundaria, elementos estructurales específicos adicionales y disposición estructural general.
- 4) Clan: sub-plegamiento que agrupa estructuras proteicas que tienen un motivo de secuencia común dentro de la repetición (o una parte de la repetición). Los motivos de secuencia se conservan debido al mismo plegamiento tridimensional y no a un ancestro común.
- 5) Familia: Grupo de proteínas con un antepasado común basado en la similitud de secuencias.

Estas proteínas se clasifican tomando en cuenta la longitud de las repeticiones en: **1)** Agregados cristalinos, formados por regiones con repeticiones de uno o dos residuos de longitud como el caso de la alfa keratina; **2)** Repeticiones fibrosas, estabilizadas por interacciones entre cadenas con repeticiones de 3-7 residuos como la fibrina de la lamprea; **3)** Repeticiones elongadas, con estructuras alargadas con repeticiones de 5-40 residuos en las que unidades repetitivas se requieren mutuamente para mantener la estructura, tal es el caso de la pectato liasa de *Pectobacterium carotovorum*; **4)** Repeticiones cerradas, compuestas por estructuras cerradas (no alargadas) con repeticiones de 30-60 residuos y **5)** Cuentas en una cuerda, estructuras cerradas (no alargadas) con repeticiones de 30-60 residuos en las

que las unidades repetitivas se necesitan unas a otras y están dispuestas circularmente con más de 50 residuos, lo bastante grandes para plegarse independientemente en dominios estables (Figura 6) [117].



**Figura 6.** Clasificación estructural de proteínas basadas en la longitud de sus repeticiones. Clase I, 1-estructura  $\beta$  cristalina de la polialanina. Clase II, 2-Triple hélice del colágeno, 3-espiral  $\alpha$ -helicoidal. Clase III, 4 –  $\beta$ -solenoide, 5-solenoide  $\alpha/\beta$  de la proteína LRR, 6-solenoide  $\alpha$ , 7-trímero de espiral  $\beta$ , 8-estructura  $\beta$  antiparalela. Clase IV, 9-Barril TIM, 10-barril  $\beta$  transmembranal, 11-Estructura de trébol  $\beta$ , 12-Estructura de propela  $\beta$  y Clase V, 13-Dominio de dedos de zinc, 14-cadena de poli ubiquitina. Imagen tomada de Kajava A. 2012 [117].

### 3.3.2. Lectinas con repeticiones en tándem

Las lectinas poseen una especificidad única y un gran potencial para aplicaciones biomédicas y biotecnológicas. En particular, las lectinas con péptidos de repetición corta, que forman dominios de unión a carbohidratos, no sólo son de gran interés para comprender la evolución de las proteínas, sino que también pueden utilizarse como andamiaje para diseñar nuevos receptores. La glicobiología sintética proporciona las

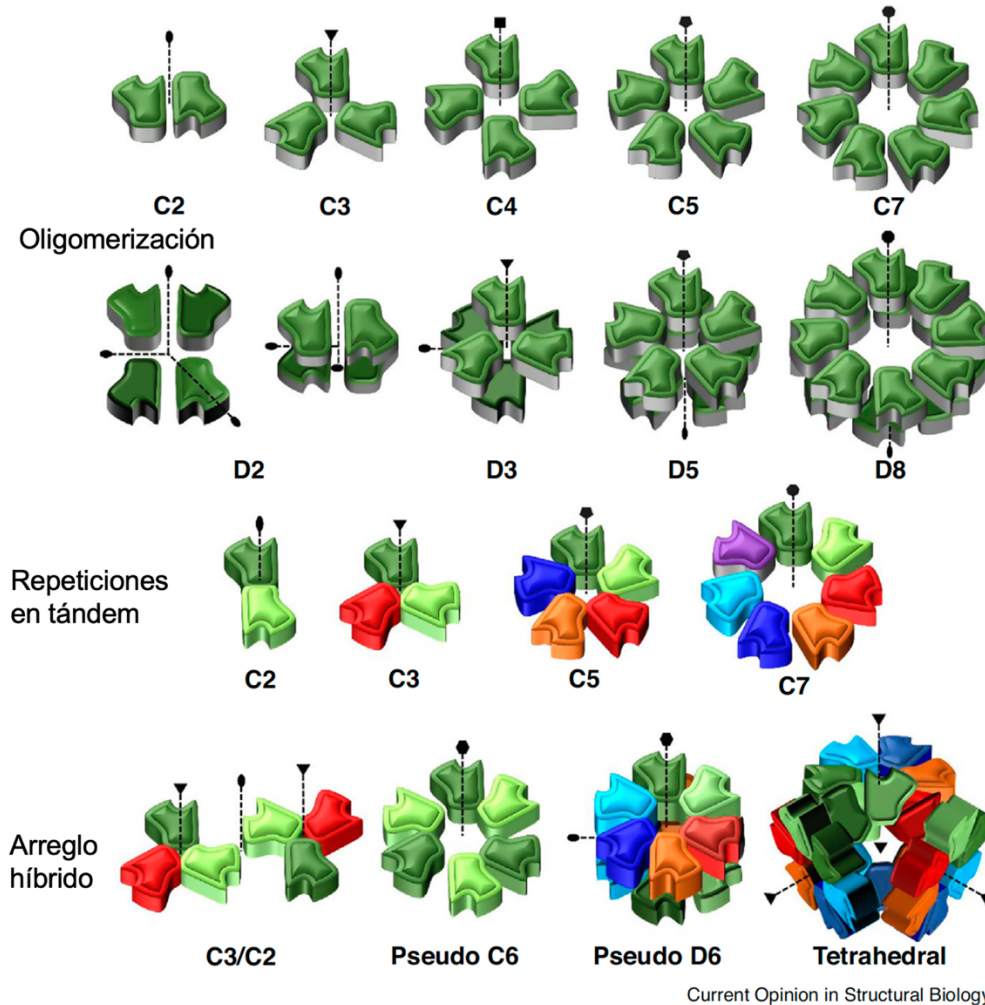
herramientas para diseñar la especificidad de las lectinas, así como su estructura, multivalencia y topología [114]. En general las lectinas son multivalentes, es decir, presentan varios sitios de unión para unirse simultáneamente a varios carbohidratos. Aun cuando la multivalencia suele asociarse a una gran avidéz por los glicoconjugados, la topología de los sitios de unión en el espacio también es importante para generar una gran especificidad, por ejemplo, hacia las ramificaciones de los N-glicanos y para la función biológica. Recientemente se ha demostrado que la disposición específica de los sitios de unión de las lectinas favorece la formación de grupos de glicolípidos y, por tanto, afecta a la estructura y la dinámica de las membranas celulares [120].

Las propiedades multivalentes de las proteínas están relacionadas con sus formas altamente simétricas. En la mayoría de los casos, la simetría de una estructura proteica se debe a la oligomerización de protómeros idénticos. En las lectinas oligoméricas, los DRC suelen estar repartidos en distintas direcciones, lo que da lugar a estructuras con simetría diedra (D2 a D8) o tetraédrica (Figura 7). Por otro lado, la simetría cíclica genera estructuras en las que todos los DRC se encuentran en el mismo lado, por lo que están perfectamente adaptadas para unirse a superficies glicosiladas, como las membranas celulares. Existen pocas lectinas cíclicas oligoméricas que adopten la simetría C2 a C7 (Figura 7). Otras lectinas generan multivalencia mediante una estrategia diferente, utilizando repeticiones en tándem para crear simetrías cíclicas y diédricas. En estos casos, se ha demostrado que el origen evolutivo de la simetría estructural se debe a la duplicación de genes y a eventos de fusión [120].

Según la base de datos Unilectin3D (<https://www.unilectin.eu/>) [121] la mayoría de las lectinas son oligoméricas como un mecanismo para obtener la multivalencia sin embargo, otras la adquieren mediante la presencia de repeticiones en tándem. Algunas, como la lectina del plátano, presentan una simetría C2 como aquellas que presentan el plegamiento Prisma  $\beta$  I. Otras presentan repeticiones en tándem con simetría C3, como en los plegamientos de Prisma  $\beta$  II en las lectinas de monocotiledóneas y las que tiene el plegamiento de trébol  $\beta$ , como en las lectinas del género *Mytilus*. La simetría cíclica mayor se observa en las lectinas con plegamiento

de propela  $\beta$ , como la taquilectina 1 de *Tachypleus tridentatus*. Es importante notar que la combinación de la repetición en tándem y la oligomerización da como resultado una valencia aún mayor (Figura 7).

Un ejemplo interesante del arreglo en repeticiones en tándem lo constituye la taquilectina 1 de *Tachypleus tridentatus*, que se caracteriza por tener seis secuencias repetidas en tándem de 31 aminoácidos de longitud cada una, formando una estructura de propela  $\beta$ . De forma similar se encuentra la taquilectina-2 (PDB: 1TL2) que presenta cinco secuencias de 47 aminoácidos repetidas en tándem, ambas involucradas en el reconocimiento y eliminación de patógenos [122,123]. De las lectinas que poseen un plegamiento de trébol  $\beta$ , se ha descrito ampliamente a CEL-III de *Cucumaria echinata* que presenta actividad hemolítica y citotóxica, esta puede contribuir a la defensa del organismo en contra de depredadores. Esta actividad es mediada por el reconocimiento de la proteína a Gal/GalNac en las células blanco, seguido por la formación de poros que permiten la permeación de iones en la membrana celular a través de la oligomerización de la proteína. El mecanismo formador de poros también se ha propuesto en distintas bacterias que producen toxinas formadoras de poros [124,125]. Esta característica de CEL-III se ha utilizado como nueva estrategia para el control de la malaria en mosquitos transgénicos [86].

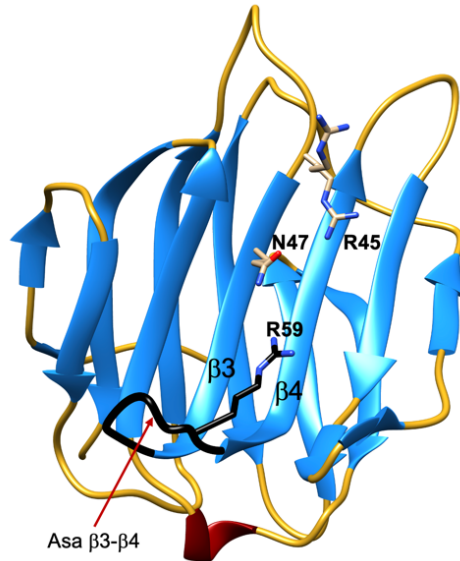


**Figura 7.** Representación esquemática de diferentes tipos de simetría observados en lectinas oligoméricas, lectinas con repeticiones en tándem y lectinas híbridas. Los dominios con secuencias idénticas están representados con colores idénticos. La simetría de la familia está indicada debajo de cada imagen y el eje de simetría (o pseudo eje de simetría) está representado por líneas punteadas. Imagen tomada de Notova S. et al. 2020 [120].



### 3.3.3. Galectinas con repeticiones en tándem

El DRC canónico de las galectinas tiene aproximadamente 130 aminoácidos, y siete residuos forman el sitio de unión al galactósido. Tiene un plegamiento de sándwich  $\beta$  que se compone de dos láminas  $\beta$  compuestas por 5 o 6 hebras antiparalelas (Figura 8). De las 15 familias de galectinas identificadas en mamíferos sólo en la Gal-4, 6, 8, 9 y 12 se han observado repeticiones en tándem. Estas contienen dos DRC distintos conectados por un asa. Mientras que Gal-4 se ha implicado en enfermedades inflamatorias intestinales, ya sea como mediador pro-inflamatorio o pro-apoptótico, Gal-6 posee un asa más pequeña que conecta sus dos DRC comparada con Gal-4 y se sabe muy poco de su función, Gal-8 interviene en enfermedades autoinmunes, como la artritis reumatoide y el lupus eritematoso y modula la progresión tumoral. Gal-9 controla la inflamación alérgica y la autoinmunidad mediada por Th1/Th17 y tiene valor pronóstico en ciertos tipos de tumores. Por último, Gal-12 desempeña un papel importante en la fisiología de los adipocitos [126]. Sólo en moluscos acuáticos se han encontrado lectinas con 4 repeticiones no idénticas en tándem que se ha observado tienen una gran importancia en la inmunidad innata al reconocer patógenos y parásitos y promover su fagocitosis. También pueden unirse selectivamente a componentes del fitoplancton, lo que sugiere que también participan en la captación y digestión intracelular de microalgas [107].



**Figura 8.** Galectina 8 de humano. Esta proteína tiene dos dominios de reconocimiento a carbohidratos, aquí se muestra el dominio N terminal, Código PDB 5AP5. Se observa el plegamiento tipo sándwich  $\beta$  típico de estas lectinas. En negro se muestra el asa y los residuos responsables del reconocimiento del galactósido en representación de palitos.

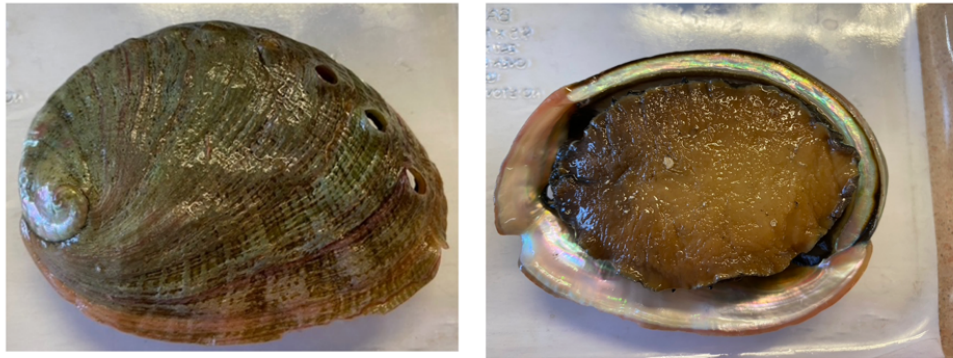
### **3.4. Descripción del abulón rojo, *Haliotis rufescens*.**

El abulón rojo (Figura 9) es un molusco marino gasterópodo que se distribuye desde California, USA hasta Baja California, México con un mayor número de ejemplares en la bahía de El Rosario. Posee un cuerpo suave rodeado de un manto, una cabeza anterior y un pie grande o musculo abductor. Su concha es gruesa, oval, color rojo ladrillo, presenta 3 a 4 poros respiratorios y puede llegar a medir hasta 27 cm. Su pie muscular permite que se adhiera a las superficies rocosas que se encuentran en su hábitat. Durante la noche se alimentan de algas por lo que en ese período pueden llegar a tener mayor actividad [127].

Este molusco constituye un recurso de alto valor económico alcanzando una producción anual de hasta 362 toneladas métricas [128]. Estos organismos tienen una gran importancia ecológica en los bosques de algas y sirven como alimento para las nutrias. En consecuencia a su sobreexplotación ha aumentado la mortandad y el

riesgo de sufrir el síndrome de marchitamiento [129]. En un esfuerzo por compensar la amenaza de la explotación, se ha logrado cultivar de forma eficiente a este molusco lo que ha permitido estudiar su genómica y profundizar en su naturaleza y biología, sin embargo la información que existe del abulón aún se mantiene incipiente [130,131].

El estudio de los moluscos es de gran interés ya que son el segundo filo más diverso de los invertebrados tanto en la tierra como en los océanos, se han descrito más de 100,000 especies con muchas diferencias morfológicas, en talla y de hábitat entre ellos. En este filo, dos de las clases más diversas son los gasterópodos y los bivalvos, los cuales contienen el 98% de las especies vivientes conocidas [132].



#### Clasificación taxonómica

Nombre científico: *Haliotis rufescens* Swainson, 1822  
Reino: Animalia  
Filo: Mollusca  
Clase: Gastropoda  
Subclase: Prosobranchia  
Orden: Archaeogastropoda  
Superfamilia Pleurotomariacea  
Familia: Haliotidae  
Género: Haliotis

**Figura 9.** Abulón rojo *Haliotis rufescens* Swainson, 1822. Molusco marino gasterópodo cultivado en Baja California, México.

## 4. JUSTIFICACIÓN

Las lectinas tienen una amplia aplicación biomédica y biotecnológica debido a la gran cantidad de funciones en las que están involucradas. De ahí la importancia de buscar lectinas con nuevas especificidades y propiedades únicas de actividad, estabilidad y especificidad. En invertebrados marinos se han estudiado muy poco este tipo de proteínas y se han encontrado nuevas lectinas que no comparten similitud con ninguna de las reportadas a la fecha. Debido a que estos organismos carecen de un sistema de defensa adaptativo, su respuesta inmune se basa principalmente en estas proteínas, las cuales forman parte del componente humoral en el que juegan un papel importante en la defensa contra patógenos. Debido a su capacidad para aglutinar células, se ha encontrado que en estos organismos las lectinas participan en procesos de reconocimiento y encapsulación de patógenos. Estas funciones biológicas han permitido el uso de las lectinas para desarrollar nuevos agentes terapéuticos que ayuden a combatir enfermedades provocadas por diversos patógenos. Un tipo muy importante de lectinas son las galectinas y los moluscos son los únicos organismos donde se han encontrado galectinas con cuatro dominios no idénticos en tándem. Existe muy poca información sobre aspectos bioquímicos, estructurales, funcionales y evolutivos de galectinas de moluscos, en especial con repeticiones en tándem, por lo que este estudio aporta información valiosa de esta nueva familia de lectinas.

## 5. HIPÓTESIS

La galectina de *H. rufescens* es una lectina multiespecífica que pertenece a una nueva familia de galectinas.

## 6. OBJETIVOS

### 6.1 *Objetivo general*

Caracterizar estructural, funcional y evolutivamente a una galectina de cuatro dominios en tándem del molusco marino *H. rufescens*.

### 6.2 *Objetivos específicos*

- Purificar a una galectina del *H. rufescens* y determinar su actividad estabilidad y especificidad por glicanos.
- Obtener la secuencia de la lectina para expresarla de forma recombinante y analizar la actividad, especificidad y estabilidad.
- Obtener información estructural que permita dilucidar el mecanismo de reconocimiento de sus ligandos.
- Analizar la historia evolutiva de las galectinas de invertebrados y vertebrados reportadas con uno, dos, tres y cuatro dominios.

## 7. MATERIAL Y MÉTODOS

### 7.1 Purificación de la galectina de *H. rufescens* (HrGal)

Los abulones rojos (*Haliotis rufescens*) se obtuvieron en Ensenada, Baja California. El hepatopáncreas del organismo se extrajo y homogenizó en 100 mL de amortiguador de fosfato de sodio 50 mM, pH 7.4, 150 mM de NaCl (PBS). El extracto obtenido se centrifugó a 3,500 x g por 30 min en una centrifuga Beckman Coulter con un rotor JA-14, y el clarificado se obtuvo con una segunda centrifugación a 15,000 x g por 30 min en un rotor JA-17. El sobrenadante se aplicó a una resina de afinidad de D(+) Galactosa-agarosa (Thermo Fisher Scientific), previamente equilibrada con el amortiguador PBS. La galectina se eluyó con 250 mM de D (+) galactosa en PBS. Posteriormente la galectina se dializó exhaustivamente contra el PBS para eliminar el ligando y poder medir su actividad.

### 7.2. Determinación de la pureza de la galectina

La pureza de la galectina se verificó por dos métodos, la electroforesis SDS-PAGE y por espectrometría de masas MALDI-TOF. Esta última nos permitió verificar también la masa de la lectina.

#### 7.2.1. Electroforesis SDS-PAGE

Las distintas fracciones obtenidas durante el proceso de purificación se analizaron mediante electroforesis en condiciones desnaturalizantes en un gel de poliacrilamida al 12%. La galectina se desnaturalizó y redujo en presencia del amortiguador de carga 1X ( $\beta$ -mercaptoetanol 10%, azul de bromofenol 0.05%, glicerol 20%, SDS 10%, Tris-HCl 0.125 M, pH 6.8). La migración electroforética se llevó a cabo con un voltaje constante de 90 V por 30 min y después una hora a 120 V en amortiguador de electroforesis 1X, en un sistema electroforético de BioRad® (Mini-PROTEAN®). Como referencia de migración se utilizó el marcador de masa molecular conocida (SDS-PAGE Molecular weight Standards BioRad®, Broad Range No. 161-0317). La visualización de las bandas proteicas se realizó mediante una tinción con Coomassie R-250.

### 7.2.2. Espectrometría de masas

Para verificar la pureza de la galectina y determinar la masa exacta se utilizó la espectrometría de masas MALDI-TOF (desorción/ionización laser asistida por una matriz-tiempo de vuelo) en un equipo Microflex Bruker Scientific LLC (Billerica, MA, USA) misma que nos permitió determinar la masa molecular de la proteína. Para poder utilizar la espectrometría de masas como criterio de pureza es necesario considerar los límites de detección del equipo, que en nuestro caso es de 50 pmol/mL y, por consiguiente, la masa de la proteína, mientras más grande es la proteína se requiere más para poder verla y para poder ver los contaminantes. Por lo que al tratarse de una proteína de alrededor de 64 kDa como se observó en la electroforésis, se requería una concentración mínima de 3.2 mg/mL para poder detectarla en el MALDI-TOF. Sin embargo, para poder detectar contaminantes de masa mayor se utilizó una concentración de 7 mg/mL y se analizaron varios intervalos de masa molecular: se buscaron proteínas de masa molecular pequeña (menor a 40 kDa), mediana (de 40 a 70 kDa) y grande (de 70 a 110 kDa) con el fin de detectar cualquier contaminante. Además, se utilizaron dos matrices, una de ácido sinapínico y una matriz universal, esto con el fin de detectar proteínas con diferente composición. Ambas matrices se prepararon como soluciones saturadas disueltas en 30% acetronitrilo, 70% agua y 0.1% de ácido trifluoroacético. Para calibrar el espectrómetro se utilizó albumina sérica bovina (BSA) (66.4 kDa) y la taumatina (22 kDa). La concentración de la proteína fue determinada con el método de BCA siguiendo las indicaciones del fabricante (Protein Assay Kit Thermo Scientific, Rockford, IL, USA) y utilizando a la BSA como proteína estándar.

### **7. 3. *Formalinización de los eritrocitos de conejo***

Se colectó sangre de conejo en una solución de NaCl 0.9% y se centrifugó a 2,500 x g por 15 min. Se realizaron varios lavados con la solución de NaCl hasta obtener un sobrenadante transparente. Los eritrocitos de conejo que se utilizaron en las pruebas de actividad, se formalizaron siguiendo el protocolo reportado por Fragkiadakis [133].

Los eritrocitos se suspendieron en un amortiguador de fosfatos 50 mM pH 7.4, 150 mM NaCl para obtener una solución 20 % v/v. En una bolsa de diálisis con tamaño de corte de 12.5 kDa, se agregan 25 mL de formaldehído al 37 % a pH 5.0, y se sumergen en 100 mL de esta solución de eritrocitos. Esta solución se colocó y se sumergió en los eritrocitos dejando interaccionar por 12 h a 4°C con agitación suave. Transcurrido este tiempo, la bolsa se perforó con una aguja y se liberó el formaldehído, lo que permite una mejor fijación de los eritrocitos, dejando interaccionar la solución de formaldehído durante 36 h a 4°C con agitación suave.

Después de la fijación, los eritrocitos de conejo formalinizados (ECF) se centrifugaron a 2500 x g, durante 15 min a 4°C y se lavaron exhaustivamente con una solución 0.9 % NaCl para eliminar los residuos de formaldehído. Con el fin de eliminar los grupos aldehídos libres de la membrana celular, los ECF (20 % v/v) se suspendieron en el amortiguador de fosfatos 50 mM pH 7.4, 150 mM NaCl con 2 % p/v de BSA (Albúmina de suero de bovino) y se incubaron 24 h a 4°C con agitación suave. Para eliminar la BSA, los eritrocitos se centrifugaron a 2,500 x g por 15 min a 4°C, y se lavaron exhaustivamente con la solución de 0.9% NaCl (las veces que sea necesario hasta que la solución se encuentre transparente). Finalmente, los eritrocitos (20% v/v) se resuspendieron con NaCl 0.9% y NaN<sub>3</sub> 0.05% pH 7.2, y se conservaron a 4°C hasta su uso.

#### **7.4 Ensayos de eritroaglutinación**

Para determinar la actividad de la lectina se realizaron ensayos de eritroaglutinación con ECF. Para determinar si HrGal es dependiente de calcio, la galectina purificada se pasó por una resina quelante (Chelating resin HR2-312 de Hampton Research) y posteriormente se analizó la actividad eritroaglutinante. De igual forma se determinó el efecto de diferentes iones y aditivos en la actividad de la lectina añadiendo 5 mM de Mg<sup>+2</sup>, Mn<sup>+2</sup>, Ca<sup>+2</sup>, Co<sup>+2</sup>, Cu<sup>+2</sup> y Zn<sup>+2</sup>, DTT 2 mM e imidazol 500 mM.

Los ensayos de actividad se realizaron en placas de 96 micro pozos con fondo en forma de "U", siguiendo un esquema de diluciones dobles seriadas del pozo 1 al 12. Se colocaron 50 µL de amortiguador de fosfato de sodio 50 mM, pH 7.4, se adicionaron 50 µL de la proteína (0.2 mg/mL), se homogenizó y diluyó en serie y



posteriormente, se agregaron 50  $\mu$ L de la solución de ECF al 2% a cada pozo, las cajas se agitaron por cinco minutos, se sellaron y se dejaron a temperatura ambiente por 90 min. Transcurrido ese tiempo se examinaron visualmente las placas en busca de aglutinación de los eritrocitos. Las pruebas se realizaron por triplicado y se utilizó como control positivo la aglutinina de germen de trigo y como control negativo el amortiguador de fosfatos con los aditivos que se analizaron. La actividad aglutinante se define como el recíproco del título, que corresponde al factor de dilución (número de pozo) más alto de proteína que presenta hemaglutinación [134].

### **7.5. Determinación de la especificidad**

La especificidad de la lectina se estimó por dos métodos: mediante ensayos de inhibición de la eritroaglutinación y por termofluorescencia.

#### **7.5.1. Ensayos de inhibición de la eritroaglutinación**

Con el fin de determinar que azúcares reconoce HrGal se realizaron pruebas de inhibición con monosacáridos, disacáridos oligosacáridos y polisacáridos e identificamos a los carbohidratos que podían inhibir su actividad hemoaglutinante sobre los eritrocitos de conejo. De igual forma que en los ensayos de eritroaglutinación, se utilizaron placas de 96 micro pozos con 25  $\mu$ L de amortiguador, se realizaron diluciones dobles seriadas del azúcar hasta el pozo 12. A cada pozo se le agregó la misma concentración de proteína (0.2 mg/mL) en un volumen de 25  $\mu$ L. Posteriormente se agitaron por 5 min y se dejaron a temperatura ambiente por 30 min, para permitir la interacción entre la proteína y su ligando. Después de la incubación con el azúcar se adicionaron 25  $\mu$ L de eritrocitos de conejo al 2%, se agitó por 5 minutos y se dejó incubando durante una hora a temperatura ambiente.

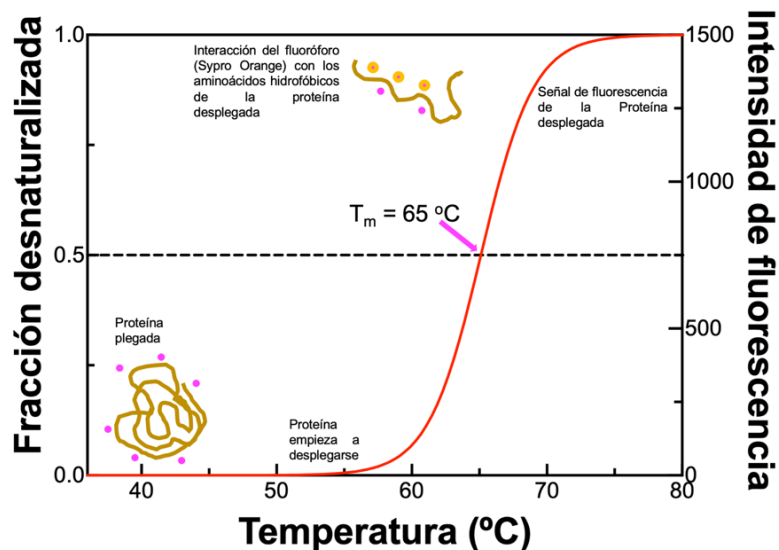
Para todos los azúcares probados, se consideró que la dilución mínima inhibitoria de la actividad era el último pozo en el que visualmente el sedimento de eritrocitos al fondo de los pozos y la zona clara alrededor de estos era de igual tamaño que la del blanco positivo. La concentración mínima inhibitoria (CMI) se definió como la concentración mínima de azúcar que permite la inhibición de la actividad hemoaglutinante.

**Tabla 2.** Azúcares empleados en los ensayos de especificidad de HrGal

<b>Azúcar</b>	<b>Concentración</b>	<b>Azúcar</b>	<b>Concentración</b>
Agar	0.01% (m/v)	Manitol	250 mM
Agarosa	0.01% (m/v)	Melibiosa	250 mM
Ácido alginico	0.01% (m/v)	N-acetil-D-glucosamina	250 mM
Amilopectina	0.05% (m/v)	N-acetil-D-galactosamina	250 mM
Carboximetil celulosa	0.01% (m/v)	$\alpha$ -L-ramnosa	250 mM
Quitina	0.01% (m/v)	$\alpha$ -Lactosa	250 mM
Quitosán	0.01% (m/v)	$\beta$ -Lactosa	250 mM
Fucoidán	0.002% (m/v)	$\beta$ -ciclodextrina	0.01% (m/v)
D(-) Arabitól	250 mM	$\kappa$ -Carragenano	0.002% (m/v)
D(-) Arabinosa	250 mM	$\iota$ -Carragenano	0.002% (m/v)
D(+) Celobiosa	250 mM	$\lambda$ -Carragenano	0.002% (m/v)
D(+) Fucosa	250 mM	Pectina	0.002% (m/v)
D(-) Fructosa	250 mM	Ácido poligalacturónico	0.002% (m/v)
D(+) Galactosa	250 mM	Rafinosa	250 mM
D(+) Ácido galacturónico	250 mM	Salicina	250 mM
D(+) Glucosa	250 mM	Almidón	0.05%
D(+) Ácido glucurónico	250 mM	Sacarosa	250 mM
D(+) Maltosa	250 mM	Xilano	0.01% (m/v)
D(+) Manosa	250 mM	Lactulosa	250 mM
D(+) Trehalosa	250 mM	Laminarina	0.002%
D(+)Xilosa	250 mM	Manan	0.002% (m/v)
Heparina	410 mM		

### 7.5.2. Fluorimetría diferencial de barrido

La fluorimetría diferencial de barrido (Differential Scanning Fluorimetry, DSF) o termofluorescencia es una técnica muy poderosa que permite evaluar la estabilidad de proteínas que se encuentren en diferentes ambientes (pH, iones, amortiguadores, aditivos, ligandos, etc) la interacción proteína-ligando, e incluso realizar un análisis isotérmico de la interacción proteína ligando. Esta consiste en utilizar un fluoróforo (el más utilizado es el SYPRO Orange) que interactúa con las regiones hidrofóbicas de la proteína. En el inicio del experimento, la proteína está plegada, por lo que la mayor cantidad de aminoácidos hidrofóbicos se encuentran en el núcleo de la proteína. Cuando esta es sometida a un incremento de temperatura se empieza a desplegar y el fluoróforo interactúa con los aminoácidos hidrofóbicos expuestos dando una señal de fluorescencia. Los cambios en la fluorescencia son monitoreados con un equipo de RT-PCR tiempo real. Después de obtener el fluorograma en función de la temperatura para cada condición, se determinó el intervalo de temperatura en el que ocurre la transición del estado plegado al desplegado. Estos datos se normalizan para obtener una curva sigmoide, que permite determinar la temperatura de fusión ( $T_m$ ) de la proteína; que corresponde al valor de temperatura en la que el 50% de las moléculas de proteína se encuentran plegadas y el 50% desplegadas (Figura 10) [135]. Para el análisis, se utilizó el programa StepOne v2.3 (Life Technologies, USA) y los datos de fluorescencia obtenidos ( $\lambda_{exc}=490$  nm y  $\lambda_{emis}=575$  nm) se emplearon para el análisis. Los datos de la transición fueron normalizados y se ajustaron a una distribución de Boltzmann utilizando el programa Prisma 9 (GraphPad Software Inc., USA) con el cual se obtuvo el punto de inflexión que corresponde al valor de  $T_m$  [136].



**Figura 10.** Esquema del ensayo de desnaturalización térmica utilizando la termofluorescencia. En un inicio la proteína está plegada y la sonda fluorescente (Sypro Orange) que se une a aminoácidos hidrofóbicos no emite una señal de fluorescencia. Conforme se incrementa la temperatura la proteína se despliega y la sonda se une a los aminoácidos hidrofóbicos expuestos emitiendo la señal de fluorescencia.

#### 7.5.2.1. Ensayo de especificidad

Para determinar la especificidad de la lectina a 1 mL de HrGal con una concentración de 0.15 mg/mL se le adicionó 1  $\mu$ L del fluoróforo SYPRO-Orange (1:1000). Se colocaron 10  $\mu$ L de la galectina en cada pozo de una placa de 96 (MicroAmp Fast, Applied biosystems) y a cada uno de ellos se le adicionaron 10  $\mu$ L de los diferentes azúcares que se utilizaron en los ensayos de inhibición de la eritroaglutinación (Tabla 2). La placa se selló con una película ópticamente inerte (Greiner VIEWseal de Jena Bioscience) y se centrifugó a 300 rpm por 2 min a 25 °C. Una vez centrifugada, la placa se colocó en un equipo de RT-PCR en tiempo real (Applied Biosystems StepOne) siguiendo un programa de inicio de 25°C por 2 min, seguido del incremento de temperatura de 1 grado cada minuto hasta llegar a 98°C.

#### 7.5.2.2. Ensayo de estabilidad

Esta técnica permite evaluar la estabilidad térmica de una proteína bajo diferentes

condiciones como son: pH, sales, aditivos, ligandos y agentes solubilizantes. Se sabe que muchas veces el pH de una solución no es el responsable de la estabilidad de una proteína sino la sal de dicho amortiguador. Es por esto que se evaluaron 96 condiciones variando el pH y la sal del amortiguador para encontrar las condiciones donde se obtuviera la mayor estabilidad. Asimismo, se probaron 96 aditivos del kit Aditive Screen, Hampton Research. El análisis de las curvas de desnaturalización se realizó como en el punto anterior. Posteriormente, se analizaron las condiciones que lograban estabilizar a HrGal con ensayos de actividad hemoaglutinante para así determinar si las condiciones de mayor estabilidad corresponden con las de mayor actividad.

#### *7.5.2.3. Desulfatación de la heparina y del carragenano*

Con el fin de evaluar el efecto de los grupos sulfato presentes en la heparina y en el carragenanos en el reconocimiento de HrGal, se llevó a cabo la desulfatación de la heparina (Biosynth Ltd., Reino Unido) ya que en el caso del carragenano no fue necesario desulfatar, ya que contábamos con tres tipos de este polisacárido que variaban en el grado de desulfatación ( $\kappa$ -,  $\iota$ - y  $\lambda$ -carragenano, siendo el primero el más sulfatado y el último el menos sulfatado). Para la desulfatación de la heparina se utilizaron dos métodos: En el primero, el polisacárido se secó por congelación en una solución alcalina, para descargar el grupo sulfato en la posición O-3 de GlcN<sub>2,3S</sub> y GlcN<sub>2,3,6S</sub>, según lo descrito por Santini et al. [137]. En el segundo método, la heparina se pasó por una resina de intercambio aniónico (Agilent Bio WAX, NP5,SS) en un sistema HPLC. La fracción que no se une se neutralizó con piridina. Esta sal de piridinio se trató con dimetilsulfóxido en 5% de metanol durante 1.5 horas a 50 °C para eliminar los grupos sulfato [138], posteriormente se pasó por una resina de intercambio cationico equilibrando con agua desionizada y siguiendo la absorbancia a 215 nm.

La desulfatación de la heparina se confirmó mediante análisis ATR-FTIR de muestras secas utilizando un espectrómetro FT-IR Nicolet iS50 (Thermo Fisher Scientific, USA). Se corrieron las líneas de base para todos los espectros. Estos espectros se normalizaron y suavizaron aplicando un filtro Savitsky-Golay de 5° orden. Por último,

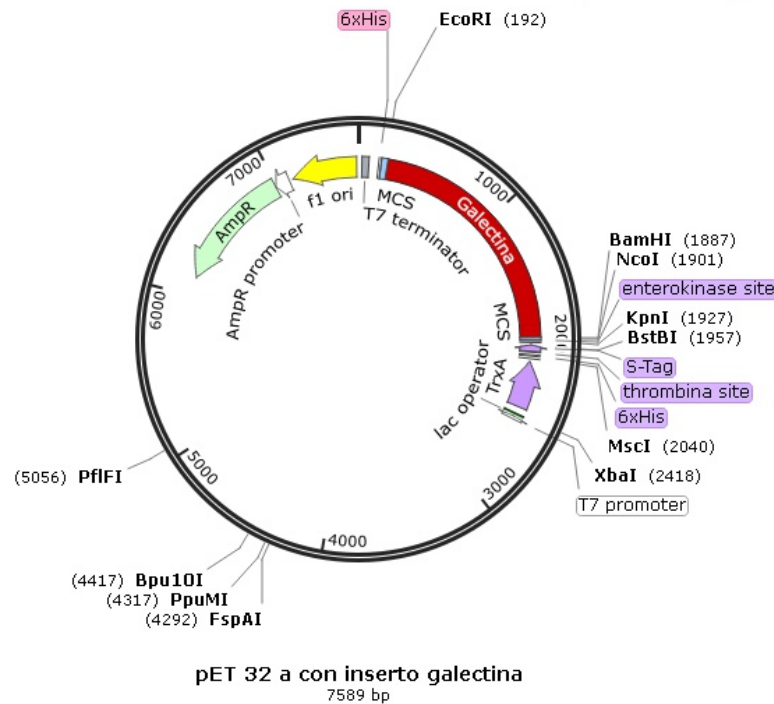
los espectros de los compuestos desulfatados se escalaron, se restaron de los espectros de los compuestos originales y se analizaron. Se registraron los picos de frecuencia que superaban el 50% de las ordenadas del espectro visible y los baricentros correspondientes a las bandas  $\nu_{as}S=O$ .

### **7.6. Determinación de la secuencia de aminoácidos**

La secuencia de aminoácidos de HrGal se determinó mediante la identificación de péptidos por espectrometría de masas. La proteína purificada se cargó en un gel SDS-PAGE al 12% teñido con azul de Coomassie R-250. La banda correspondiente a la lectina se cortó y destiñó con 200  $\mu$ L de una solución de  $NH_4HCO_3$  (250 mM en acetonitrilo al 50%), luego se redujo con 50 mM de TCEP (tris(2-carboxietil) fosfina) durante 10 min a 60 °C y las cisteínas se alquilaron con 100 mM de yodoacetamida, 1 h a 25 °C en la oscuridad. Posteriormente la lectina se sometió a una digestión con tripsina (Trypsin Singles, grado proteómico, Sigma-Aldrich) a 37 °C durante toda la noche. Los péptidos se mezclaron con una matriz de  $\alpha$ -ciano-4-hidroxicinámica en 30% de acetonitrilo, 70% de agua y 0.1% de ácido trifluoroacético para su posterior análisis con un espectrómetro MALDI-TOF Microflex Bruker. Para la calibración se utilizó el estándar de calibración de péptidos II de Bruker. Los espectros se registraron en modo reflectrón de iones positivos y se utilizó el software FlexAnalysis 3.0 (Bruker Daltonics) para el registro de espectros de masas y la detección de picos. Los fragmentos se analizaron con el programa MASCOT 2.4.

### **7.7. Clonación y expresión de la galectina recombinante de *H. rufescens* (rHrGal)**

El gen de la galectina (KJ183034) fue sintetizado en la compañía GenScript, la construcción se solicitó clonada en pET-32a con los sitios de restricción *KpnI*, en el extremo N-terminal y *EcoRI*, en el extremo C-terminal. El inserto posee la secuencia (ENLYFQG) de reconocimiento para el corte con la proteasa del virus del mosaico del tabaco (TEV) en el extremo N-terminal. La construcción de rHrGal se obtendrá como una proteína de fusión con tiorredoxina y una etiqueta de histidinas (Figura 11).



**Figura 11.** Inserción del gen de HrGal (KJ183034) en el vector de expresión pET-32a. El gen de 1665 pb fue insertado entre los sitios de restricción *KpnI* y *EcoRI*. Tamaño final del plásmido de 7601 pb. En rojo se marca la región que corresponde a la secuencia del inserto, en morado se señala a tiorredoxina (TrxA) y en verde al gen de resistencia del antibiótico a ampicilina (AmpR).

Una vez verificado el inserto se transformó en la cepa BL21 de *E. coli*. Se creció en medio LB con ampicilina 100 µg/ml a 37 °C con agitación de 180 rpm durante toda la noche. Se realizó una dilución 1:50 con medio fresco y se incubó nuevamente a 37 °C con agitación de 180 rpm, se monitoreó la densidad óptica hasta alcanzar 0.8 DO<sub>600</sub>, posteriormente se indujo la expresión con 0.5 mM de isopropil-β-D-1 tio galactopiranosido (IPTG), a una temperatura de 37°C y se continuó analizando la cinética de expresión durante 24 horas, tomando una muestra de cultivo a las 0, 2, 4, 6, 8, 12, 16, 20 y 24 horas después de la inducción. Posteriormente, se evaluaron las cepas BL21 y *Rosetta-gami* a temperaturas menores (18 y 30 °C) una vez que se indujeran con 0.25 mM de IPTG. La expresión

de rHrGal se monitoreo tomando muestras de 9 puntos distintos (0, 2, 4, 6, 8, 12, 16, 20 y 24 horas) después de la inducción. Cada muestra fue centrifugada a 15000 x g por 10 min a una temperatura de 4°C y se descartó el sobrenadante. Los botones obtenidos se lisaron con amortiguador de lisis (Tris-HCl 30 mM, pH 7.4, NaCl 200 mM y PMSF 1 mM) y se sonicaron a intervalos de 10 s con espacios de 30 s para completar un tiempo de 8 min, a 4°C. Posteriormente, se centrifugaron a 15000 x g por 10 min a una temperatura de 4°C y tanto el sobrenadante como el precipitado se analizaron en geles de poliacrilamida al 10% en condiciones desnaturalizantes.

### **7.8. Purificación de la rHrGal**

El botón celular, se lisó a intervalos de 10 s con espacios de 30 s para completar un tiempo de 8 min, a 4°C con 30 mL de amortiguador de lisis (Tris-HCl 30 mM, pH 7.4, NaCl 200 mM y PMSF 1 mM). El lisado se centrifugó a 26000 x g por 30 min a 4°C, para obtener un nuevo pellet en donde se encontraba la galectina de forma insoluble. Después, se identificó la menor concentración de urea que podía solubilizar a la galectina. Con el ensayo de solubilización en condiciones desnaturalizantes, agregando a los lisados de rHrGal distintas concentraciones de urea (0.5, 2, 4, 6 y 8 M) y dejando por 24 h a 4 °C para permitir la solubilización, luego se centrifugaron las muestras para obtener en el sobrenadante la fracción soluble y, en el precipitado la fracción que continuaba insoluble. Se tomaron muestras de cada una de las fracciones para cargarlas en un gel al 10%. De las distintas fracciones se seleccionó la condición en donde se encontraba la mayor cantidad de rHrGal soluble.

Después de haber identificado la condición de urea, el pellet obtenido se solubilizó durante toda la noche a 4°C con amortiguador de lisis desnaturalizante (Tris-HCl 30 mM, pH 7.4, NaCl 200 mM, urea 6 M, PMSF 1 mM), la fracción solubilizada resultante se sónico y centrifugó con las mismas condiciones descritas en la primera lisis. El sobrenadante obtenido, se filtró con un filtro de 0.45 µm y se inyectó en una columna de afinidad de níquel HisTrap™ HP columns (GE Healthcare, Uppsala, Sweden). La columna fue equilibrada con el amortiguador de equilibrio (Tris 50 mM, pH 7.4, 200 mM NaCl, 1mM de PMSF, 1-8 M de urea y 30 mM imidazol). Se realizó un gradiente de elución, probando distintas concentraciones de



imidazol desde 30 hasta 500 mM. Cada fracción se analizó por medio de electroforesis para determinar la pureza de la galectina. Posteriormente, se realizaron ensayos de replegamiento de rHrGal con un gradiente de urea, disminuyendo la concentración de urea (4, 3, 2, 1 y 0 M) en una bolsa de diálisis con tamaño de corte de 12.5 kDa, contra el amortiguador urea-TBS-glicerol-DTT por 24 horas a 4 °C. Posteriormente, se realizó una digestión enzimática para eliminar la etiqueta de histidinas. En este proceso, se dializó la proteína con el amortiguador de digestión (Tris 50 mM, pH 8, NaCl 300 mM, MgCl<sub>2</sub> 5 mM, DTT 0.25 mM, y 5% de glicerol) y la adición de la proteasa del virus del tabaco recombinante (TEV). La digestión se llevó a cabo a razón de 1:20 con la proteína a 4 °C toda la noche. Nuevamente, se realizó una purificación con la columna de Ni<sup>2+</sup> para retirar las histidinas y la tiorredoxina fusionada. Finalmente, se comprobó el corte por electroforesis SDS-PAGE 12% y por espectrometría de masas MALDI-TOF utilizando una matriz de ácido sinapínico, como se ha descrito anteriormente.

### **7.9. Dicroísmo circular**

Esta técnica nos permite a estimar el contenido de estructura secundaria y cambios conformacionales en una proteína. Debido a que las proteínas son macromoléculas ópticamente activas, muestran una absorción diferencial de la luz circular polarizada. El dicroísmo circular (DC) en el UV lejano, se basa predominantemente en la excitación de transiciones electrónicas en los grupos amida, por lo que se obtienen espectros de DC característicos para los diferentes elementos estructurales (hélices, hebras beta, giros y estructuras desordenadas) [127,128].

Las mediciones de DC de HrGal y rHrGal fueron realizadas en un espectropolarímetro J-1500 (Jasco Inc.), equipado con un Peltier como sistema de control de temperatura a 25 °C. Ambas proteínas HrGal y rHrGal se dializaron contra el amortiguador Mes-Citratos 10 mM, NaCl 10 mM a pH 7,4 y luego se filtraron con filtros de tamaño de poro de 0.22 µm. Los espectros en el UV lejano se midieron de 190 a 260 nm en una celda de cuarzo de 0.1 cm de longitud de trayectoria. Se promediaron tres repeticiones para obtener los espectros finales de la proteína a una concentración de 0.25 mg/mL. Los espectros de DC de la muestra fueron

normalizados y restada la línea base. Los espectros fueron expresados como dicroísmo circular molar ( $\Delta\epsilon$ ) en  $M^{-1}cm^{-1}$ , utilizando la siguiente fórmula:

$$\Delta\epsilon = \theta\lambda \times 0.1 \times MRW / L \times C \times 3298$$

Donde:

- $\Delta\epsilon$  = Dicroísmo circular molar ( $M^{-1}cm^{-1}$ )
- $\theta\lambda$  = Elipticidad en miligrados (mili°) a la longitud de onda correspondiente
- MRW = Masa media por residuo (Da) que se calcula: masa molecular de la proteína (Da)/(número de aminoácidos-1)
- L = paso de luz (cm)
- C = concentración de la proteína (g/ml)

Los espectros fueron escalados a elipticidad molar  $[\theta]$  ( $deg \cdot cm^2 \cdot d mol^{-1}$ ), empleando la ecuación 1.

Los resultados obtenidos se expresaron como elipticidad molar por residuo medio ( $[\theta]$  grados  $cm^2 d mol^{-1}$ ). La estimación del contenido de estructura secundaria se obtuvo mediante el análisis de los espectros con el servidor BeStSel [141]

### **7.10. Análisis del modelo estructural y estudios de acoplamiento molecular**

La estructura 3D de HrGal fue obtenida usando el programa de inteligencia artificial AlphaFold [142]. Una vez que se obtuvo la estructura, corregimos detalles geométricos con el programa Coot versión 0.8. Finalmente, la calidad estereoquímica general del modelo se determinó con el programa PROCHECK [143]. Las simulaciones de acoplamiento se realizaron con el servidor HDOCK [144], usando al  $\iota$ -carragenano como su ligando. Logramos identificar y analizar los residuos potencialmente involucrados en el reconocimiento del ligando con el programa Chimera [145].

### **7.11. Análisis filogenético de HrGal**

Las secuencias de galectinas tetradominio, fueron recuperadas usando el algoritmo BLASTP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), a partir de la base de datos de

proteínas no redundantes en el sitio de NCBI, usando la secuencia de aminoácidos de *H. rufescens* depositada en Genbank AHX26603 como anzuelo. El análisis incluyó búsquedas en TBLASTN contra el ADNc de 10 especies de moluscos con la secuencia del genoma completo para recuperar todas las galectinas existentes. Solo se incluyeron aquellas secuencias cuyo valor de  $E^1$  resulto menor a  $10^{-5}$ . Además, se identificaron y alinearon los dominios de reconocimiento a carbohidratos (DRCs) individuales de 66 especies de invertebrados y 16 de vertebrados, incluyendo las secuencias de humano, rata, ratón, y *Xenopus laevis*. Se consideraron un total de 448 DRCs de 209 galectinas para el estudio. En el caso de recuperar alguna secuencia de aminoácidos incompleta o que mostrara inserciones atípica o deleciones, se recuperaba también la secuencia de ADNc o genómico para predecir la estructura del gen intron/exón, con el software de predicción de genes, Softberry FGENESH+ [146] (<http://www.softberry.com/>), considerando la información de homología de proteínas cuya similitud se relacionaba a secuencias completas de galectinas. Las especies seleccionadas se encuentran enlistadas con sus números de accesos en la sección de anexos, Tabla S1 y S2. Todos los alineamientos de secuencias de aminoácidos se completaron usando ClustalX [147] y se corrigieron manualmente con BIOEDIT [148]. Para identificar la relación evolutiva de las galectinas de cuatro dominios con su secuencia completa, se construyó un árbol filogenético sin raíz, usando el método de máxima verosimilitud con el modelo de Le-Gascuel, y en el caso de los dominios individuales se usó el modelo de Whelan y Goldman, usando el programa MEGA X [149]. Se estimó la confiabilidad de la ramificación utilizando un bootstrap de 500 replicas.

---

<sup>1</sup> El valor de E es un parámetro que sirve de filtro para los resultados del BLAST. Describe el número de resultados que uno puede esperar ver por casualidad al buscar en una base de datos de tamaño particular. La coincidencia es mas significativa con valores mas bajos de E [179].

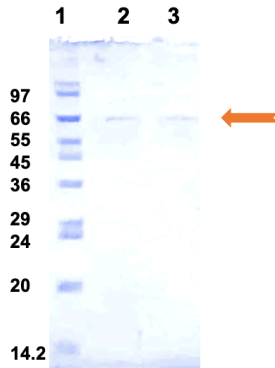
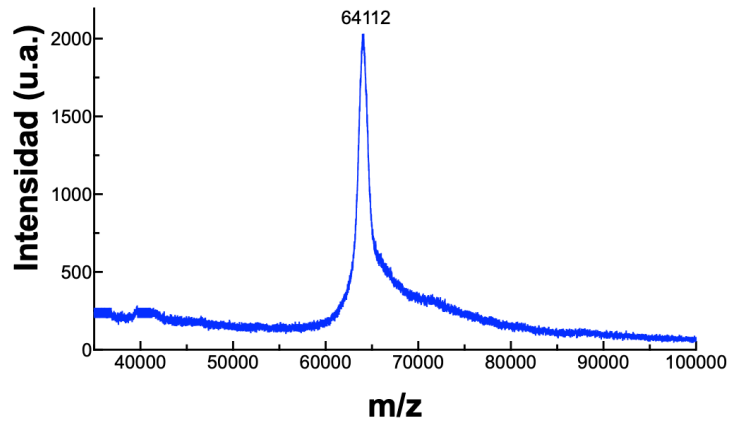
## 8. RESULTADOS Y DISCUSIÓN

Las galectinas son una familia de lectinas evolutivamente conservadas que han sido ampliamente estudiadas por su gran importancia biomédica. Sin embargo, a la fecha existe muy poca información sobre este tipo de proteínas de moluscos marinos que son los únicos organismos donde se han encontrado galectinas con cuatro dominios en tándem. Con el fin de aportar información valiosa de esta nueva familia de lectinas purificamos a una galectina del molusco marino *H. rufescens*, misma que se pudo obtener también en forma recombinante.

### **8.1. Purificación de HrGal a partir de la fuente natural**

Utilizando una resina de D-galactosa-agarosa se purificó una galectina del hepatopáncreas del abulón rojo *Haliotis rufescens* a la que denominamos HrGal. La pureza y masa de la lectina se verificó por electroforesis en condiciones desnaturizantes y reductoras y por espectrometría de masas MALDI-TOF (Figura 12A y 12B respectivamente) en donde se observó que la proteína tenía una masa molecular de 64.1 kDa. Para confirmar la identidad de HrGal, se determinó su secuencia de aminoácidos por digestión enzimática con tripsina y el posterior análisis de los péptidos por espectrometría de masas. La detección de los picos se realizó con el programa FlexAnalysis 3.0 (Bruker Daltonics) y el análisis de estos con el servidor MASCOT 2.4 (Matrix Science) instalado de forma local. Este análisis arrojó 8 péptidos con un score de 98% de precisión, uno de ellos corresponde con el extremo carboxilo terminal (Figura 13). Utilizando el genoma de *H. rufescens* que ya está reportado (GCA\_003343065.1), el servidor MASCOT identificó con un porcentaje de cobertura de 28.3 % a una proteína previamente reportada (GenBank AHX26603) [17] que corresponde a una galectina de *H. rufescens* que consta de 555 aminoácidos con una masa de 63.8 kDa y un punto isoeléctrico teórico de 4.73 de la que se tenía muy poca información. La diferencia en masa respecto a la obtenida experimentalmente con la proteína purificada del organismo puede deberse a alguna modificación post-traducciona, las cuales son muy comunes en estos organismos, principalmente las acetilaciones o las glicosilaciones. Con el fin de identificar en la secuencia de HrGal a los aminoácidos canónicos involucrados en el reconocimiento

del ligando (H,N,R,N,W,E,R) [150] se hicieron alineamientos con las secuencias de galectinas de diferentes fuentes que contaran con un código PDB para posteriormente poder analizar su estructura: De humano (PDB 5DUV), de ratón, *Mus musculus* ( PDB 2DYC), del pez *Conger myriaster* (PDB 1C1L), de la rana *Xenopus laevis* (PDB 3WUD), de vaca, *Bos taurus* (PDB 1SLT), de rata, *Rattus norvegicus* (PDB 3M2M) del nemátodo *Toxascaris leonina* (PDB 4HL0) y de pollo *Gallus gallus* (PDB 4WVV). En la Figura 14 observamos que los aminoácidos canónicos se conservan en HrGal (marcados con una estrella rosa). Sin embargo, fue interesante notar que, en la secuencia de HrGal dichos aminoácidos se repetían 4 veces con la única variación del último de ellos que podía encontrarse como K, K, R o L. Con el fin de determinar si HrGal pertenecía a la familia de proteínas con repeticiones en tándem, realizamos un análisis con el servidor Meta-Repeat-Finder (<https://bioinfo.crbm.cnrs.fr/index.php?route=tools>) [151] Este análisis mostró que efectivamente, HrGal presenta cuatro repeticiones, que constan de los aminoácidos: P<sup>10</sup> al N<sup>144</sup>, P<sup>147</sup> al D<sup>282</sup>, D<sup>283</sup> al L<sup>420</sup> y P<sup>421</sup> al F<sup>455</sup> respectivamente, con un porcentaje de similitud entre ellas de 58 %. Esto se puede observar claramente en la secuencia de aminoácidos donde se aprecia que las repeticiones no están compuestas por los mismos aminoácidos (Figura 15 y 16). Los resultados obtenidos con este servidor los validamos con el programa Tally-2.3 [152] que mostró que dichos resultados tenían un 99.6% de confiabilidad, esto implica que cada repetición en tándem cuenta con los aminoácidos canónicos para el reconocimiento de galactósidos y por tanto HrGal cuenta con 4 dominios de reconocimiento a carbohidratos o DRC's (Figura 16).

**A)****B)**

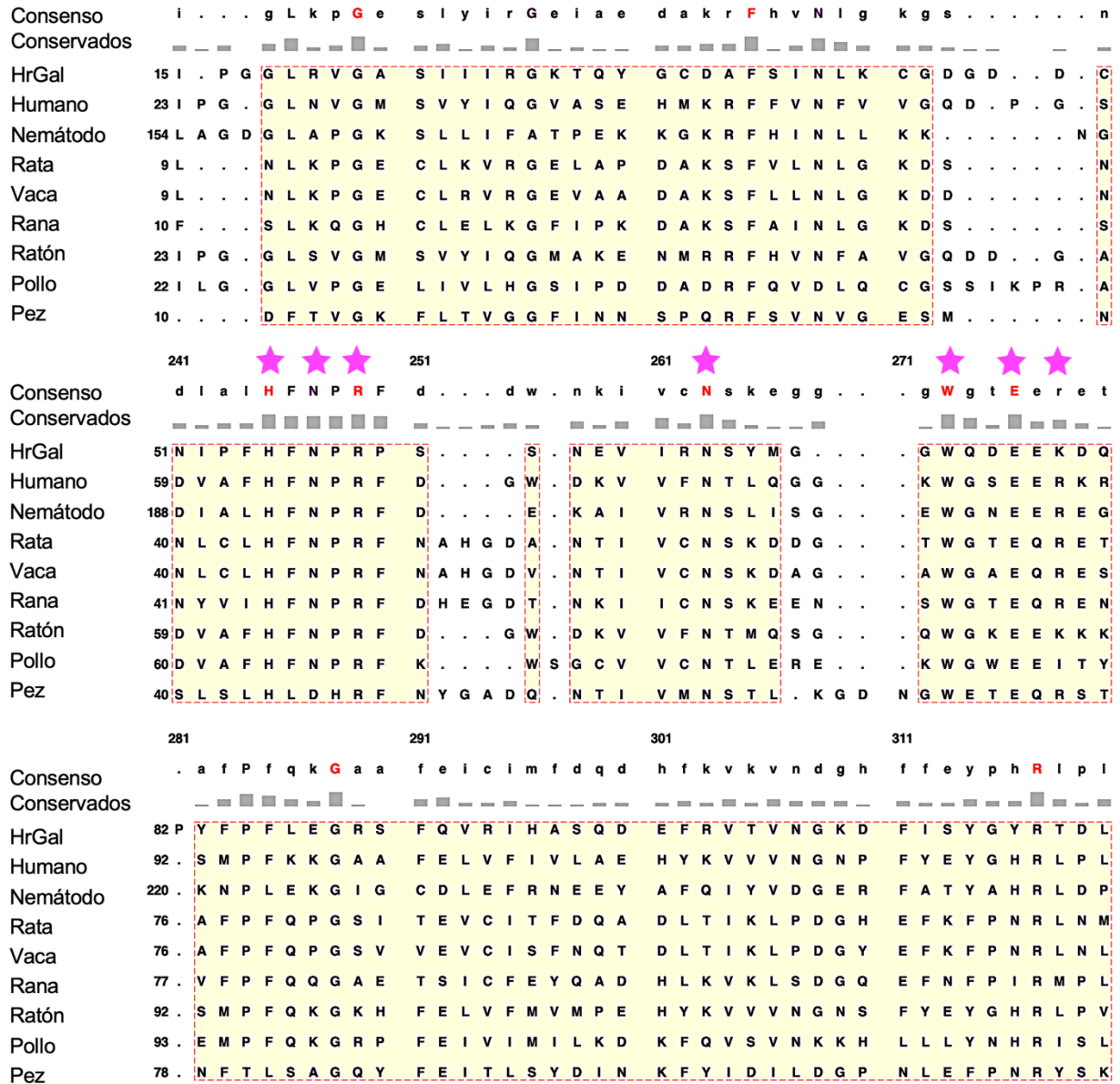
**Figura 12.** Purificación de HrGal. A) Electroforesis SDS-PAGE. Carril 1, Marcador de peso molecular, carril 2 y 3, lectina pura. La flecha muestra una banda de un peso aproximado de 66 kDa. B) Espectro de desorción/ionización laser asistida por una matriz-tiempo de vuelo donde se observa la masa de la lectina (64.1 kDa). La matriz utilizada fue el ácido sinapínico. Los estándares para la calibración del espectrómetro de masas fueron taumatina (22.0 kDa) y BSA (66.43 kDa).

```

YLQLGSGAEYS DVTIQNSMTNPYFEEF K
YLTINGDISICDVEFCEPLPDDFM K
NTHEDGSWGDEELEQPNFPFES K
MSFSYSNDPPEHIYIPGGL R
IPSNLQPGGWVVVNGTAK K
DQPYFPFLEGR R
VCHLMLIGDANFFEPEFF
GAMFDALFIAQED R

```

**Figura 13.** Péptidos identificados con el servidor MASCOT después de la hidrólisis de la proteína con tripsina.

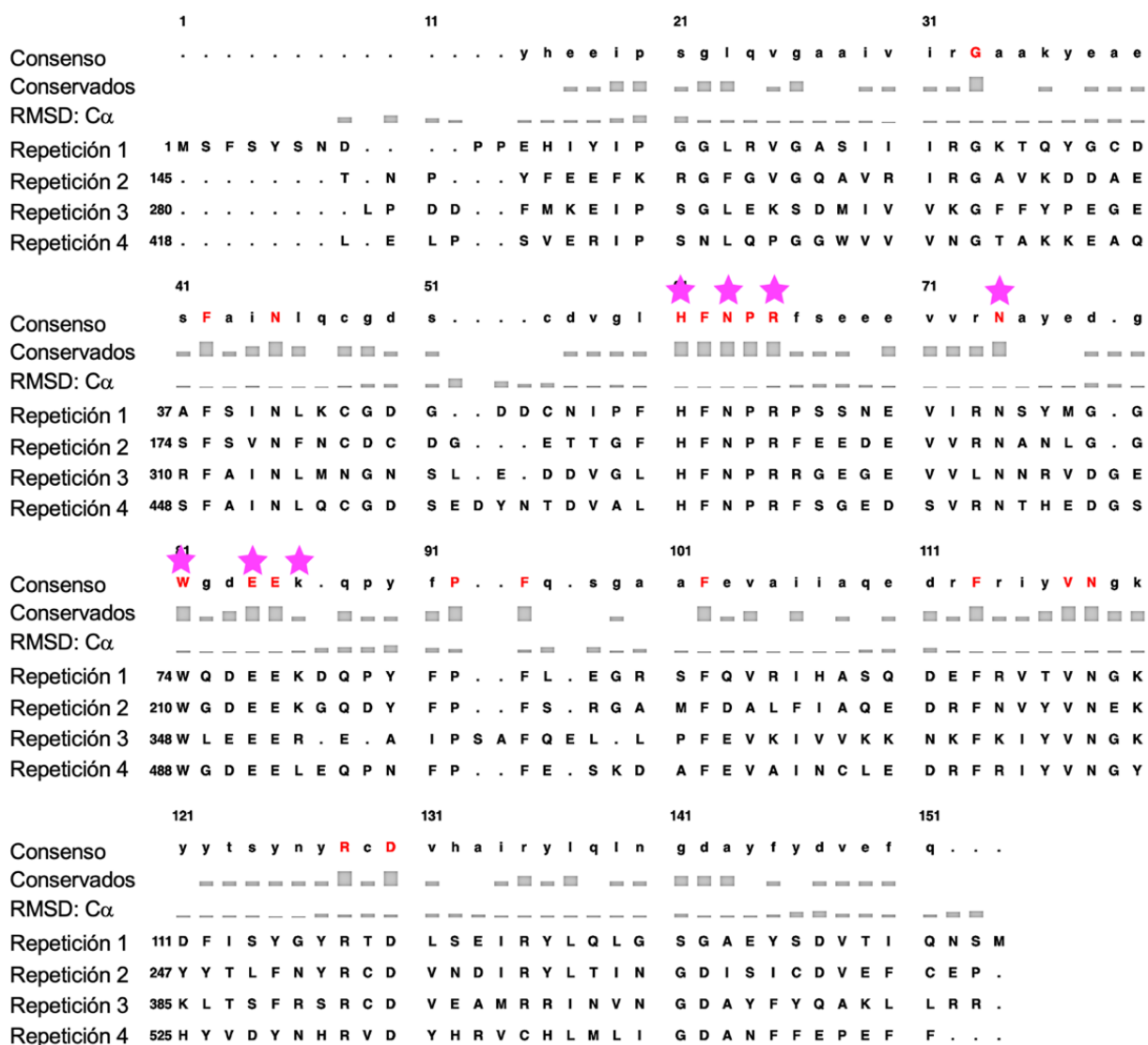


**Figura 14.** Alineamiento de secuencias de galectinas de diferentes organismos contra HrGal. Los aminoácidos conservados se señalan en letra color rojo. La estrella rosa muestra los aminoácidos canónicos involucrados en la interacción con el ligando. Las secciones enmarcadas en los cuadros rosa forman hebras beta en la estructura de cada galectina.

1	M	S	F	S	Y	S	N	D	P	P	E	H	I	Y	I	P	G	G	L	R	20
21	V	G	A	S	I	I	I	R	G	K	T	Q	Y	G	C	D	A	F	S	I	40
41	N	L	K	C	G	D	G	D	D	C	N	I	P	F	H	F	N	P	R	P	60
61	S	S	N	E	V	I	R	N	S	Y	M	G	G	W	Q	D	E	E	K	D	80
81	Q	P	Y	F	P	F	L	E	G	R	S	F	Q	V	R	I	H	A	S	Q	100
101	D	E	F	R	V	T	V	N	G	K	D	F	I	S	Y	G	Y	R	T	D	120
121	L	S	E	I	R	Y	L	Q	L	G	S	G	A	E	Y	S	D	V	T	I	140
141	Q	N	S	M	T	N	P	Y	F	E	E	F	K	R	G	F	G	V	G	Q	160
161	A	V	R	I	R	G	A	V	K	D	D	A	E	S	F	S	V	N	F	N	180
181	C	D	C	D	G	E	T	T	G	F	H	F	N	P	R	F	E	E	D	E	200
201	V	V	R	N	A	N	L	G	G	W	G	D	E	E	K	G	Q	D	Y	F	220
221	P	F	S	R	G	A	M	F	D	A	L	F	I	A	Q	E	D	R	F	N	240
241	V	Y	V	N	E	K	Y	Y	T	L	F	N	Y	R	C	D	V	N	D	I	260
261	R	Y	L	T	I	N	G	D	I	S	I	C	D	V	E	F	C	E	P	L	280
281	P	D	D	F	M	K	E	I	P	S	G	L	E	K	S	D	M	I	V	V	300
301	K	G	F	F	Y	P	E	G	E	R	F	A	I	N	L	M	N	G	N	S	320
321	L	E	D	D	V	G	L	H	F	N	P	R	R	G	E	G	E	V	V	L	340
341	N	N	R	V	D	G	E	W	L	E	E	R	E	A	I	P	S	A	F	360	
361	Q	E	L	L	P	F	E	V	K	I	V	V	K	K	N	K	F	K	I	Y	380
381	V	N	G	K	K	L	T	S	F	R	S	R	C	D	V	E	A	M	R	R	400
401	I	N	V	N	G	D	A	Y	F	Y	Q	A	K	L	L	R	R	L	E	L	420
421	P	S	V	E	R	I	P	S	N	L	Q	P	G	G	W	V	V	V	N	G	440
441	T	A	K	K	E	A	Q	S	F	A	I	N	L	Q	C	G	D	S	E	D	460
461	Y	N	T	D	V	A	L	H	F	N	P	R	F	S	G	E	D	S	V	R	480
481	N	T	H	E	D	G	S	W	G	D	E	E	L	E	Q	P	N	F	P	F	500
501	E	S	K	D	A	F	E	V	A	I	N	C	L	E	D	R	F	R	I	Y	520
521	V	N	G	Y	H	Y	V	D	Y	N	H	R	V	D	Y	H	R	V	C	H	540
541	L	M	L	I	G	D	A	N	F	F	E	P	E	F	F						

**Figura 15.** Secuencia de aminoácidos de HrGal. Con los diferentes colores se muestran las 4 repeticiones en tándem que no son idénticas, identificadas con el servidor Meta-Repeat-Finder [151]. En magenta se muestran los aminoácidos canónicos involucrados en el reconocimiento del azúcar por las galectinas.





**Figura 16.** Alineamiento de las secuencias de las cuatro repeticiones en tándem de HrGal identificadas con el servidor Meta-Repeat-Finder [152]. Los aminoácidos conservados se encuentran en letra color rojo. Los residuos canónicos que interactúan con el ligando, señalados con una estrella color rosa se localizan en cada repetición por lo que constituyen DRC's.

## **8.2. Actividad hemoaglutinante de HrGal nativa**

La actividad de la lectina se verificó mediante ensayos de eritroaglutinación utilizando eritrocitos de conejo formalinizados. Para determinar si dicha actividad era dependiente de calcio, HrGal se pasó por una resina quelante y se determinó su actividad en ausencia y presencia de 1mM de  $\text{CaCl}_2$ . La proteína resultó ser activa sin calcio (1,600 U/mg); sin embargo, incrementó su actividad 4 veces en presencia de dicho ion (6,600 U/mg). Se sabe que en algunos casos la interacción del  $\text{Ca}^{2+}$  con las proteínas produce cambios conformacionales que modulan su actividad. Por otro lado, en algunas enzimas extracelulares, la interacción del  $\text{Ca}^{2+}$  en ciertas asas muy flexibles favorece la estabilidad de la proteína y puede además proporcionar protección contra la digestión proteolítica [153].

## **8.3. Ensayos de especificidad de HrGal**

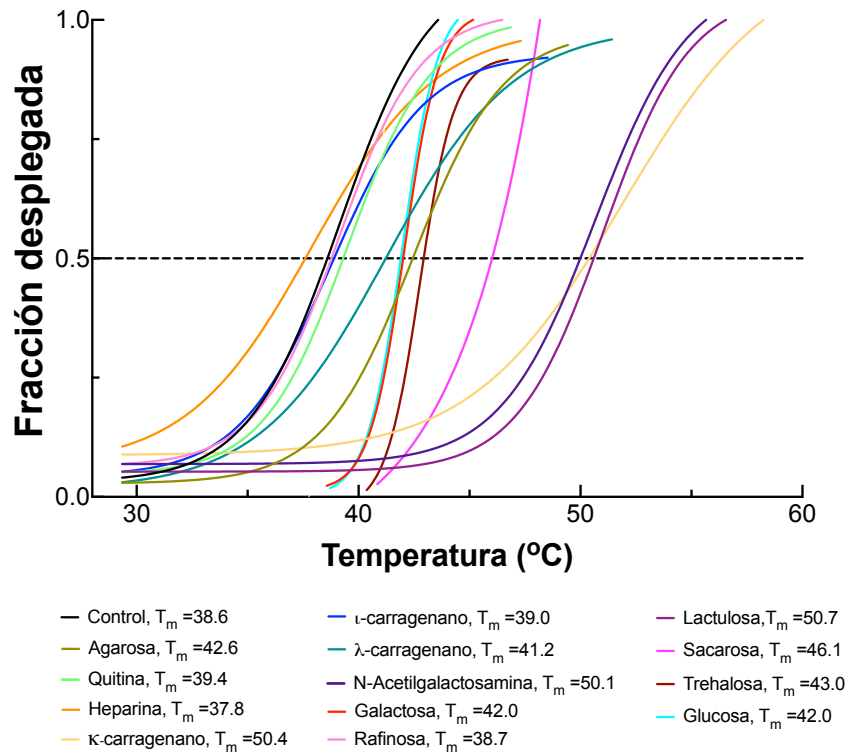
La especificidad de la lectina se determinó utilizando los azúcares que se muestran en la Tabla 2. Aun cuando los ensayos de inhibición de la aglutinación de eritrocitos continúan siendo una técnica muy utilizada para determinar la especificidad de las lectinas, esta suele ser poco sensible cuando se trata de lectinas con repeticiones en tándem, ya que estas tienen varios dominios de reconocimiento (DRC's) con diferente especificidad lo que les permite reconocer a un azúcar sólo con uno o dos de los DRC y para inhibir la aglutinación de eritrocitos es necesario que todos los sitios estén ocupados, por lo que aparentemente no reconoce un azúcar determinado. Por lo anterior y debido a que HrGal tenía 4 DRC's no idénticos decidimos utilizar también la fluorimetría diferencial de barrido. Con esta técnica observamos que algunos glucósidos como la sacarosa y la trehalosa fueron reconocidos por la galectina, lo que se puede observar en la Figura 17. De los azúcares analizados, trece de ellos fueron reconocidos por la HrGal, lo que se puede apreciar por su incremento en la  $T_m$ , pero sólo los azúcares sulfatados heparina y el  $\iota$ -carragenano lograron inhibir por completo su actividad hemaglutinante, con una concentración mínima inhibitoria de 4 mM para la heparina y de  $6.25 \times 10^{-5}$  % para el  $\iota$ -carragenano (Tabla 3).

**Tabla 3.** Ensayos de especificidad de HrGal utilizando diversos azúcares

<b>Azúcar</b>	<b>CMI</b>	<b>Azúcar</b>	<b>CMI</b>
Agar	N.I.	Manitol	N.I.
Agarosa	N.I. ✓	Melibiosa	N.I.
Ácido algínico	N.I.	N-acetil-D-glucosamina	N.I.
Amilopectina	N.I.	N-acetil-D-galactosamina	N.I.
Carboximetil celulosa	N.I.	α-L-ramnosa	N.I.
Quitina	N.I. ✓	α-Lactosa	N.I.
Quitosán	N.I.	β-Lactosa	N.I.
Fucoidán	N.I.	β-ciclodextrina	N.I.
D(-) Arabitol	N.I.	κ-Carragenano	N.I. ✓
D(-) Arabinosa	N.I.	ι-Carragenano	6.25 X 10 <sup>-5</sup> %
D(+) Celobiosa	N.I.	λ-Carragenano	N.I. ✓
D(+) Fucosa	N.I.	Pectina	N.I.
D(-) Fructosa	N.I.	Ácido poligalacturónico	N.I.
D(+) Galactosa	N.I. ✓	Rafinosa	N.I. ✓
D(+) Ácido galacturónico	N.I. ✓	Salicina	N.I.
D(+) Glucosa	N.I. ✓	Almidón	N.I.
D(+) Ácido glucurónico	N.I.	Sacarosa	N.I. ✓
D(+) Maltosa	N.I.	Xilano	N.I.
D(+) Manosa	N.I.		
D(+) Trehalosa	N.I. ✓		
D(+)Xilosa	N.I.		
Heparina	4 mM		
Lactulosa	N.I. ✓		
Laminarina	N.I.		
Manan	N.I.		

N.I., No inhibió la actividad eritroaglutinante; CMI, Concentración mínima de azúcar requerida para inhibir la aglutinación de eritrocitos de conejo por HrGal.

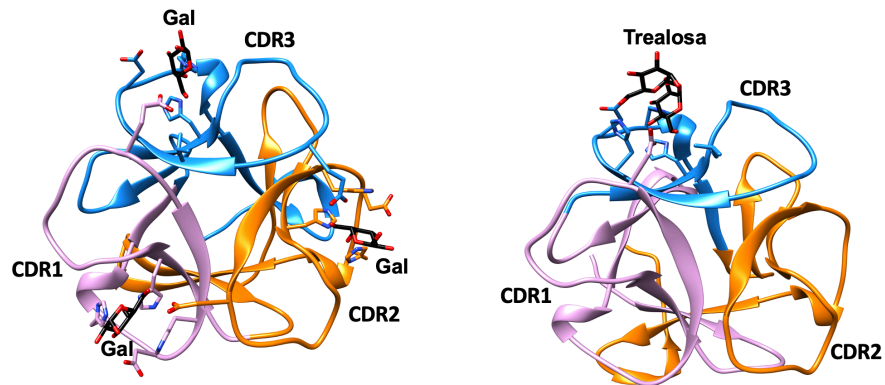
✓, Glicanos que reconoce HrGal pero que no inhiben la aglutinación de eritrocitos. La interacción fue determinada por ensayos de termofluorescencia.



**Figura 17.** Interacción de HrGal con diferentes azúcares. De los azúcares probados en los ensayos de especificidad, trece de ellos fueron reconocidos por HrGal incrementando su  $T_m$ , pero sólo la heparina y el  $\iota$ -carragenano inhibieron la actividad aglutinante.

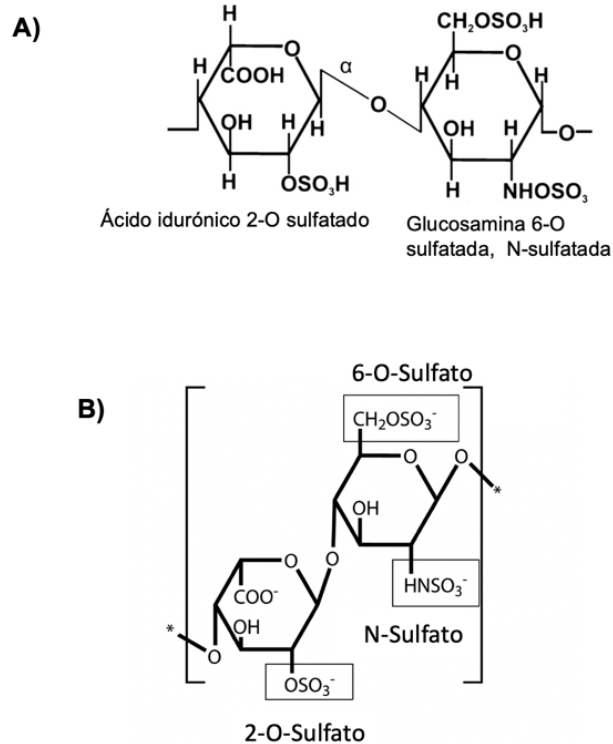
Es importante mencionar que como se observa en la Figura 16, los cuatro dominios de HrGal tienen diferentes secuencias de aminoácidos, y esta característica puede permitir el reconocimiento de diferentes azúcares en cada uno de ellos. En nuestro grupo hemos observado este comportamiento en otras lectinas que tienen repeticiones no idénticas en tándem como los de HrGal; tal es el caso de la lectina MCL del molusco marino *M. californianus*. Esta lectina tiene tres DRC's en tándem, mostró especificidad por galactósidos y no se inhibió su actividad hemaglutinante en presencia de glucósidos sin embargo, algunos dominios fueron capaces de reconocer trehalosa (disacárido compuesto por dos glucosas) (Figura 18) y debido a que no se saturaban todos los sitios, no se inhibía la eritroaglutinación [61]. Por lo

anterior es necesario utilizar las dos técnicas (inhibición de la eritroaglutinación y termofluorescencia) para analizar la especificidad de este tipo de lectinas ya que son complementarias. Es interesante notar que disacáridos como la trehalosa y la sacarosa frecuentemente son utilizados como agentes estabilizadores; sin embargo, el mecanismo de cómo logran este efecto ha sido tema de constante investigación de diferentes aspectos tanto estructurales como dinámicos [154], en este caso en particular es porque están estabilizando sitios de reconocimiento.



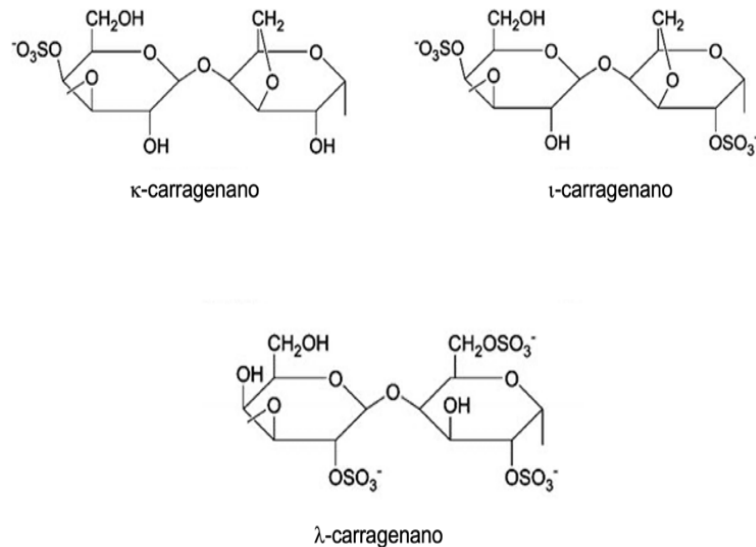
**Figura 18.** Lectina de *Mytilus californianus*. Esta lectina tiene tres repeticiones en tándem que adquieren un plegamiento tipo trebol  $\beta$ , donde cada hoja del trébol corresponde a un DRC. Esta lectina tiene especificidad por galactósidos, sin embargo, en ausencia de estos puede reconocer otros glicanos con menos afinidad. En este caso el CDR 3 fue capaz de reconocer a la trehalosa, pero al unirse sólo a un sitio de reconocimiento no inhibe la eritroaglutinación.

De los polisacáridos que inhibieron la eritroaglutinación, la heparina pertenece a la familia de los glicosaminoglicanos sulfatados y es ampliamente utilizada como anticoagulante. Esta presenta pesos moleculares que varían entre 5-40 kDa. Su estructura consiste en unidades repetitivas de disacáridos compuestos por residuos de ácido urónico (L-idurónico (IdoA) o ácido D-glucurónico (GlcA)) y N-acetil-D-glucosamina (Figura 19A) [155,156]. La posición de los grupos sulfato se muestran en la Figura 19B.



**Figura 19.** Estructura de la heparina. A) Disacárido más abundante en la estructura de la heparina B) Posición de los grupos sulfato.

Por otro lado, los carragenanos son polisacáridos sulfatados lineales de galactosa y 3-6-anhidro-D-galactosa (Figura 20). Los tres tipos de carragenanos más importantes son el kappa ( $\kappa$ ), iota ( $\iota$ ) y lambda ( $\lambda$ ) que difieren en el grado de sulfatación, siendo el primero el menos sulfatado y el último el más sulfatado [157]. Resultó particularmente interesante observar que los carragenanos  $\kappa$  y  $\lambda$  no mostraron efectos inhibitorios en la actividad aglutinante de HrGal, sin embargo, analizando los ensayos de termofluorescencia (Figura 17) observamos que HrGal sí reconoce a los tres tipos de carragenano, siendo el  $\kappa$  el que incrementó más la  $T_m$  (11.8 °C) contra 2.1 °C del  $\lambda$  y 0.4 °C del  $\iota$ . Este último fue el único que inhibió la eritroaglutinación, por lo que el  $\kappa$  y  $\lambda$  deben unirse sólo a algunos de los DRC dejando al menos dos DRC's que permitan la eritroaglutinación, mientras que el  $\iota$  no deja DRC's libres.



**Figura 20.** Estructura del carragenano κ, ι y λ. El polisacárido está formado por galactosa y 3-6-anhidro-D-galactosasa y difieren en el grado de sulfatación.

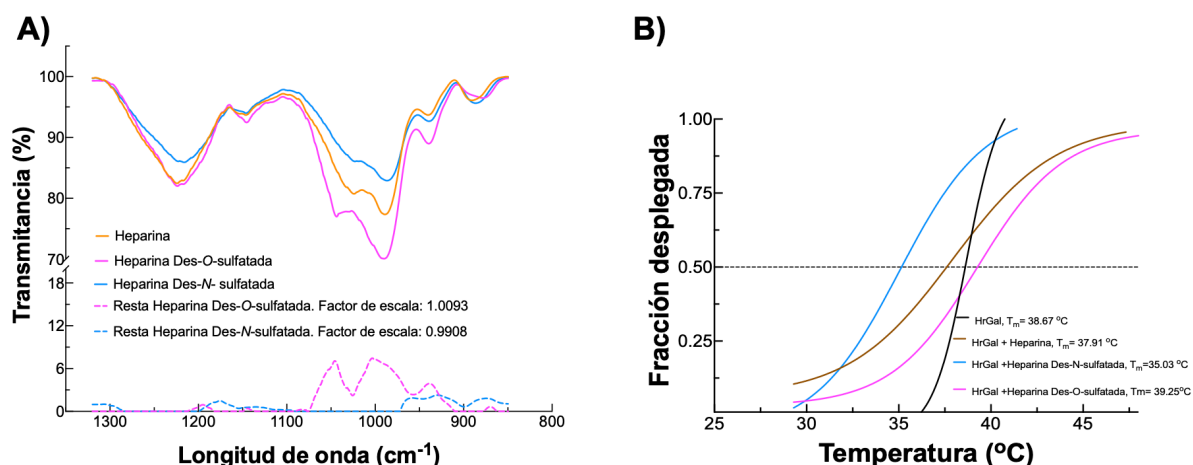
Los polisacáridos sulfatados tienen una gran variedad de actividades biológicas que dependen de la localización de los grupos sulfato [158]. Si bien se sabe que las lectinas que reconocen glicosaminoglucanos (GAG) más estudiadas son las selectinas, que interactúan con los GAG de forma dependiente de cationes divalentes [159], se han reportado algunas otras lectinas que reconocen azúcares sulfatados y tienen actividad inmunomoduladora [160,161]. De los pocos reportes de galectinas con especificidad por glicanos sulfatados, se pueden mencionar a las galectinas 1, 3 y 8 de humano, cuya importancia como agentes terapéuticos contra el cáncer se ha determinado en diversos estudios [162–164]. En la galectina 8 el reconocimiento por azúcares sulfatados se le atribuye principalmente al extremo N-terminal del DRC, a los aminoácidos Arg<sup>45</sup>, Gln<sup>47</sup> y Arg<sup>59</sup> y a la región del asa extendida entre la hebra S3 y la S4 que interactúan directa o indirectamente con el sulfato (Figura 8) contribuyendo a la fuerte unión de la galectina con los oligosacáridos sulfatados [165,166].

### 8.3.1. Evaluación del efecto de la presencia de los grupos sulfato en la actividad de HrGal

Con el fin de determinar si los grupos sulfato eran necesarios para el reconocimiento del ligando por parte de HrGal se decidió desulfatar a la heparina mediante dos procedimientos diferentes: Se des-O sulfató por un método de liofilización y se des-N sulfató mediante solvólisis. En ambos casos, el ATR-FTIR confirmó un desplazamiento de las frecuencias de los picos más intensos y una modificación de los baricentros de la banda, como se esperaba según Cabassi et al. [167] (Figura 21A). Para el análisis de las diferencias entre la heparina sulfatada y la desulfatada nos concentramos sólo en las señales que corresponden a las vibraciones del sulfato, que son las que se muestran en la Tabla 4. En la Figura 21B se observa reconocimiento de HrGal por la heparina completa, la Des-O-sulfatada y la Des-N-sulfatada. Con la heparina completa hay una ligera disminución en la  $T_m$  (0.76 °C) sin embargo, lo más sobresaliente es el cambio en la pendiente de la sigmoide con respecto al control (HrGal sin heparina) que mostró una pendiente de 0.9 °C y una  $T_m$  de 38.67 °C mientras que la lectina en presencia de heparina mostró una pendiente más suave (2.81 °C) con una  $T_m$  de 37.91°C. En otras lectinas hemos observado que una pendiente más prolongada está relacionada con oligomerización de la proteína inducida por el reconocimiento del ligando, sin embargo, dicha interacción generalmente está acompañada por un incremento en el valor de la  $T_m$ . Un efecto favorable en cuanto al incremento de la  $T_m$  se observó con la heparina Des-O-sulfatada que incrementó ligeramente el valor de la  $T_m$  (en 0.5 °C) por lo que prefiere a la heparina sulfatada en N. Esto también se observó con el decremento de la  $T_m$  con la heparina Des-N-sulfatada cuya  $T_m$  disminuyó 3.6 °C. La mayoría de los ligandos al interaccionar con la proteína la estabilizan, provocando un aumento de la temperatura de fusión de dicha proteína. Sin embargo, algunos ligandos las desestabilizan uniéndose principalmente al estado desplegado disminuyendo la temperatura de fusión, compitiendo con un cofactor endógeno (estabilizador) o a través de efectos potencialmente inespecíficos. Los ligandos que estabilizan las proteínas pueden denominarse ligandos N, y los ligandos que desestabilizan las proteínas pueden denominarse ligandos U [168,169]. Sin embargo para entender que



está pasando cuando HrGal interacciona con la heparina son necesarios más estudios y sobre todo estudios estructurales que permitan visualizar cómo es el reconocimiento ya que se trata de cuatro dominios cada uno de ellos separados por asas formadas por 6 residuos (Asa DRC1-DRC2, N<sup>142</sup>, S<sup>143</sup>, M<sup>144</sup>, T<sup>145</sup>, N<sup>146</sup>, P<sup>147</sup>; Asa DRC2-DRC3, E<sup>278</sup>, P<sup>279</sup>, L<sup>280</sup>, P<sup>281</sup>, D<sup>282</sup>, D<sup>283</sup>; Asa DRC3-DRC4, R<sup>416</sup>, R<sup>417</sup>, L<sup>418</sup>, E<sup>419</sup>, L<sup>420</sup>, P<sup>421</sup>) que son muy flexibles y como se trata de un polisacárido no sabemos de qué forma se organizan para reconocerlo. Sin embargo, si se puede concluir que el grado de sulfatación y la posición de dichos grupos sulfato es importante en el reconocimiento. Algo similar se observó con los carragenanos  $\kappa$ ,  $\iota$  y  $\lambda$ , donde el que poseía el menor grado de sulfatación (k) aumentó más la  $T_m$ , pero no inhibió la eritroaglutinación, por lo que no se unió a todos los DRC.



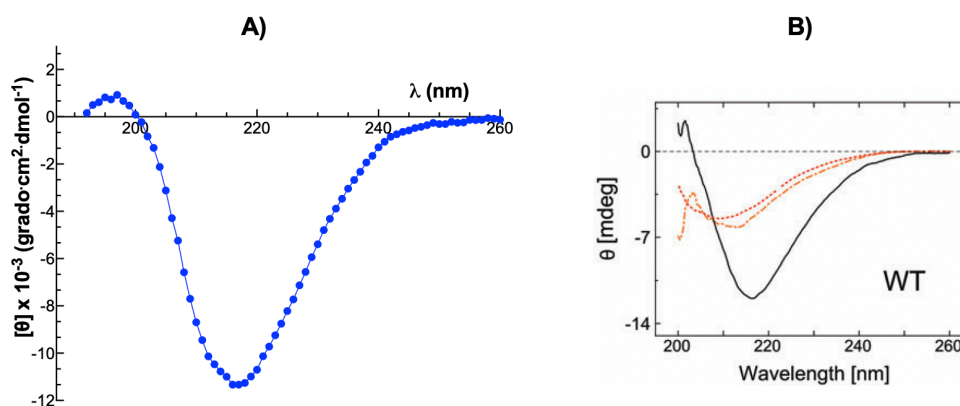
**Figura 21.** Desulfatación de la heparina. A) Espectros ATR-FTIR de la heparina y de la heparina químicamente des-N y des-O-sulfatada en la región que corresponde a las bandas  $\nu_{as}S=O$ . A todos los espectros se les restó la línea base de forma adaptativa, se suavizaron con un filtro Savitsky-Golay de quinto orden y se normalizaron. Los espectros de las heparinas desulfatadas se escalaron individualmente restándolos del de la heparina. Todos los espectros se obtuvieron a partir de películas secas. B) Se observa la variación en el reconocimiento de la heparina completa y desulfatada en N y en O por la HrGal.

**Tabla 4.** Frecuencias de infrarrojo de las bandas  $\nu_{as}S=O$  de la heparina completa y de la químicamente desulfatada indicando sus baricentros.

Compuesto	Frecuencia de picos ( $cm^{-1}$ )	Baricentro de la banda
Heparina	1,141.9	1,118.8
	1,158.7	
	1,228.6	
Heparina Des-O-sulfatada	1,121.6	1,101.9
	1,143.3	
	1,145.5	
Heparina Des-N-sulfatada	1,098.9	1,131.4
	1,148.6	
	1,211.8	
	1,219	
	1,201.8	
	1,211.9	
1,218.8		

#### 8.4. Determinación de la estructura secundaria de HrGal por Dicroísmo circular

Las mediciones obtenidas con el espectropolarímetro en la región del UV lejano (190-260 nm) reveló que es la hebra- $\beta$  la que predomina en la estructura de HrGal (15.9% antiparalela y 23% paralela) (Figura 22). Este contenido de estructura se ha descrito anteriormente en el plegamiento tipo sandwich- $\beta$  presente en galectinas de diferentes fuentes [46,170].



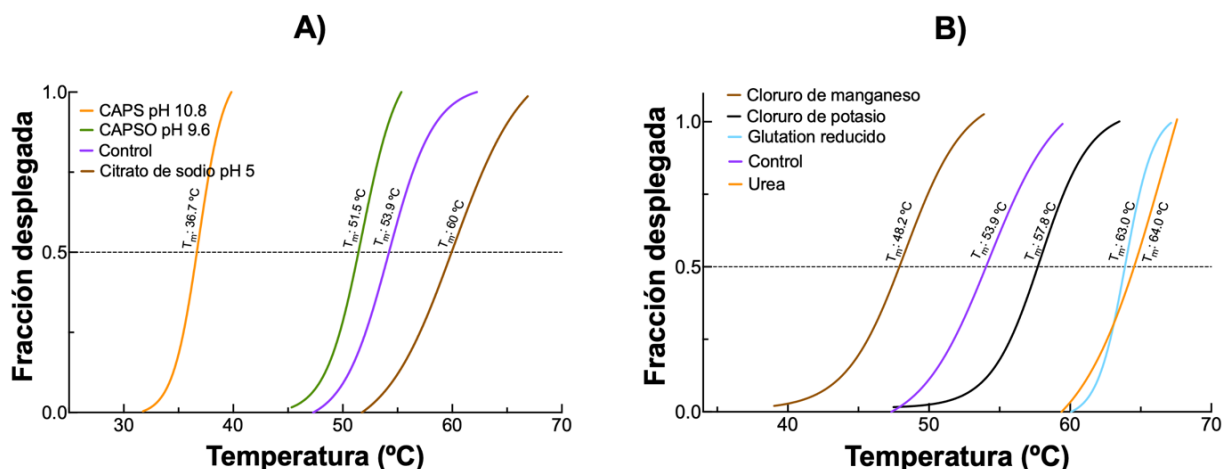
**Figura 22.** Espectro de dicroísmo circular en la región del UV lejano de HrGal (A). En (B) se observa el espectro de la galectina 1 de humano (línea negra continua) que es muy parecido al de HrGal (tomado de Guardia et al, 2014).

### **8.5. Ensayos de estabilidad utilizando Fluorimetría diferencial de barrido**

Es importante mencionar que durante todo el proceso de purificación la proteína se precipitaba rápidamente desde el momento de la elución de la resina de afinidad. Por lo anterior, decidimos realizar ensayos de estabilidad utilizando la técnica de fluorimetría diferencial de barrido con el fin de encontrar las mejores condiciones de solubilidad y estabilidad de HrGal; Para esto se probaron 96 amortiguadores y el kit de aditivos que consta también de 96 condiciones. En la Figura 23A se muestran los amortiguadores en los que la galectina mostró una mayor estabilidad. Solamente con tres de los 96 amortiguadores probados se obtuvieron curvas sigmoidales. En presencia del amortiguador de citratos la  $T_m$  incrementó 6.1°C respecto al control (HrGal en amortiguador de fosfatos, que es la condición en la que se purificó a la proteína) mientras que con CAPS pH 10.8 y CAPSO pH 9.6 la  $T_m$  disminuye hasta 17.2°C, lo que nos indica que en esta condición la proteína se encuentra menos estable. Sin embargo, también puede indicar que la proteína es más soluble, ya que en algunas ocasiones una  $T_m$  mayor también puede indicar oligomerización o agregación, por lo que se tiene que considerar la forma de la sigmoide [135]. De forma interesante, con CAPSO 50 mM pH 9.6, la  $T_m$  disminuyó 2.4°C; pero por la forma de la sigmoide que podría indicar solubilización, se decidió elegirla para probarla en los ensayos de la actividad específica, observando un incremento del doble (13,300 U/mg) en comparación con el control que se encontraba en amortiguador de fosfatos (Tabla 5). En la condición con amortiguador de citratos pH 5, la lectina no mostró actividad, en este caso el aumento en la  $T_m$  debe ser por procesos de agregación ya que estamos muy cerca de su punto isoeléctrico.

La estabilidad térmica de HrGal se vio favorecida no solo por el amortiguador, también se observó que algunos iones, urea y agentes reductores favorecían su estabilidad (Figura 23B). De forma interesante, se observa que la urea estabiliza a la proteína, sin embargo, la forma de la sigmoide sugiere que, en menor medida, al

probar su actividad observamos que, aun en presencia de urea 1M, la lectina logró mantener su actividad (13300 U/mg). Esto puede deberse a que los agentes caotrópicos pueden inducir re-arreglos a las proteínas para obtener una conformación más estable, al evitar interacciones que favorezcan la agregación [135]. Además, se ha observado por ejemplo en la Galectina 1 de humano [171], la presencia de varios puentes disulfuro que mantienen plegada a la lectina aún en presencia de concentraciones altas de urea. Sin embargo, en este ensayo también destacó el efecto del glutatión reducido, que es un agente reductor que favoreció su estabilidad, lo que puede indicar la presencia de cisteínas libres en la estructura de HrGal. Considerando este efecto, decidimos evaluar el efecto de Tris-(2-carboxyetil) fosfina (TCEP), que es otro agente reductor y también de un agente derivatizante de cisteínas (MMTS, S-metilmetanotiosulfanato) que es un compuesto pequeño que puede bloquear de forma reversible las cisteínas en las moléculas que contengan grupos tiol a sulfonilato. El primero incrementó la actividad cuatro veces (26,600 U/mg), mientras que el último aumento la actividad ocho veces (5,300 U/mg). Estos resultados pueden sugerir que si bien varias de las cisteínas presentes en la lectina pueden estar formando puentes disulfuro, algunas de ellas podrían estar libres provocando la inestabilidad de la galectina al interaccionar con otras cisteínas libres, formando puentes disulfuro no nativos, lo que termina agregando a la proteína (Tabla 5). El efecto mencionado, se puede evitar al adicionar algún agente reductor. A partir de este estudio determinamos que para mantener estable a la lectina era necesario mantenerla en CAPSO 50 mM pH 9 y DTT 1mM.



**Figura 23.** Estudios de estabilidad de HrGal mediante ensayos de termofluorescencia. (A) Efecto del pH y sales amortiguadoras en la estabilidad de la HrGal. Se utilizaron 96 amortiguadores y se muestran sólo las que mostraron efectos positivos en la estabilidad (B) Efecto de diferentes aditivos, se utilizó el kit de Hampton Research que incluye 96 aditivos.

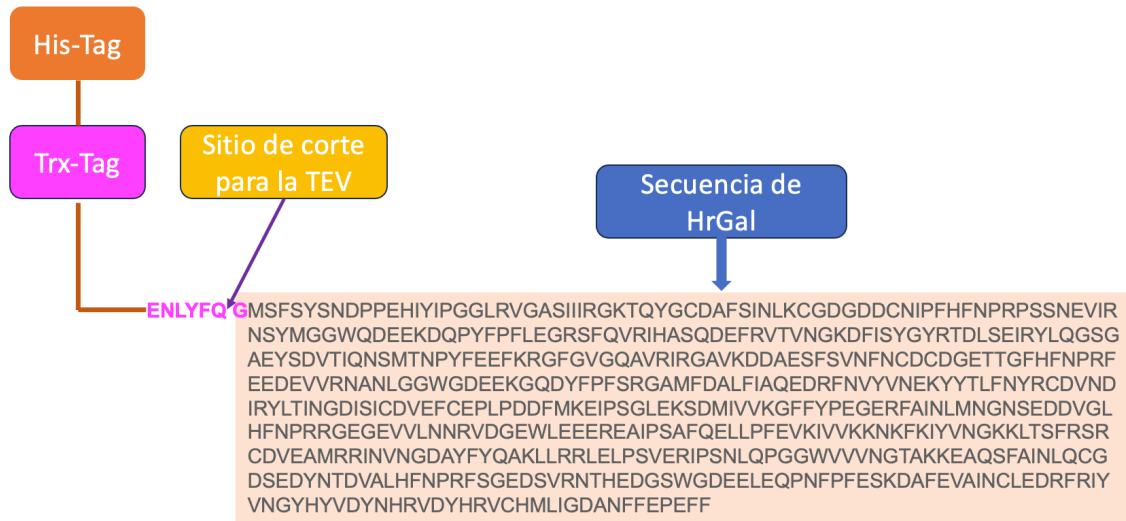
**Tabla 5.** Ensayo de eritroaglutinación de la HrGal con distintos aditivos.

Aditivo	Actividad* con $\text{CaCl}_2$ (1mM)
HrGal en Fosfatos 50 mM pH 7.4 (control)	6666
Urea 1 M	13330
TCEP 1mM	26600
MMTS 1mM	53300
CAPSO 50 mM pH 9	13300

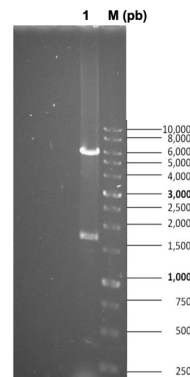
\*La actividad se expresó como el recíproco de la dilución más alta que mostró aglutinación.

### 8.6. Clonación y expresión de la galectina recombinante rHrGal

Para la obtención de la galectina de *Haliotis rufescens* de forma recombinante (rHrGal) se consideró la construcción que se muestra en la Figura 24.



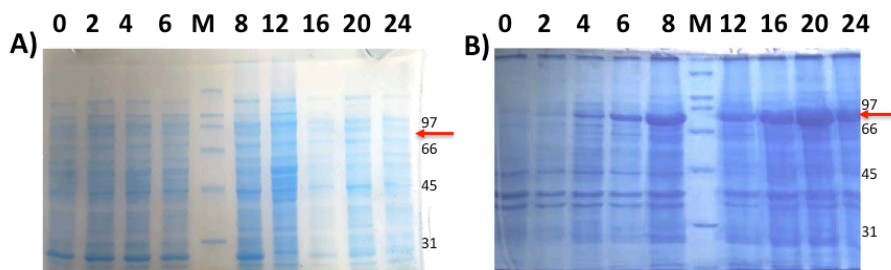
**Figura 24.** Construcción y diseño del plásmido de expresión. El producto de expresión tiene una masa de 81.5 kDa incluyendo el sitio de corte y las etiquetas de fusión. Tras el corte con la proteasa TEV la masa de la lectina digerida a la que llamamos rHrGal es de 63.918 kDa



**Figura 25.** Ensayo de restricción. Digestión con enzimas de restricción del ADN plasmídico con *KpnI* y *EcoRI*. Electroforesis en gel de agarosa al 2% teñido con bromuro de etidio.

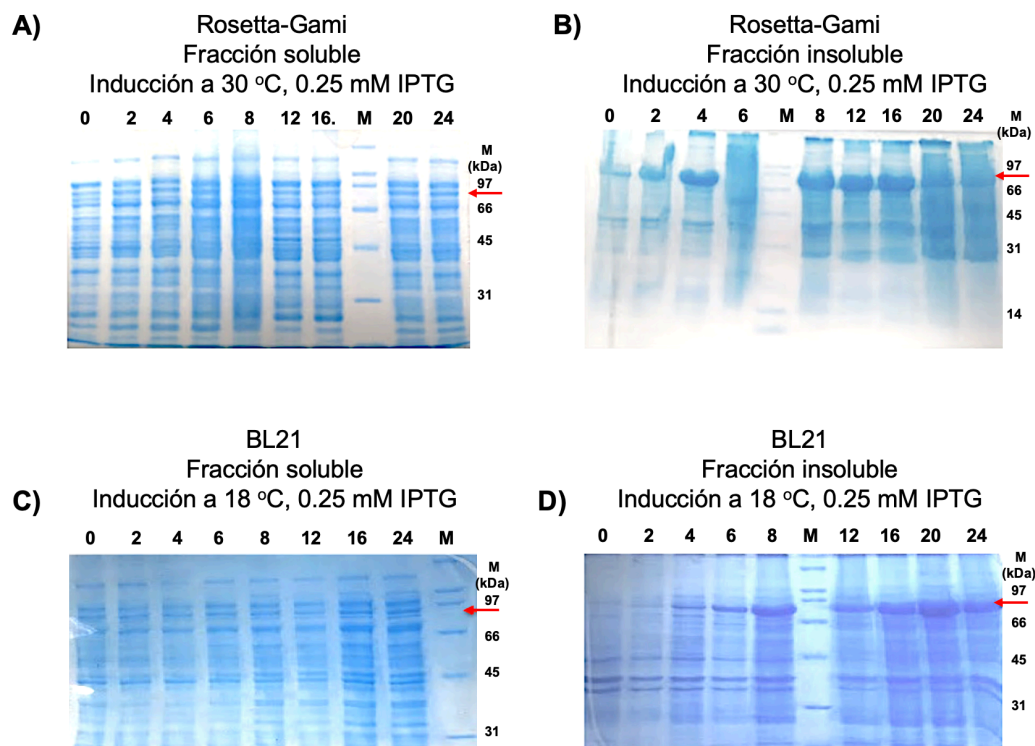
Para confirmar el inserto del gen de rHrGal, el plásmido recombinante se digirió con las enzimas de restricción *KpnI* y *EcoRI*. La muestra se corrió en un gel de agarosa

donde se observaron 2 bandas, que corresponden al inserto de 1,701 pb y el vector (pET-32a) de 5.9 kb (Figura 25). Posteriormente, se continuó con la cinética de expresión heteróloga de rHrGal en la cepa BL21, y se observó que la proteína se encontraba en la fracción insoluble (Figura 26 A y B).



**Figura 26.** Cinética de expresión de rHrGal en la cepa BL21 a 37 °C con 0.5 mM de IPTG. A) Fracción soluble B) Fracción insoluble. La electroforesis se realizó en un gel de poliacrilamida al 10%. Se observa el control a las 0 horas (sin inducción), en los tiempos 2, 4, 6, 8, 12, 16, 20 y 24 horas (inducido). M corresponde con el marcador de masa molecular. Con la flecha roja se muestra la fracción con la masa esperada de aproximadamente 81.5 kDa que corresponde a la rHrGal con la etiqueta de tiorredoxina.

Con el fin de obtener a la galectina en la fracción soluble se probaron diferentes condiciones de temperatura, concentración de IPTG, tiempo de incubación y diferentes cepas [172], sin embargo la proteína siempre se encontraba en la fracción insoluble por lo que se decidió trabajar con la fracción insoluble. Las dos cepas que mostraron ser más prometedoras fueron la Rosseta-Gami y BL21 (Figura 27). Finalmente, se decidió utilizar BL21 a 18°C con 0.25 mM de IPTG con 20 horas post-inducción para la expresión de rHrGal y su posterior purificación.



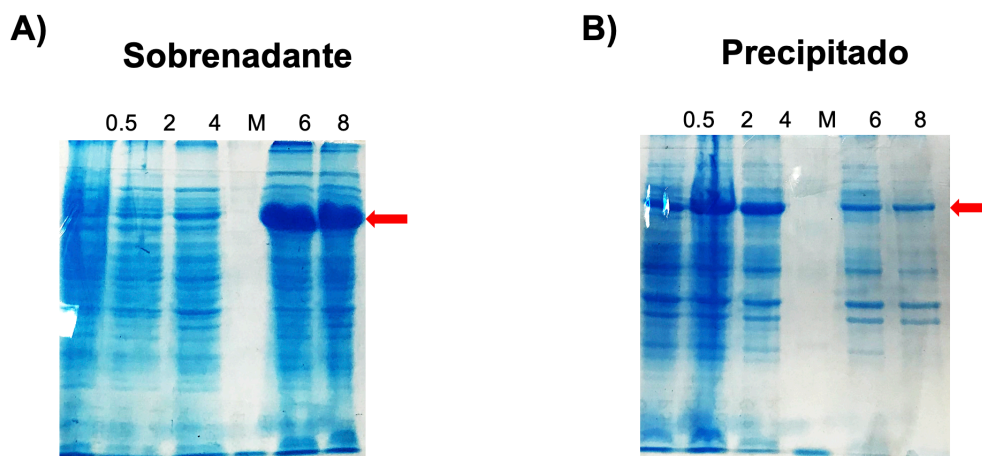
**Figura 27.** Cinética de expresión de rHrGal. Panel A y B, resultados obtenidos con la cepa *Rosetta-Gami* (A, fracción soluble y B, fracción insoluble). En el panel C y D, resultados obtenidos con la cepa BL21 (C, fracción soluble y D, fracción insoluble). Con ambas cepas se muestran el tiempo 0 (sin inducción) y en los tiempos 2, 4, 6, 8, 12, 16, 20, 24 horas (inducido). La letra M corresponde al marcador de masa molecular. Con la flecha roja se muestra la banda con una masa aproximada de 81.5 kDa que corresponde a rHrGal más la etiqueta de tiorredoxina.

### 8.7. Purificación de la rHrGal

Con la fracción insoluble se continuó con la purificación de la galectina. Primero determinamos cual era la concentración mínima de urea necesaria para solubilizar a la proteína. Para este ensayo se evaluaron cinco concentraciones (0.5, 2, 4, 6 y 8 M) y después de 24 h de incubación, se cargó el sobrenadante (Figura 28 A) y el precipitado (Figura 28 B) en geles de poliacrilamida al 10% y se observó que la concentración mínima que se necesitaba era 6M de urea (TBS-Urea 6M). Después, la proteína se purificó en condiciones desnaturizantes de urea 6 M en una columna

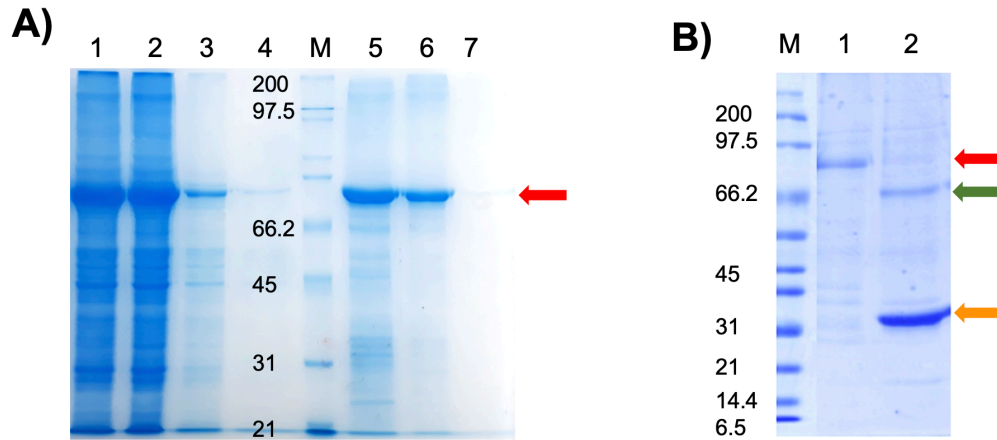


de afinidad de  $\text{Ni}^{2+}$ . Para eluir a la galectina se utilizó un gradiente de imidazol (20, 150, 250 y 500 mM) donde observamos que rHrGal+trx eluye desde 150 mM de imidazol. (Figura 29A). También observamos que la resina se saturó ya que la lectina también se encuentra en lo que no se unió a la lectina por lo que esta se volvió a aplicar a la resina para recuperarla toda.

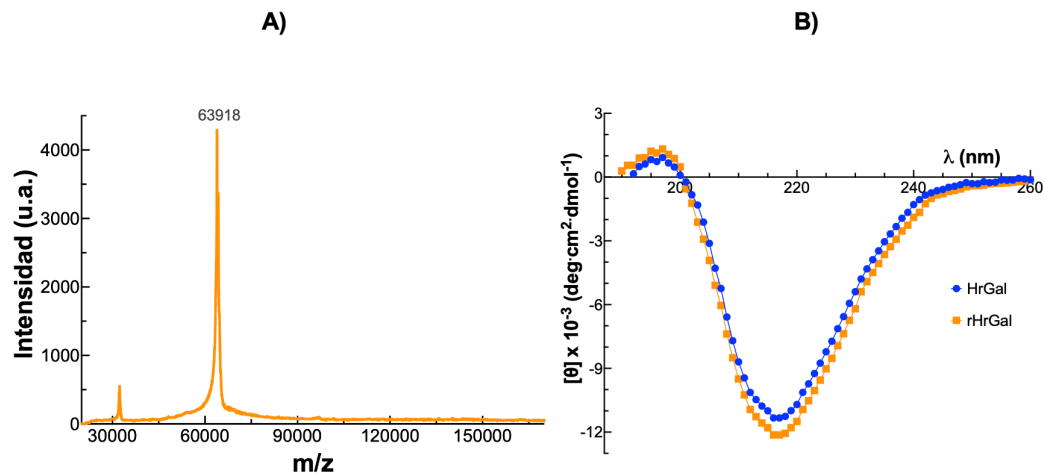


**Figura 28.** Purificación de rHrGal. **A)** Concentración óptima de urea para solubilizar a rHrGal+Trx, los carriles 1-5 corresponden con el sobrenadante de la galectina después de incubarse con 0.5, 2, 4, 6, 8 M de urea respectivamente. **B)** Precipitado de rHrGal+Trx después de incubarse con 0.5, 2, 4, 6, 8 M de urea respectivamente.

Las fracciones eluidas con 150 y 250 mM de imidazol se unieron y replegaron mediante diálisis realizando un gradiente inverso de urea de 6, 4, 2, y 0 M con DTT 10 mM y glicerol al 10% para su corte con la enzima TEV, que elimina las histidinas y la tiorredoxina (Figura 29B). La fracción 5 que se observa en la Figura 29B se recirculó en la resina de níquel donde la tiorredoxina y la TEV se unieron, mientras que la galectina no se unió. De esta manera se obtuvo a la a rHrGal pura como se demostró por espectrometría de masas MALDI-TOF (Figura 30A) donde se observó la masa de 63918 Da. En la Figura 30B se observa que la lectina recombinante presenta el mismo plegamiento que la proteína endógena, con un predominio de hebras  $\beta$  en su estructura.



**Figura 29.** A) Purificación de rHrGal+Trx, los diferentes carriles corresponden con **1**, fracción cruda, **2**, fracción que no se unió a la resina de níquel, **3**, lavado sin imidazol, **4**, **5**, **6** y **7** corresponden con lo que eluye con 20, 150, 250 y 500 mM de imidazol respectivamente. **B)** Corte con TEV para eliminar la etiqueta de fusión, los carriles corresponden con **M**, Marcador de masa molecular, **1**, rHrGal+Trx (antes del corte con TEV), **2**, Fracción obtenida de rHrGal después de la incubación con TEV y antes de pasar por la columna de níquel. Con la flecha roja se observa lo que no se cortó con la TEV (rHrGal+trx), Con la flecha verde se muestra a la galectina escindida (~66 kDa) y con la flecha naranja se muestra a la TEV (30.1) kDa.

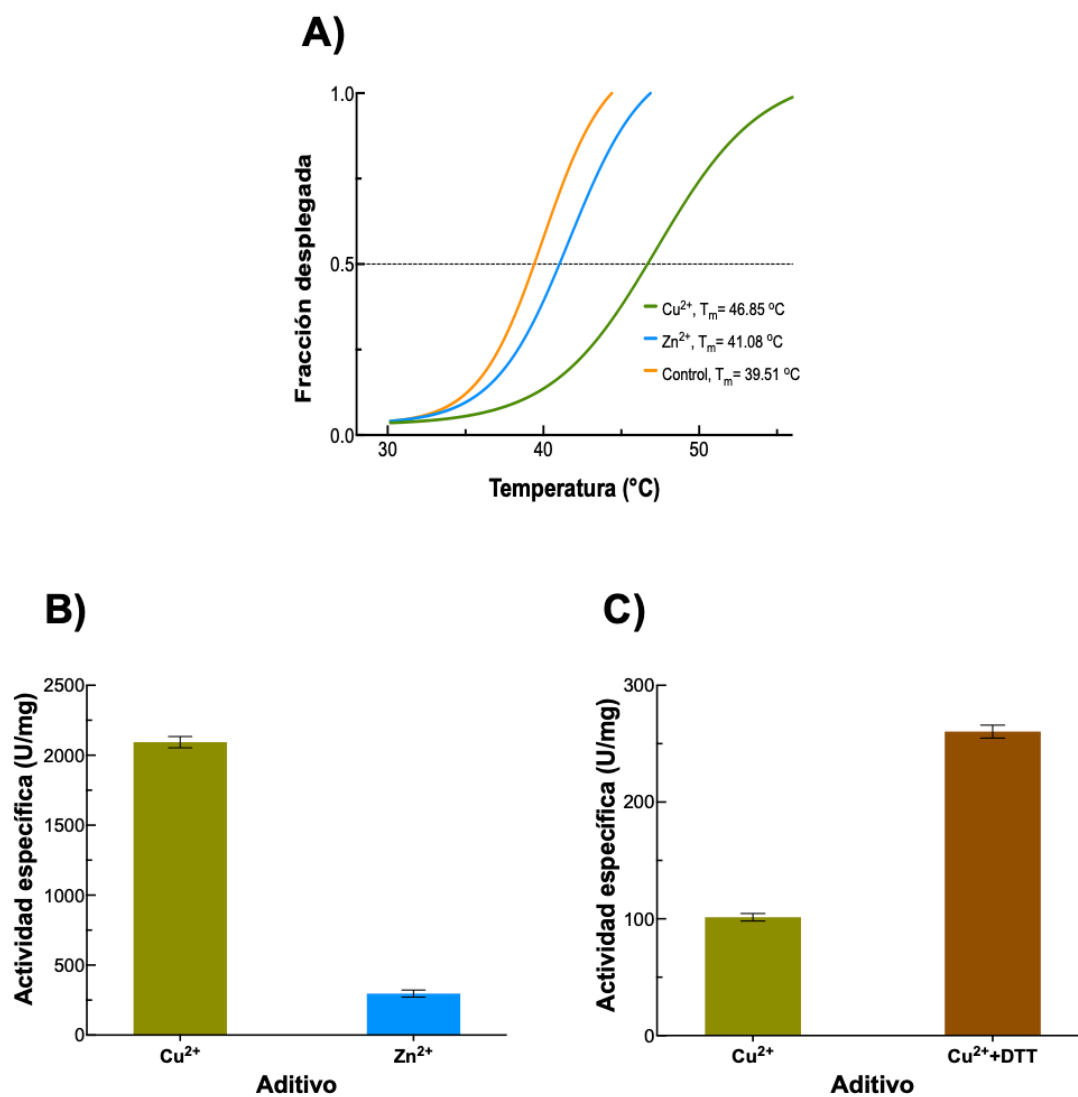


**Figura 30.** Caracterización de la galectina recombinante. A) Espectro de masas MALDI-TOF donde se observa la masa de 63.9 kDa. La señal alrededor de 30 kDa corresponde al ion  $m/z$  B) Comparación del espectro de dicroísmo circular en el UV lejano de la proteína recombinante comparada con la obtenida de la fuente natural.

### 8.8. *Aglutinación de eritrocitos HrGal*

Fue interesante notar que aún en presencia de TBS-urea 6M e imidazol 500 mM la galectina mostraba actividad (1,550 U/mg). Sin embargo, al eliminar la urea, la galectina perdía su actividad. Debido a que rHrGal no era activa después del replegamiento, fue necesario encontrar una condición óptima para que pudiera mantener su actividad, por lo que probamos distintos iones y aditivos (DTT, imidazol,  $\text{Ca}^{+2}$ ,  $\text{Co}^{+2}$ ,  $\text{Mn}^{+2}$ ,  $\text{Mg}^{+2}$ ,  $\text{Cu}^{+2}$  y  $\text{Zn}^{+2}$ ). De los ensayos de termofluorescencia observamos que el  $\text{Cu}^{+2}$  y el  $\text{Zn}^{+2}$  incrementaban la  $T_m$  de rHrGal (Figura 31A), por lo que decidimos adicionarlos inmediatamente después de eluir a rHrGal de la resina de níquel-sefarosa con la proteína desplegada a una concentración de 2 mM. Observamos que la lectina incrementó su actividad a 2,100 U/mg en presencia de cobre (Figura 31B). Sin embargo, cuando estos iones se adicionaban con la proteína

plegada esta seguía inactiva. Estos resultados sugieren que los iones pueden ser esenciales ya que podrían estabilizar la hendidura donde ocurre el reconocimiento del ligando al interactuar con las asas que unen las hebras S3 y S4. Anteriormente, se ha reportado que la galectina 4 podría estar siendo estabilizada por calcio en el asa S4-S5 aumentando su afinidad por el ligando. Esta región se ha descrito como crucial en el reconocimiento de azúcares sulfatados, como se observó en galectinas 8 y 4 [157,173]. A pesar de que las galectinas son una familia de lectinas que se distingue por no depender de iones para su actividad, los iones con los que se presentó actividad hemoaglutinante son componentes traza naturales del medio marino, y se ha demostrado que los moluscos acumulan principalmente  $\text{Cu}^{+2}$  y  $\text{Zn}^{+2}$  en sus tejidos blandos, ya sea músculo o vísceras, por lo que HrGal pudiera estar interactuando con estos iones para estabilizar su plegamiento [174,175]. Si bien observamos que al replegar a rHrGal en presencia de iones  $\text{Cu}^{+2}$  y DTT permanece activa (Figura 31C), dicha actividad se perdía en periodos cortos de tiempo por lo que seguimos probando diferentes condiciones y finalmente encontramos que la condición óptima para realizar el replegamiento de la proteína era por goteo de la lectina pura en el amortiguador de replegamiento (Tris-HCl 50 mM pH 7.4, NaCl 1.5 M,  $\beta$ -mercaptoetanol 1 mM, sulfato de cobre 2 mM, glicerol 5%, 150 mM de galactosa) con agitación suave a 4 °C. La dilución de la proteína purificada en este amortiguador se realizó en una relación 1:10 considerando que al final la proteína debería quedar diluída a 0.5 mg/mL. El replegamiento de la proteína se completó mediante diálisis hasta eliminar totalmente la urea y la galactosa (al ser su ligando inhibe la actividad eritroaglutinante) y disminuyendo gradualmente la concentración de NaCl hasta 150 mM. Un paso vital durante el replegamiento de la lectina fue la adición de galactosa que al ser su ligando favoreció el plegamiento adecuado de la lectina manteniendo la integridad del sitio de reconocimiento a carbohidratos. Con esta técnica de replegamiento recuperamos por completo a la lectina de los cuerpos de inclusión y logramos mantenerla soluble y activa en amortiguador Tris-HCl 50 mM pH 7.4, NaCl 150 mM, DTT 1 mM.



**Figura 31.** Ensayos de estabilización de rHrGal. A) Efecto de la adición de cobre y zinc en la estabilidad de rHrGal. B) Actividad de rHrGal desplegada en presencia de 2 mM de Cu<sup>+2</sup> y Zn<sup>+2</sup>. C) Actividad de rHrGal replegada con 2 mM de cobre y 1 mM de DTT. Se realizaron tres repeticiones por experimento y se muestra la desviación estándar.

### 8.9. Análisis estructural de HrGal

En la Figura 32A se observa el modelo estructural de HrGal obtenido con el servidor AlphaFold [142]. La cadena polipeptídica está organizada en cuatro dominios que corresponden con las cuatro repeticiones en secuencia identificadas con el servidor Meta-Repeat-Finder [152] (Figuras 15 y 16). Al poseer cada dominio los aminoácidos canónicos para el reconocimiento del azúcar constituyen dominios de reconocimiento a carbohidratos o DRC's.

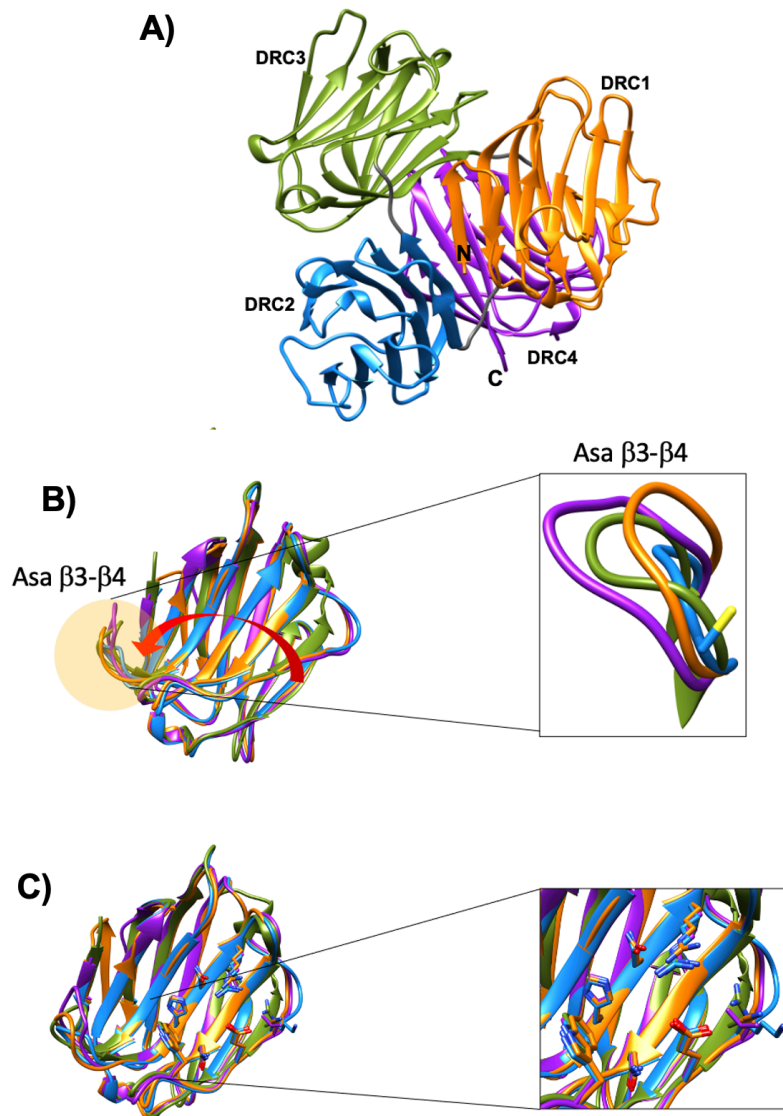
En la Figura 32B se muestra la superposición de los cuatro dominios donde se aprecia una gran similitud estructural entre ellos con RMSD entre 0.72 y 0.87 Å. Cada DRC presenta el plegamiento de sándwich  $\beta$  típico de las galectinas. La mayor diferencia en los DRC's se encuentra en el asa que conecta las hebras  $\beta$ 3- $\beta$ 4 que se considera de gran importancia para el reconocimiento de los ligandos, esta tiene diferente longitud en cada dominio. En el DRC1 consta de 10 residuos (Cys<sup>44</sup>-Pro<sup>53</sup>), en el DRC2 de 6 residuos (Cys<sup>181</sup>-Glu<sup>186</sup>), en el DRC3 de 6 residuos (Gly<sup>318</sup>-Asp<sup>323</sup>) y en el DRC4 de 10 residuos (Cys<sup>455</sup>-Asp<sup>464</sup>) (Figura 32B acercamiento). Las asas  $\beta$ 3- $\beta$ 4 de los diferentes dominios están estabilizadas por un puente disulfuro con excepción de la del DRC3 (Figura 32B, color verde). Las asas más grandes están en los DRC1 y DRC4. Estas diferencias en longitud pueden provocar diferencias en el reconocimiento de azúcares en cada dominio.

En el modelo estructural también se observó que la galectina posee 3 puentes disulfuro y 4 cisteínas libre (Tabla 6). Es probable que la presencia de los puentes disulfuro que se encuentran estabilizando el asa que conecta las hebras  $\beta$ 3- $\beta$ 4 sean los que impidan que la proteína se despliegue por completo y mantenga su actividad aún con 6M de urea. Además, la presencia de las 4 cisteínas libres puede ser responsable de la pérdida de actividad cuando se elimina la urea, esto debido a que al pH en el que se mantenía la lectina (pH 7.4) estas cisteínas pueden formar puentes disulfuro intermoleculares provocando agregación, de ahí que se requiera la presencia de un agente reductor (ver sección 8.3 y Figura 23B). Al realizar un análisis del área accesible al disolvente por residuo utilizando el programa Areaimol presente en el servidor CCP4 [176], observamos que las cisteínas libres 35, 272 y 393 estaban ocultas por lo que no eran las responsables de la inestabilidad, sin

embargo la Cys<sup>183</sup> tiene un área accesible al disolvente de 81 Å<sup>2</sup> por lo que está muy expuesta y esta debe ser la que esté provocando la pérdida de actividad y la necesidad de agentes reductores (Tabla 5). Además, esta cisteína está localizada en una zona de vital importancia para la función de la lectina que es el asa que conecta las hebras β3-β4 (Figura 32B en azul).

En la Figura 32C se observan los aminoácidos canónicos conservados en todas las galectinas y que son responsables del reconocimiento de β-galactósidos.

Un aspecto interesante de la estructura de HrGal fue la presencia de un asa y una hebra β adicional en el DRC1 que no están presentes en los otros DRC (Figura 33). Algo similar se observó en el DRC del extremo amino de la galectina 8 de humano, que sólo cuenta con dos DRC los cuales comparten el 38% de identidad a nivel de aminoácidos y tiene una alta afinidad por sialil 3'-O-lactosa [165]. Esta especificidad fue principalmente atribuida al DRC N-terminal, en donde la Gln<sup>47</sup> juega un papel importante en el reconocimiento de azúcares sulfatados [166].



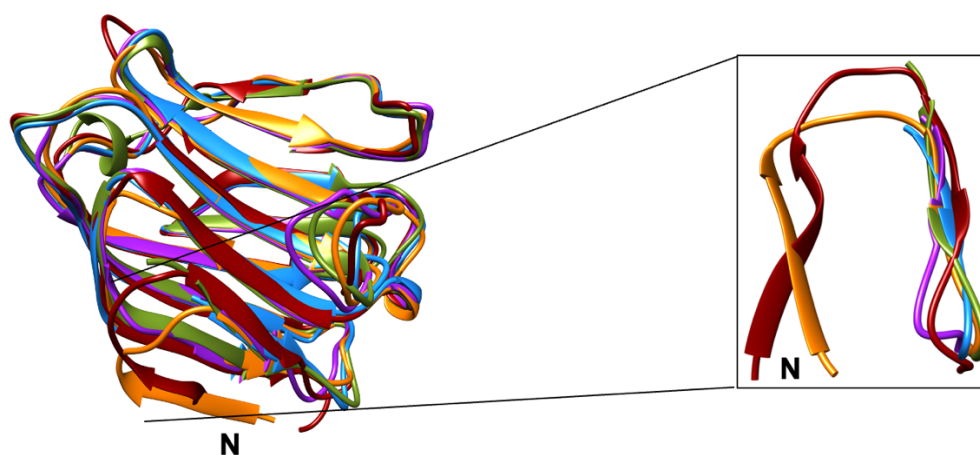
**Figura 32.** Modelo teórico de HrGal obtenido con el servidor AlphaFold. A) Diagrama de listones que muestra los cuatro dominios en tándem. B) Superposición de los dominios de HrGal, donde se observa el plegamiento de sándwich- $\beta$  típico de las galectinas. Los DRC1-4 se observan en color naranja, azul, verde y morado respectivamente. La flecha roja señala la hendidura donde se acomoda el ligando y el círculo amarillo el asa  $\beta 3-\beta 4$  que se considera de vital importancia para el reconocimiento. Un acercamiento de esta asa muestra las diferencias en tamaño en cada DRC y la presencia de una Cys libre en el DRC2 que está expuesta al disolvente. C) Se observa que los cuatro dominios



presentan los aminoácidos conservados que participan en el reconocimiento de su ligando.

**Tabla 6.** Cisteínas libres y formando puentes disulfuro en cada DRC de HrGal

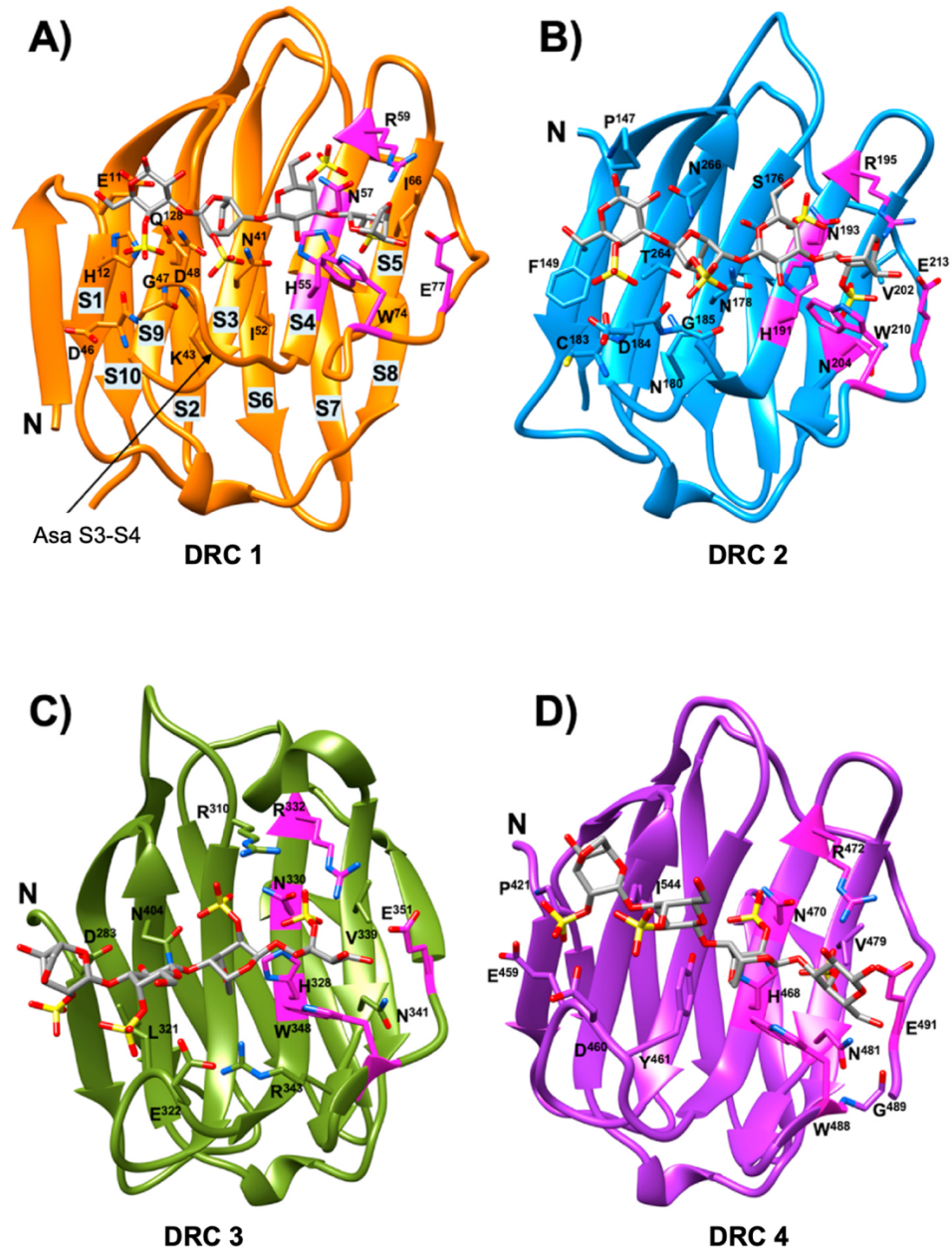
	<b>Puentes disulfuro</b>	<b>Cys libres</b>
DRC 1	Cys <sup>44</sup> ---Cys <sup>50</sup>	Cys <sup>35</sup>
DRC 2	Cys <sup>181</sup> ---Cys <sup>255</sup>	Cys <sup>183</sup> y Cys <sup>272</sup>
DRC 3	----	Cys <sup>393</sup>
DRC 4	Cys <sup>539</sup> ---Cys <sup>455</sup>	



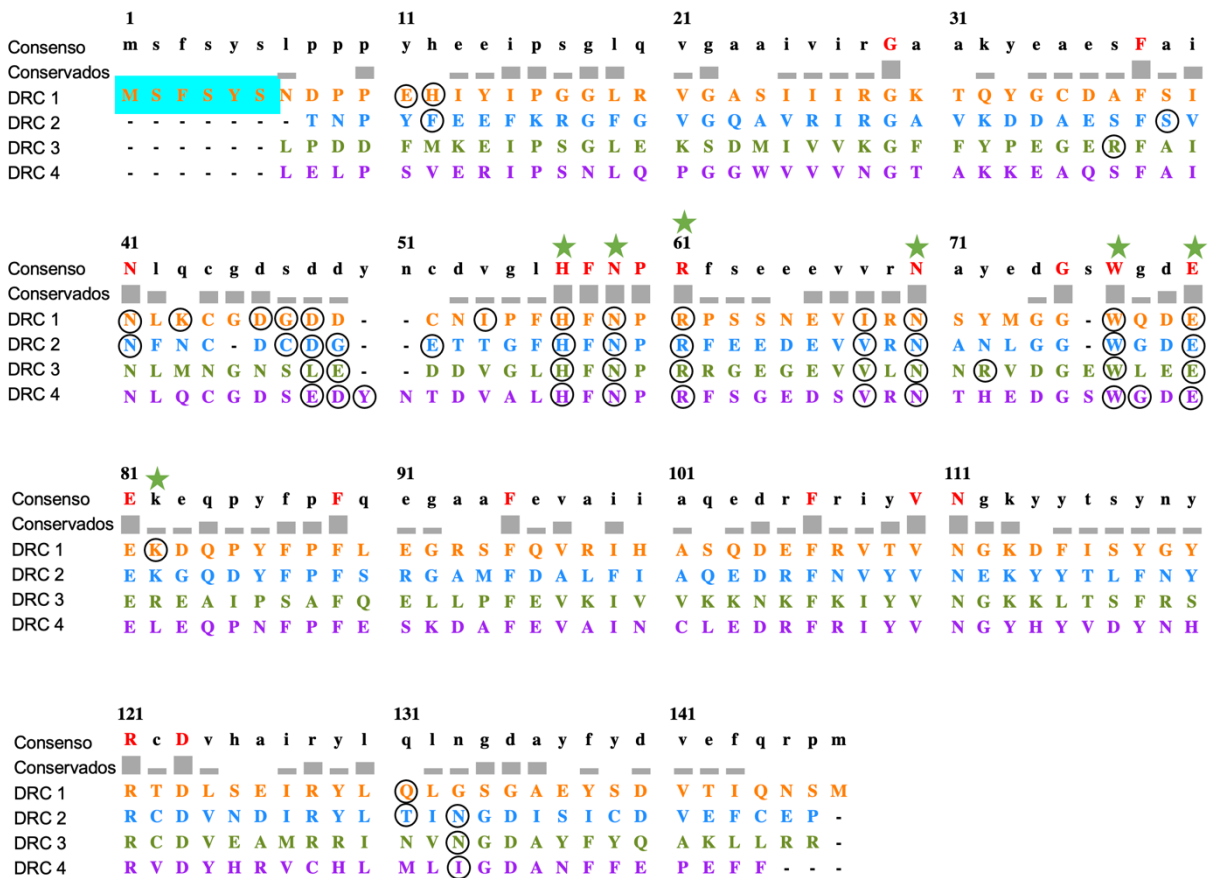
**Figura 33.** Dominio N-terminal de la galectina 8 de humano (rojo) sobrelapada con los diferentes dominios de HrGal. A la derecha se muestra el acercamiento del asa y la hebra beta adicionales en la región N-terminal de ambas galectinas.

### **8.10. Acoplamiento molecular con $\iota$ -carragenano**

Con el fin de entender como reconoce la HrGal al  $\iota$ -carragenano, se llevó a cabo una simulación de acoplamiento molecular con cada DRC. El DRC1 mostró 15 aminoácidos que interactúan con su ligando (Figura 34A). En contraste, el DRC2 mostró 16 aminoácidos (Figura 34B), mientras que en los DRC3 y 4, 13 residuos interactúan con el ligando (Figuras 34C y 34D). Los puentes de hidrógeno que se establecen entre el  $\iota$ -carragenano y los DRCs son 35, 32, 25 y 21 en el DRC1 al DRC4, respectivamente. Estas variaciones en los aminoácidos presentes en el sitio de reconocimiento, así como las diferentes interacciones que éstos favorecen pueden explicar las diferencias en especificidad que cada dominio puede tener. Esto se observó en la galectina 8, donde el DRC1 (N-terminal) es responsable del reconocimiento de azúcares sulfatados, mientras que el DRC2 mostró una especificidad por otros azúcares. Las sustituciones en este sitio pueden permitir un perfil distinto de reconocimiento de su ligando [177]. El alineamiento de los dominios de HrGal mostrado en la Figura 32 resalta los residuos involucrados en la interacción con el ligando (en círculos), y el asa adicional, que se considera responsable de reconocer los azúcares sulfatados (cian). En todos los DRCs, la mayoría de los aminoácidos canónicos interactúan con el ligando (Figura 31 resaltados en magenta). Además, algunos aminoácidos de los DRCs son esenciales para reconocer  $\iota$ -carragenano, tales como Glu<sup>11</sup>, His<sup>12</sup>, Asn<sup>41</sup>, Asp<sup>48</sup> y Gln<sup>128</sup> del DRC1 y los equivalentes de los dominios del 2-4.



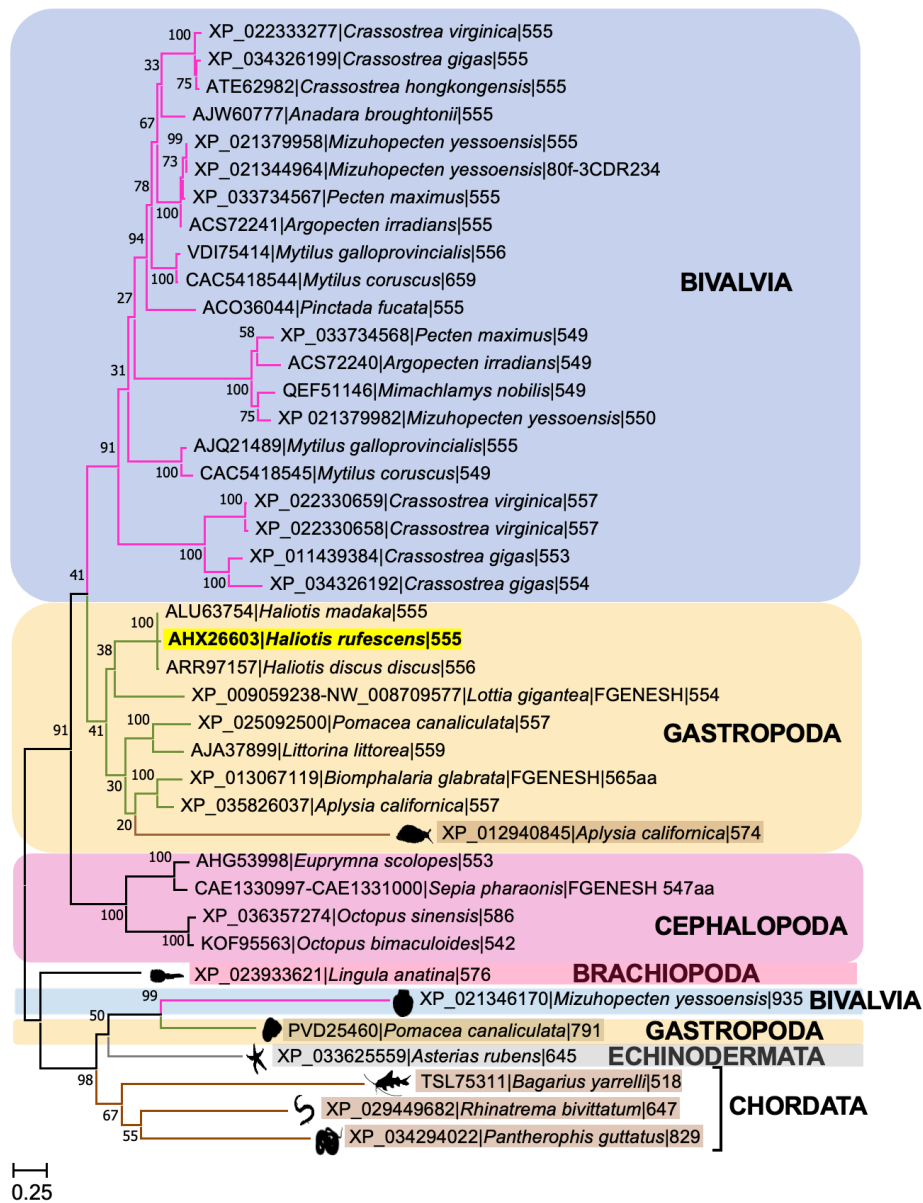
**Figura 34.** Acoplamiento molecular de cada dominio de HrGal con el  $\alpha$ -carragenano. Detalle de la hendidura de unión, donde se indican los residuos de aminoácidos que interactúan con el ligando. A), B), C) y D) corresponden al DRC 1, 2, 3 y 4 respectivamente. Los aminoácidos canónicos involucrados en la interacción de las galectinas con su ligando se muestran en magenta.



**Figura 35.** Alineamiento de secuencias de los cuatro dominios de unión a carbohidratos (DRCs) no idénticos de HrGal. Los aminoácidos conservados se muestran en la secuencia consenso en rojo. Los residuos canónicos que interactúan con el ligando, presentes en todas las galectinas, se marcan con una estrella verde. El asa adicional, considerada responsable del reconocimiento de azúcares sulfatados, se muestra en cian. Los posibles residuos que interactúan con el ι-carragenano están marcados con un círculo.

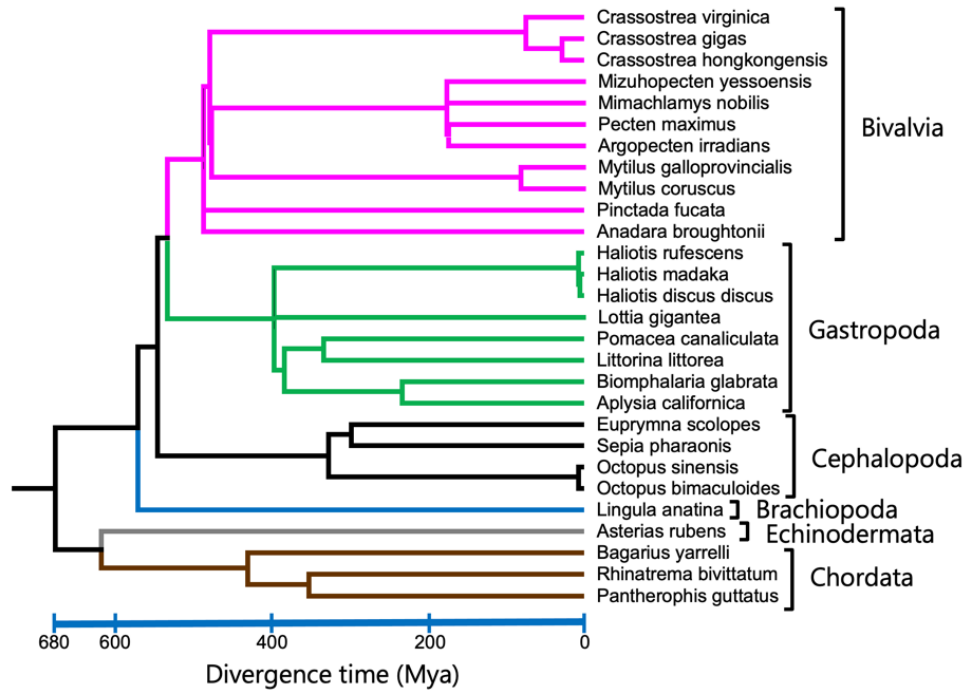
### 8.11. Análisis filogenético de HrGal

A partir de la base de datos de proteínas no redundantes del NCBI y empleando la herramienta BlastP (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), se recuperaron 41 secuencias de galectinas tetradominio homólogas a la galectina de *Haliotis rufescens* (AHX26603). De estas secuencias, 36 corresponden a diferentes especies de moluscos (bivalvos, gastropodos y cefalópodos), tres pertenecen a cordados (*Pantherophis guttatus* (XP\_034294022), *Bagarius yarrelli* (TSL75311), *Rhinatrema bivittatum* (XP\_029449682)), una al equinodermo *Asterias rubens* (XP\_033625559), y otra al brachiopodo *Lingula anatina* (XP\_023933621). La Figura 33 muestra un árbol filogenético construido a partir de un alineamiento múltiple con las secuencias antes mencionadas. En él puede observarse que todas las secuencias de moluscos (con la excepción de dos), se agrupan en una misma rama dentro del árbol, lo que sugiere que se trata de un grupo monofilético de galectinas tetradominio. Es interesante observar que la topología del árbol filogenético obtenida con las galectinas tetradominio (Figura 36), es muy similar a los taxones formados con los tiempos de divergencia recuperados en el árbol construido con TimeTree 5 (Figura 37). Solamente la clase de los bivalvos exhibió una topología más compleja debido a que los genes ancestrales de estas galectinas tetradominio han sufrido varias duplicaciones en esta clase de moluscos. De hecho, en este análisis se consideraron los diez genomas de moluscos que se han secuenciado por completo, donde es interesante observar que en los cefalópodos (*Octopus bimaculoides* y *Octopus sinesis*) y gasterópodos (*Biomphalaria glabrata*, *Lottia gigantea* y *Pomaccea canaliculata*), se encontró solamente un gen para las galectinas tetradominio; excepto *Aplysia californica* que tiene genes para dos galectinas tetradominio, pero uno de ellos es divergente (Figura 36, Figura 37 y Tabla S1-sección de anexos). En el caso contrario, los bivalvos poseen dos (*Pecten maximus*), tres (*Crassostrea virginica* y *Crassostrea gigas*) o incluso cuatro (*Mizuhopecten yessoensis*) genes de galectinas tetradominio.



**Figura 36.** Análisis filogenético de galectinas bidominio y tetradominio. En el árbol filogenético se encuentran 41 secuencias completas de galectinas tetradominio. Se construyó un árbol filogenético sin raíz, usando el método de máxima verosimilitud con el modelo de Le-Gascuel en MEGA X. Se utilizó un bootstrap de 500 réplicas. Los nombres de las secuencias se indican con el número de acceso de GenBank, seguido de la especie y la longitud de la secuencia. Los clados de bivalvos están sombreadas en celeste, el de gasterópodos en amarillo, cefalópodos en rosa. La secuencia de *H. rufescens* se

encuentra resaltada en amarillo, así como todas las especies que tenían origen independiente de las galectinas tetradominio de moluscos (*Pantherophis guttatus*, *Bagarius yarrelli*, *Rhinatrema bivittatum*, *Asterias rubens*, *Pomacea canaliculata*, *Aplysia californica*, *Mizuhopecten yessoensis* y *Lingula anatina*).



**Figura 37.** Escala de tiempo para la divergencia de diferentes moluscos según la calibración de consenso recuperada principalmente de [4] (<http://www.timetree.org>).

Para obtener evidencia adicional que permita tener una mejor idea sobre el posible origen de estas galectinas tetradominio en moluscos, los dominios individuales de cada galectina tetradominio se cortaron *in silico*, y sus secuencias se alinearon entre ellas (Figura 35).

Adicionalmente, se recuperaron de la base de datos de proteínas no redundantes del NCBI, todos los dominios individuales de galectinas encontrados en moluscos, así como los dominios individuales de galectinas presentes en las 100 secuencias con mayor identidad a la galectina de *Haliotis rufescens* (AHX26603). Como referencia, se incluyeron también en el análisis todos los dominios individuales

de galectinas identificadas por el grupo del Dr. Denis Houzelstein (2004) [178], presentes en humanos, rata, y ratón, así como en *Xenopus laevis* y *X. tropicalis*. De esta manera, se obtuvo un alineamiento con 448 dominios individuales de galectinas, pertenecientes todos a animales, ya sea en forma de galectinas monodominio, bidominio, tridominio, tetradominio o incluso, con mayor número de dominios. La Figura 35 muestra un árbol filogenético construido a partir de los dominios individuales. Es interesante observar que, en este árbol cada uno de los dominios individuales de las galectinas tetradominio de moluscos (nombrados como DRC 1 a DRC 4), se agrupan también en ramas independientes: –DRC1 (azul), DRC2 (rojo), DRC 3 (verde) y DRC 4 (naranja), sugiriendo fuertemente que los cuatro dominios de las galectinas tetradominio de moluscos forman un grupo monofilético. Las galectinas tetradominio que no forman parte de este grupo monofilético en moluscos, están señaladas en la Figura 33 con una pequeña imagen ilustrando la silueta de la especie a la que pertenece la secuencia. En la Figura 35 estas mismas siluetas muestran la ubicación de estas secuencias divergentes, en donde claramente puede observarse que se ubican en ramas separadas, mostrando que su origen es independiente del resto de las galectinas tetradominio de moluscos. En el caso del equinodermo *Asterias rubens* (XP\_033625559), sus dominios CRD1 y CRD3 son casi idénticos, al igual que los dominios CRD2 y CRD4. Esto sugiere que su galectina tetradominio se originó a partir de una galectina bidominio que sufrió una duplicación en tándem. Un evento similar de duplicación en tándem que ocurrió de forma independiente dio origen a la galectina tetradominio divergente de *Pomacea canaliculata* (PVD25460); de hecho, es interesante que los dominios independientes de esta última proteína se agrupan dentro de los clados que comprenden a las galectinas bidominio de moluscos. Este resultado muestra también que, con excepción de la galectina tetradominio divergente de *Pomacea canaliculata* (PVD25460), el resto de las galectinas tetradominio de moluscos no se originaron a partir de una duplicación de las galectinas bidominio de moluscos, agrupadas en la Figura 34 como ramas de color morado (2D-CRD1) y rosa (2D-CRD2). Esto último concuerda con los resultados previamente publicados por Vasta et al. (2015) [46], quien después de un análisis filogenético que incluyó 11 galectinas tetradominio y 9



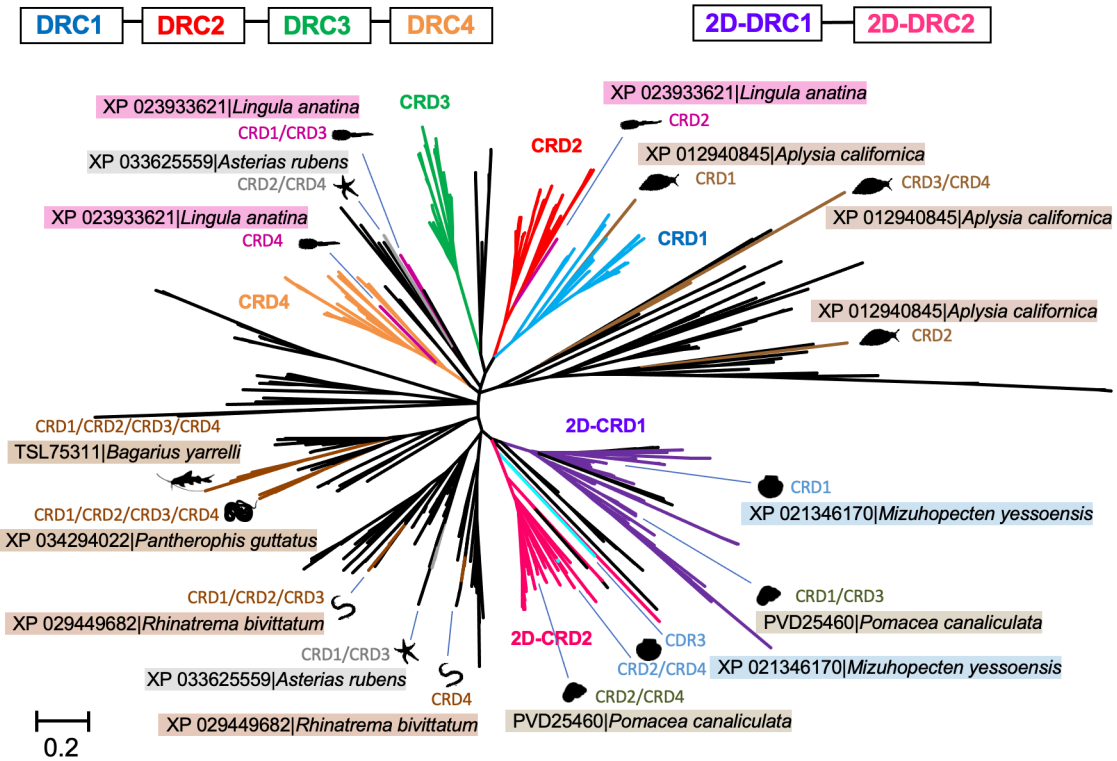
galectinas bidominio de moluscos, concluyó que las galectinas bidominio y tetradominio conforman dos grupos que se originaron de forma independiente. La galectina tetradominio divergente de *Mizuhopecten yessoensis* (XP\_021346170), también se originó muy probablemente a partir de las galectinas bidominio de moluscos, pero no mediante una simple duplicación en tándem como en el caso de la secuencia divergente de *Pomacea canaliculata* (PVD25460).

Las galectinas de cuatro dominios de los vertebrados, *Pantherophis guttatus* (XP\_034294022) y *Bagarius yarrelli* (TSL75311), sus cuatro CRD's se agrupan todos entre ellos mismos de manera independiente, lo que sugiere que, en cada galectina, ocurrieron duplicaciones recientes de un mismo dominio que dieron lugar a proteínas tetradominio con dominios casi idénticos. En el caso del vertebrado *Rhinatrema bivittatum* (XP\_029449682), tres de sus dominios son casi idénticos (CRD1, CRD2, CRD3), pero el cuarto CRD4 es divergente y se coloca en una posición muy separada de las otras en el árbol. Notablemente, estas tres galectinas tetradominio están muy separadas entre ellas, lo que implica que se originaron cada una de ellas de manera independiente y que no tienen relación con las galectinas tetradominio de moluscos.

En el caso de *Aplysia californica* (XP\_012940845), tres de sus cuatro CRD's son divergentes (CRD1, CRD2 y CRD3) con respecto a las galectinas tetradominio de moluscos. Sorpresivamente, su CRD1 se agrupa dentro de la misma rama que las del grupo monofilético de los moluscos, su CRD2 se encuentra muy separado de los otros, mientras que el CRD3 y CRD4 pudieron haber originado por una duplicación reciente. Otra excepción es el braquiópodo *Lingula anatina* (XP\_023933621), ya que el CRD2 y CRD4 se encuentran dentro de las ramas correspondientes al grupo monofilético de las galectinas tetradominio de moluscos, sin embargo, sus CRD1 y CRD3 posiblemente son resultado de una duplicación reciente, pero a partir de un dominio de galectina distinto (localizado en una rama diferente).

## GALECTINAS TETRADOMINIO

## GALECTINAS BIDOMINIO



**Figura 38.** Relación filogenética de dominios individuales de galectinas. El árbol filogenético sin raíz fue construido con el método de máxima verosimilitud con el modelo de Whelan y Goldman, en MEGA X. Se utilizó un bootstrap de 500 réplicas. En el árbol filogenético radial se encuentran los 448 CRD analizados. Las galectinas bidominio y tetradominio de invertebrados marinos se encuentran identificados por color, tal y como se muestra en las cajas. Las secuencias que presentaban excepciones se encuentran resaltadas, *Pantherophis guttatus*, *Bagarius yarrelli*, *Rhinatrema bivittatum*, *Asterias rubens*, *Pomacea canaliculata*, *Aplysia californica*, *Mizuhopecten yessoensis* y *Lingula anatina*.

En el árbol filogenético del material suplementario (Figura S1 en la sección de anexo), se pueden observar más a detalle cada una de las 448 secuencias estudiadas. Es interesante observar en esta figura, que las galectinas bidominio de moluscos (2D-CRD1/mollusk, 2D-CRD2/mollusk) y de nematodos (2D-CRD1/nematoda y 2D-CRD2/nematoda) formaban clados monofiléticos separados

que sugieren un origen independiente en cada uno de sus respectivos clados. A partir de esta figura, es claro que ninguno de los dominios de las galectinas tetradominio de moluscos se agrupa con alguno de los dominios presentes en vertebrados. Esto nos permite inferir que ningún dominio de las galectinas tetradominio de moluscos es ortólogo a alguno de los presentes en vertebrados.

## 9. CONCLUSIONES

Se logró la purificación de una galectina con cuatro dominios en tándem y una masa molecular de 64.1 kDa del molusco marino *Haliotis rufescens* a la que nombramos HrGal y se establecieron las condiciones para mantener su estabilidad y actividad.

HrGal mostró especificidad por galactósidos y glucósidos, pero sólo los azúcares sulfatados heparina y el  $\iota$ -carragenano inhibieron su actividad hemaglutinante. Los ensayos de inhibición de la actividad hemoaglutinante de HrGal con heparina y carragenano, con diferentes grados de sulfatación y con variaciones en la posición de los grupos sulfato, permitieron establecer que no sólo es importante la presencia de los grupos sulfato sino la posición de éstos para inhibir su actividad hemoaglutinante

Debido a que de la fuente natural se obtiene HrGal con rendimiento muy bajos, ésta se obtuvo de forma recombinante (rHrGal). Sin embargo, la proteína se purificó a partir de los cuerpos de inclusión, solubilizándolos con TBS-Urea 6M. Además, se determinó que la presencia de  $\text{Cu}^{+2}$ , DTT y galactosa eran necesarios para un apropiado re-plegamiento.

En el modelo estructural de HrGal se observó que la Cys<sup>183</sup> localizada en el DRC2 está expuesta al disolvente además de estar localizada en el asa  $\beta$ 3- $\beta$ 4, por lo que debe ser la responsable de la pérdida de la actividad de la lectina en ausencia de un agente reductor. Un asa y una hebra  $\beta$  adicional presentes sólo en el DRC1 y que también se observa en la galectina 8 de humano, es considerada como la responsable del reconocimiento de los azúcares sulfatados. Los DRC de HrGal aun cuando conservan los aminoácidos canónicos presentes en todas las galectinas no son idénticos. Esas diferencias aunadas a la diferencia en el asa  $\beta$ 3- $\beta$ 4 que se considera de gran importancia para el reconocimiento de los ligandos y que en cada dominio tiene una longitud y arreglo diferente, pueden provocar diferencias en especificidad en cada DRC, lo que se tendría que probar más adelante.

Los estudios de filogenia permitieron determinar que en los diferentes taxones ocurrieron distintas duplicaciones y fusiones de dominios que dieron lugar a proteínas con más de un dominio. Así, las distintas galectinas de 2, 3, 4 y más

dominios, muy probablemente se originaron de manera independiente de las galectinas multidominio que se encuentran en vertebrados. Encontramos que las secuencias de moluscos forman un grupo monofilético de galectinas de tetradominio no relacionadas con las galectinas de vertebrados.

## 10. REFERENCIAS

- [1] A.-C. Villani, S. Sarkizova, N. Hacohen, Systems Immunology: Learning the Rules of the Immune System, *Annu. Rev. Immunol.* 36 (2018) 813–842. <https://doi.org/10.1146/annurev-immunol-042617-053035>.
- [2] C.A. Janeway, R. Medzhitov, Innate Immune Recognition, *Annu. Rev. Immunol.* 20 (2002) 197–216. <https://doi.org/10.1146/annurev.immunol.20.083001.084359>.
- [3] R.C.F. Cheung, J.H. Wong, W. Pan, Y.S. Chan, C. Yin, X. Dan, T.B. Ng, Marine lectins and their medicinal applications, *Appl. Microbiol. Biotechnol.* 99 (2015) 3755–3773. <https://doi.org/10.1007/s00253-015-6518-0>.
- [4] X.K. Jin, W.W. Li, L. Cheng, S. Li, X.N. Guo, A.Q. Yu, M.H. Wu, L. He, Q. Wang, Two novel short C-type lectin from Chinese mitten crab, *Eriocheir sinensis*, are induced in response to LPS challenged, *Fish Shellfish Immunol.* 33 (2012) 1149–1158. <https://doi.org/10.1016/j.fsi.2012.08.027>.
- [5] J. Yang, L. Wang, H. Zhang, L. Qiu, H. Wang, L. Song, C-type Lectin in *Chlamys farreri* (CfLec-1) mediating immune recognition and opsonization, *PLoS One.* 6 (2011) 16–19. <https://doi.org/10.1371/journal.pone.0017089>.
- [6] N. Sharon, H. Lis, History of lectins: From hemagglutinins to biological recognition molecules, *Glycobiology.* 14 (2004) 53–62. <https://doi.org/10.1093/glycob/cwh122>.
- [7] W.C. Boyd, E. Shapleigh, Specific Precipitating Activity of Plant Agglutinins (Lectins), *Science* (80- ). 119 (1954) 419–419. <https://doi.org/10.1126/science.119.3091.419>.
- [8] N. Sharon, H. Lis, Lectins as Cell Recognition Molecules, *Am. Assoc. Adv. Sci.* 246 (2016) 227–234. <http://www.jstor.org/stable/1704348>.

- [9] Y. Huang, J.M. Tan, Z. Wang, S.W. Yin, X. Huang, W. Wang, Q. Ren, Cloning and characterization of two different L-type lectin genes from the Chinese mitten crab *Eriocheir sinensis*, *Dev. Comp. Immunol.* 46 (2014) 255–266. <https://doi.org/10.1016/j.dci.2014.04.015>.
- [10] S.M.G. Filho, J.D. Cardoso, K. Anaya, E.S. Do Nascimento, J.T.J.G. De Lacerda, R. Mioso, T.S. Gadelha, C.A. De Almeida Gadelha, Marine sponge lectins: Actual status on properties and biological activities, *Molecules.* 20 (2015) 348–357. <https://doi.org/10.3390/molecules20010348>.
- [11] W. Senghoi, P. Runsaeng, P. Utarabhand, FmLC5, a putative galactose-binding C-type lectin with two QPD motifs from the hemocytes of *Fenneropenaeus merguensis* participates in shrimp immune defense, *J. Invertebr. Pathol.* 150 (2017) 136–144. <https://doi.org/10.1016/j.jip.2017.09.013>.
- [12] R.M. Moura, A.F.S. Queiroz, J.M.S.L.L. Fook, A.S.F. Dias, N.K.V. Monteiro, J.K.C. Ribeiro, G.E.D.D. Moura, L.L.P. Macedo, E.A. Santos, M.P. Sales, CvL, a lectin from the marine sponge *Cliona varians*: Isolation, characterization and its effects on pathogenic bacteria and *Leishmania promastigotes*, *Comp. Biochem. Physiol. - A Mol. Integr. Physiol.* 145 (2006) 517–523. <https://doi.org/10.1016/j.cbpa.2006.08.028>.
- [13] X. Dan, W. Liu, T.B. Ng, Development and Applications of Lectins as Biological Tools in Biomedical Research, *Med. Res. Rev.* 36 (2016) 221–247. <https://doi.org/10.1002/med.21363>.
- [14] E. Fischer, N. Quoc Khang, G. Letendre, R. Brossmer, A lectin from the Asian horseshoe crab *Tachypleus tridentatus*: purification, specificity and interaction with tumour cells, *Glycoconj. J.* 11 (1994) 51–58. <https://doi.org/10.1007/BF00732432>.
- [15] R.C.F. Cheung, J.H. Wong, W.L. Pan, Y.S. Chan, C.M. Yin, X.L. Dan, H.X. Wang, E.F. Fang, S.K. Lam, P.H.K. Ngai, L.X. Xia, F. Liu, X.Y. Ye, G.Q. Zhang, Q.H. Liu, O. Sha, P. Lin, C. Ki, A.A. Bekhit, A.E.D. Bekhit, D.C.C. Wan, X.J. Ye, J. Xia, T.B. Ng, Antifungal and antiviral products of marine organisms, *Appl. Microbiol. Biotechnol.* 98 (2014) 3475–3494. <https://doi.org/10.1007/s00253-014-5575-0>.

- [16] A. Silva-Becerril, A. Quintero-Martínez, A. Hernández-Santoyo, Structural and functional analysis of a tandem repeat galacturonic acid-binding lectin from the sea hare *Aplysia californica*, *Fish Shellfish Immunol.* 132 (2023) 108513. <https://doi.org/10.1016/j.fsi.2022.108513>.
- [17] W. Maldonado-Aguayo, J. Teneb, C. Gallardo-Escárate, A galectin with quadruple-domain from red abalone *Haliotis rufescens* involved in the immune innate response against to *Vibrio anguillarum*, *Fish Shellfish Immunol.* 40 (2014) 1–8. <https://doi.org/10.1016/j.fsi.2014.06.013>.
- [18] M.D. Cooper, M.N. Alder, The evolution of adaptive immune systems, *Cell.* 124 (2006) 815–822. <https://doi.org/10.1016/j.cell.2006.02.001>.
- [19] D.D. Chaplin, Overview of the immune response, *J. Allergy Clin. Immunol.* 125 (2010) S3–S23. <https://doi.org/10.1016/j.jaci.2009.12.980>.
- [20] J.A. Hoffmann, Phylogenetic Perspectives in Innate Immunity, *Science* (80- ). 284 (1999) 1313–1318. <https://doi.org/10.1126/science.284.5418.1313>.
- [21] J. Ye, F. Zhou, A.M.Q. Al-Kareef, H. Wang, Anticancer agents from marine sponges, *J. Asian Nat. Prod. Res.* 17 (2015) 64–88. <https://doi.org/10.1080/10286020.2014.970535>.
- [22] N.M. Gowda, U. Goswami, M. Islam Khan, T-antigen binding lectin with antibacterial activity from marine invertebrate, sea cucumber (*Holothuria scabra*): Possible involvement in differential recognition of bacteria, *J. Invertebr. Pathol.* 99 (2008) 141–145. <https://doi.org/10.1016/j.jip.2008.04.003>.
- [23] J. Kurtz, Memory in the innate and adaptive immune systems, *Microbes Infect.* 6 (2004) 1410–1417. <https://doi.org/10.1016/j.micinf.2004.10.002>.
- [24] Y.A. Marin, R. Salazar-Lugo, El sistema inmune de los invertebrados, *Rev. Electrónica Vet.* 10 (2009). [#0](http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:El+sistema+inmune+de+los+invertebrados+-+The+immune+system+of+the+invertebrates+).
- [25] A.E. Ward, B.M. Rosenthal, Evolutionary responses of innate immunity to adaptive immunity, *Infect. Genet. Evol.* 21 (2014) 492–496. <https://doi.org/10.1016/j.meegid.2013.12.021>.

- [26] M. Gerdol, P. Venier, An updated molecular basis for mussel immunity, *Fish Shellfish Immunol.* 46 (2015) 17–38. <https://doi.org/10.1016/j.fsi.2015.02.013>.
- [27] L. Cerenius, K. Söderhäll, Crayfish immunity – Recent findings, *Dev. Comp. Immunol.* 80 (2018) 94–98. <https://doi.org/10.1016/j.dci.2017.05.010>.
- [28] B. Allam, E. Pales Espinosa, Bivalve immunity and response to infections: Are we looking at the right place?, *Fish Shellfish Immunol.* 53 (2016) 4–12. <https://doi.org/10.1016/j.fsi.2016.03.037>.
- [29] B. Allam, E.P. Espinosa, *Mucosal immunity in mollusks*, Elsevier Inc., 2015. <https://doi.org/10.1016/B978-0-12-417186-2.00012-1>.
- [30] E. de Oliveira Figueiroa, C.R. Albuquerque da Cunha, P.B.S. Albuquerque, R.A. de Paula, M.A. Aranda-Souza, M.S. Alves, A. Zagmignan, M.G. Carneiro-da-Cunha, L.C. Nascimento da Silva, M.T. dos Santos Correia, Lectin-Carbohydrate Interactions: Implications for the Development of New Anticancer Agents, *Curr. Med. Chem.* 24 (2017) 3667–3680. <https://doi.org/10.2174/0929867324666170523110400>.
- [31] K. Lebed, A.J. Kulik, L. Forró, M. Lekka, Lectin–carbohydrate affinity measured using a quartz crystal microbalance, *J. Colloid Interface Sci.* 299 (2006) 41–48. <https://doi.org/10.1016/j.jcis.2006.01.053>.
- [32] K. Drickamer, Two distinct classes of carbohydrate-recognition domains in animal lectins, *J. Biol. Chem.* 263 (1988) 9557–9560.
- [33] and A.V. Maureen E. Taylor, Kurt Drickamer, Ronald L. Schnaar, Marilyn E. Etzler, *Discovery and Classification of Glycan-Binding Proteins*, in: et al. Varki A, Cummings RD, Esko JD (Ed.), *Essentials Glycobiol.*, 3rd ed, n.d.
- [34] S.S.K. Santosh, K Gupta, Madan, L Verma, *Introduction to Animal Lectins*, in: G.R. Vasta PhD, H. Ahmed PhD (Eds.), *Anim. Lectins A Funct. View*, 1st ed., CRC Press, 2008. <https://doi.org/10.1201/9781420006971>.
- [35] K. Drickamer, *Classification of Animal Lectins*, (n.d.). <http://www.imperial.ac.uk/research/animallecins/ctld/lectins.html>.
- [36] K. Drickamer, *Information about Calnexin and Calreticulin.*, (2014). <https://www.imperial.ac.uk/research/animallecins/ctld/classes/CNX1.html> (accessed January 26, 2021).



- [37] K. Drickamer, Information about M-type lectins., (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/M-type1.html> (accessed January 26, 2021).
- [38] I.A. Taylor ME, Drickamer K, Major Biological Functions of GBPs, in: et al. Varki A, Cummings RD, Esko JD (Ed.), Essentials Glycobiol. [Internet]., 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2022. <https://doi.org/10.1101/glycobiology.4e.28>.
- [39] K. Drickamer, Information about L-type lectins, (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/L-type1.html> (accessed January 26, 2021).
- [40] S. Varki, A., and Kornfeld, P-Type Lectins., in: et al. Varki A, Cummings RD, Esko JD (Ed.), Essentials Glycobiol. [Internet]., 3rd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2017. <https://www.ncbi.nlm.nih.gov/books/NBK453066/> (accessed January 27, 2021).
- [41] K. Drickamer, Information about the P-type lectins., (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/P-type1.html> (accessed January 27, 2021).
- [42] R. Cummings, R., McEver, C-Type Lectins., in: et al. Varki A, Cummings RD, Esko JD (Ed.), Essentials Glycobiol. [Internet]., 3rd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2017. <https://www.ncbi.nlm.nih.gov/books/NBK453028/> (accessed January 27, 2020).
- [43] L.F. Cummings RD, Galectins, in: et al. Varki A, Cummings RD, Esko JD (Ed.), Essentials Glycobiol. [Internet], 2nd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2009. <https://www.ncbi.nlm.nih.gov/pbidi.unam.mx:2443/books/NBK1944/>.
- [44] V.G. Cummings RD, Liu FT, Galectins., in: et al. Varki A, Cummings RD, Esko JD (Ed.), Essentials Glycobiol. [Internet], 3rd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2017. <https://www.ncbi.nlm.nih.gov/books/NBK453091/> doi: 10.1101/glycobiology.3e.036.
- [45] S.H. Barondes, D.N. Cooper, M.A. Gitt, H. Leffler, Galectins. Structure and

- function of a large family of animal lectins., *J. Biol. Chem.* 269 (1994) 20807–10. <http://www.ncbi.nlm.nih.gov/pubmed/8063692>.
- [46] G.R. Vasta, C. Feng, M.A. Bianchet, T.R. Bachvaroff, S. Tasumi, Structural, functional, and evolutionary aspects of galectins in aquatic mollusks: From a sweet tooth to the Trojan horse, *Fish Shellfish Immunol.* 46 (2015) 94–106. <https://doi.org/10.1016/j.fsi.2015.05.012>.
- [47] K. Drickamer, Information about I-type lectins, (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/I-type1.html> (accessed October 7, 2020).
- [48] P. Varki, A., Schnaar, R., & Crocker, I-Type Lectins, in: et al. Varki A, Cummings RD, Esko JD (Ed.), *Essentials Glycobiol.* [Internet]., 3rd ed., Cold Spring Harbor Laboratory Press., 2017. <https://doi.org/10.1101/glycobiology.3e.035>.
- [49] K. Drickamer, Information about R-type lectins., (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/R-type1.html> (accessed October 8, 2020).
- [50] L.S.R. Cummings RD, R-Type Lectins., in: et al. Varki A, Cummings RD, Esko JD (Ed.), *Essentials Glycobiol.* [Internet]., 3rd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2017. <https://doi.org/10.1101/glycobiology.3e.031>.
- [51] K. Drickamer, F-box lectins., (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/U-type1.htm> (accessed October 8, 2020).
- [52] N. Lannoo, W.J. Peumans, E.J.M. Van Damme, Do F-box proteins with a C-terminal domain homologous with the tobacco lectin play a role in protein degradation in plants?, *Biochem. Soc. Trans.* 36 (2008) 843–847. <https://doi.org/10.1042/BST0360843>.
- [53] K. Drickamer, Ficolins, (2014). [www.imperial.ac.uk/research/animallelectins/ctld/classes/Ficolin1.htm](http://www.imperial.ac.uk/research/animallelectins/ctld/classes/Ficolin1.htm) (accessed October 8, 2020).
- [54] I. Jarlhelt, K. Pilely, J.B. Clausen, M.-O. Skjoedt, R. Bayarri-Olmos, P. Garred,

- Circulating Ficolin-2 and Ficolin-3 Form Heterocomplexes, *J. Immunol.* 204 (2020) 1919–1928. <https://doi.org/10.4049/jimmunol.1900694>.
- [55] K. Drickamer, Chitinase-like lectins, (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/Chitin1.htm> (accessed October 8, 2020).
- [56] T. Zhao, Z. Su, Y. Li, X. Zhang, Q. You, Chitinase-3 like-protein-1 function and its role in diseases, *Signal Transduct. Target. Ther.* 5 (2020) 201. <https://doi.org/10.1038/s41392-020-00303-7>.
- [57] K. Drickamer, Fucolectins., (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/Fucolectin1.htm> (accessed October 12, 2020).
- [58] S. Mahajan, T.N.C. Ramya, F-type Lectin Domains: Provenance, Prevalence, Properties, Peculiarities, and Potential, in: *Biochem. Biophys. Roles Cell Surf. Mol. Adv. Exp. Med. Biol.*, 2018: pp. 345–363. [https://doi.org/10.1007/978-981-13-3065-0\\_24](https://doi.org/10.1007/978-981-13-3065-0_24).
- [59] K. Drickamer, Intelectins, (2014). <https://www.imperial.ac.uk/research/animallelectins/ctld/classes/Intelectin1.htm> (accessed October 12, 2020).
- [60] B. Mali, J. Soza-Ried, M. Frohme, U. Frank, Structural but not functional conservation of an immune molecule: A tachylectin-like gene in *Hydractinia*, *Dev. Comp. Immunol.* 30 (2006) 275–281. <https://doi.org/10.1016/j.dci.2005.04.004>.
- [61] E. García-Maldonado, P. Cano-Sánchez, A. Hernández-Santoyo, Molecular and functional characterization of a glycosylated Galactose-Binding lectin from *Mytilus californianus*, *Fish Shellfish Immunol.* 66 (2017) 564–574. <https://doi.org/10.1016/j.fsi.2017.05.057>.
- [62] I. V. Chikalovets, S.N. Kovalchuk, A.P. Litovchenko, V.I. Molchanova, M. V. Pivkin, O. V. Chernikov, A new Gal/GalNAc-specific lectin from the mussel *Mytilus trossulus*: Structure, tissue specificity, antimicrobial and antifungal activity, *Fish Shellfish Immunol.* 50 (2016) 27–33. <https://doi.org/10.1016/j.fsi.2016.01.020>.

- [63] Y. Fujii, N. Dohmae, K. Takio, S.M.A. Kawsar, R. Matsumoto, I. Hasan, Y. Koide, R.A. Kanaly, H. Yasumitsu, Y. Ogawa, S. Sugawara, M. Hosono, K. Nitta, J. Hamako, T. Matsui, Y. Ozeki, A lectin from the mussel *Mytilus galloprovincialis* has a highly novel primary structure and induces glycan-mediated cytotoxicity of globotriaosylceramide-expressing lymphoma cells, *J. Biol. Chem.* 287 (2012) 44772–44783. <https://doi.org/10.1074/jbc.M112.418012>.
- [64] M. Gerdol, P. Venier, An updated molecular basis for mussel immunity, *Fish Shellfish Immunol.* 46 (2015) 17–38. <https://doi.org/10.1016/j.fsi.2015.02.013>.
- [65] E. Catanzaro, C. Calcabrini, A. Bishayee, C. Fimognari, Antitumor potential of marine and freshwater lectins, *Mar. Drugs.* 18 (2020). <https://doi.org/10.3390/md18010011>.
- [66] J. Gardères, M.-L. Bourguet-Kondracki, B. Hamer, R. Batel, H. Schröder, W. Müller, Porifera Lectins: Diversity, Physiological Roles and Biotechnological Potential, *Mar. Drugs.* 13 (2015) 5059–5101. <https://doi.org/10.3390/md13085059>.
- [67] R.F. Carneiro, J.T. Viana, R.C.F. Torres, L.T. da Silva, A.L. Andrade, M.A. de Vasconcelos, U. Pinheiro, E.H. Teixeira, C.S. Nagano, A.H. Sampaio, A new mucin-binding lectin from the marine sponge *Aplysina fulva* (AFL) exhibits antibiofilm effects., *Arch. Biochem. Biophys.* 662 (2019) 169–176. <https://doi.org/10.1016/j.abb.2018.12.014>.
- [68] S. Kurata, S. Ariki, S. ichiro Kawabata, Recognition of pathogens and activation of immune responses in *Drosophila* and horseshoe crab innate immunity, *Immunobiology.* 211 (2006) 237–249. <https://doi.org/10.1016/j.imbio.2005.10.016>.
- [69] W.S. Fredrick, S. Ravichandran, Hemolymph proteins in marine crustaceans, *Asian Pac. J. Trop. Biomed.* 2 (2012) 496–502. [https://doi.org/10.1016/S2221-1691\(12\)60084-7](https://doi.org/10.1016/S2221-1691(12)60084-7).
- [70] V.T. Dang, K. Benkendorff, T. Green, P. Speck, Marine Snails and Slugs: a Great Place To Look for Antiviral Drugs: TABLE 1, *J. Virol.* 89 (2015) 8114–8118. <https://doi.org/10.1128/jvi.00287-15>.

- [71] T. Ogawa, M. Watanabe, T. Naganuma, K. Muramoto, Diversified Carbohydrate-Binding Lectins from Marine Resources, *J. Amino Acids*. 2011 (2011) 1–20. <https://doi.org/10.4061/2011/838914>.
- [72] Z. Bai, L. Zhao, X. Chen, Q. Li, J. Li, A galectin contributes to the innate immune recognition and elimination of pathogens in the freshwater mussel *Hyriopsis cumingii*, *Dev. Comp. Immunol.* 73 (2017) 36–45. <https://doi.org/10.1016/j.dci.2017.03.008>.
- [73] R. Praparatana, S. Maskaew, R. Thongsoi, P. Runsaeng, P. Utarabhand, Galectin, another lectin from *Fenneropenaeus merguensis*, contributed in shrimp immune defense, *J. Invertebr. Pathol.* 190 (2022) 107738. <https://doi.org/10.1016/j.jip.2022.107738>.
- [74] Y.T. Zhu, X. Zhang, S.C. Wang, W.W. Li, Q. Wang, Antimicrobial functions of EsLeCH, a C-type lectin, Via JNK pathway in the Chinese mitten crab, *Eriocheir sinensis*, *Dev. Comp. Immunol.* 61 (2016) 225–235. <https://doi.org/10.1016/j.dci.2016.04.007>.
- [75] X. Zhang, J. Lu, C. Mu, R. Li, W. Song, Y. Ye, C. Shi, L. Liu, C. Wang, Molecular cloning of a C-type lectin from *Portunus trituberculatus*, which might be involved in the innate immune response, *Fish Shellfish Immunol.* 76 (2018) 216–223. <https://doi.org/10.1016/j.fsi.2018.01.051>.
- [76] A. Kazanjian, M. Fariñas, Actividades biológicas del extracto acuoso de la esponja *Aplysina lacunosa* (Porifera: Aplysinidae)., *Rev. Biol. Trop.* 54 (2006) 189–200.
- [77] Y. Huang, X. Huang, Z. Wang, J.M. Tan, K.M. Hui, W. Wang, Q. Ren, Function of two novel single-CRD containing C-type lectins in innate immunity from *Eriocheir sinensis*, *Fish Shellfish Immunol.* 37 (2014) 313–321. <https://doi.org/10.1016/j.fsi.2014.02.001>.
- [78] M. Wang, L. Wang, M. Huang, Q. Yi, Y. Guo, Y. Gai, H. Wang, H. Zhang, L. Song, A galectin from *Eriocheir sinensis* functions as pattern recognition receptor enhancing microbe agglutination and haemocytes encapsulation, *Fish Shellfish Immunol.* 55 (2016) 10–20. <https://doi.org/10.1016/j.fsi.2016.04.019>.
- [79] L.D. Hung, B.M. Ly, V.T. Hao, D.T. Trung, V.T.D. Trang, P.T.H. Trinh, N.T.D.

- Ngoc, T.M. Quang, Purification, characterization and biological effect of lectin from the marine sponge *Stylissa flexibilis* (Lévi, 1961), *Comp. Biochem. Physiol. Part - B Biochem. Mol. Biol.* 216 (2018) 32–38. <https://doi.org/10.1016/j.cbpb.2017.11.008>.
- [80] Z. Lu, Z. Ren, C. Mu, R. Li, Y. Ye, W. Song, C. Shi, L. Liu, C. Wang, Characterisation and functional analysis of an L-type lectin from the swimming crab *Portunus trituberculatus*, *Gene*. 664 (2018) 27–36. <https://doi.org/10.1016/j.gene.2018.04.041>.
- [81] R.F. Carneiro, P.H.P. de Lima, R.P. Chaves, R. Pereira, A.L. Pereira, M.A. de Vasconcelos, U. Pinheiro, E.H. Teixeira, C.S. Nagano, A.H. Sampaio, Isolation, biochemical characterization and antibiofilm effect of a lectin from the marine sponge *Aplysina lactuca*, *Int. J. Biol. Macromol.* 99 (2017) 213–222. <https://doi.org/10.1016/j.ijbiomac.2017.02.020>.
- [82] P. Kong, L. Wang, H. Zhang, X. Song, Z. Zhou, J. Yang, L. Qiu, L. Wang, L. Song, A novel C-type lectin from bay scallop *Argopecten irradians* (AiCTL-7) agglutinating fungi with mannose specificity, *Fish Shellfish Immunol.* 30 (2011) 836–844. <https://doi.org/10.1016/j.fsi.2011.01.005>.
- [83] S.K. Ng, Y.T. Huang, Y.C. Lee, E.L. Low, C.H. Chiu, S.L. Chen, L.C. Mao, M.D.T. Chang, A recombinant horseshoe crab plasma lectin recognizes specific pathogen-associated molecular patterns of bacteria through rhamnose e115296, *PLoS One*. 9 (2014) 1–20. <https://doi.org/10.1371/journal.pone.0115296>.
- [84] S. Jayanthi, S. Shanthi, B. Vaseeharan, N. Gopi, M. Govindarajan, N.S. Alharbi, S. Kadaikunnan, J.M. Khaled, G. Benelli, Growth inhibition and antibiofilm potential of Ag nanoparticles coated with lectin, an arthropod immune molecule, *J. Photochem. Photobiol. B Biol.* 170 (2017) 208–216. <https://doi.org/10.1016/j.jphotobiol.2017.04.011>.
- [85] X. Wei, D. Yang, H. Li, H. Jiang, X. Liu, Q. Zhang, J. Yang, Sialic acid-binding lectins (SABLs) from *Solen grandis* function as PRRs ensuring immune recognition and bacterial clearance, *Fish Shellfish Immunol.* 72 (2018) 477–483. <https://doi.org/10.1016/j.fsi.2017.11.026>.

- [86] S. Yoshida, Y. Shimada, D. Kondoh, Y. Kouzuma, A.K. Ghosh, M. Jacobs-Lorena, R.E. Sinden, Hemolytic C-type lectin CEL-III from sea cucumber expressed in transgenic mosquitoes impairs malaria parasite development, *PLoS Pathog.* 3 (2007) 1962–1970. <https://doi.org/10.1371/journal.ppat.0030192>.
- [87] X.W. Zhang, X. Man, X. Huang, Y. Wang, Q.S. Song, K.M. Hui, H.W. Zhang, Identification of a C-type lectin possessing both antibacterial and antiviral activities from red swamp crayfish, *Fish Shellfish Immunol.* 77 (2018) 22–30. <https://doi.org/10.1016/j.fsi.2018.03.015>.
- [88] P.A. Luk'yanov, O. V. Chernikov, S.S. Kobelev, I. V. Chikalovets, V.I. Molchanova, W. Li, Carbohydrate-binding proteins of marine invertebrates, *Russ. J. Bioorganic Chem.* 33 (2007) 161–169. <https://doi.org/10.1134/s1068162007010190>.
- [89] C. Xiong, W. Li, H. Liu, W. Zhang, J. Dou, X. Bai, Y. Du, X. Ma, A normal mucin-binding lectin from the sponge *Craniella australiensis*, *Comp. Biochem. Physiol. - C Toxicol. Pharmacol.* 143 (2006) 9–16. <https://doi.org/10.1016/j.cbpc.2005.11.008>.
- [90] M. Engel, M. Bachmann, H.C. Schröder, B. Rinkevich, Z. Kljajic, G. Uhlenbruck, W.E.G. Müller, A novel galactose- and arabinose-specific lectin from the sponge *Pellina semitubulosa*: isolation, characterization and immunobiological properties, *Biochimie.* 74 (1992) 527–537. [https://doi.org/10.1016/0300-9084\(92\)90150-D](https://doi.org/10.1016/0300-9084(92)90150-D).
- [91] R.R. Dresch, C.B. Lerner, B. Mothes, V.M.T. Trindade, A.T. Henriques, M.M. Vozári-hampe, Biological activities of ACL-I and physicochemical properties of ACL-II , lectins isolated from the marine sponge *Axinella corrugata*, *Comp. Biochem. Physiol. Part B.* 161 (2012) 365–370. <https://doi.org/10.1016/j.cbpb.2012.01.001>.
- [92] O.D. Hendrickson, A. V. Zherdev, Analytical Application of Lectins, *Crit. Rev. Anal. Chem.* 48 (2018) 279–292. <https://doi.org/10.1080/10408347.2017.1422965>.
- [93] R. Dresch, G. Zanetti, F. Irazoqui, V. Sendra, N. Zlocowski, A. Bernardi, R.

- Rosa, A. Battastini, A. Henriques, M. Vozári-Hampe, Staining tumor cells with biotinylated ACL-I, a lectin isolated from the marine sponge, *Axinella corrugata*, *Biotech. Histochem.* 88 (2013) 1–9. <https://doi.org/10.3109/10520295.2012.717304>.
- [94] Y. Fujii, T. Fujiwara, Y. Koide, I. Hasan, S. Sugawara, S. Rajia, S.M.A. Kawsar, D. Yamamoto, D. Araki, R.A. Kanaly, Y. Ogawa, H. Fujita, Y. Ozeki, Internalization of a novel, huge lectin from *Ibacus novemdentatus* (slipper lobster) induces apoptosis of mammalian cancer cells, *Glycoconj. J.* 34 (2017) 85–94. <https://doi.org/10.1007/s10719-016-9731-x>.
- [95] E. García-Maldonado, Estudio de la oligomerización inducida por el reconocimiento a su ligando de una lectina de *Mytilus californianus*, implicaciones en su función., Posgrado en Ciencias Biomédicas, UNAM, 2017.
- [96] J.H. Liao, C.T.H. Chien, H.Y. Wu, K.F. Huang, I. Wang, M.R. Ho, I.F. Tu, I.M. Lee, W. Li, Y.L. Shih, C.Y. Wu, P.A. Lukyanov, S.T.D. Hsu, S.H. Wu, A Multivalent Marine Lectin from *Crenomytilus grayanus* Possesses Anti-cancer Activity through Recognizing Globotriose Gb3, *J. Am. Chem. Soc.* 138 (2016) 4787–4795. <https://doi.org/10.1021/jacs.6b00111>.
- [97] G. Li, S. Mei, J. Cheng, T. Wu, J. Luo, *Haliotis discus discus* sialic acid-binding lectin reduces the oncolytic vaccinia virus induced toxicity in a glioblastoma mouse model, *Mar. Drugs.* 16 (2018) 1–10. <https://doi.org/10.3390/md16050141>.
- [98] X.-W. Wang, J.-X. Wang, Diversity and multiple functions of lectins in shrimp immunity, *Dev. Comp. Immunol.* 39 (2013) 27–38. <https://doi.org/10.1016/j.dci.2012.04.009>.
- [99] S. Mahajan, T.N.C. Ramya, F-type Lectin Domains: Provenance, Prevalence, Properties, Peculiarities, and Potential, (2018) 345–363.
- [100] G.R. Vasta, L. Mario Amzel, M.A. Bianchet, M. Cammarata, C. Feng, K. Saito, F-Type Lectins: A highly diversified family of fucose-binding proteins with a unique sequence motif and structural fold, involved in self/non-self-recognition, *Front. Immunol.* 8 (2017). <https://doi.org/10.3389/fimmu.2017.01648>.
- [101] M.R.F. Marques, M.A. Barracco, Lectins, as non-self-recognition factors, in



- crustaceans, *Aquaculture*. 191 (2000) 23–44. [https://doi.org/10.1016/S0044-8486\(00\)00417-8](https://doi.org/10.1016/S0044-8486(00)00417-8).
- [102] S.N. Kovalchuk, I. V. Chikalovets, O. V. Chernikov, V.I. Molchanova, W. Li, V.A. Rasskazov, P.A. Lukyanov, CDNA cloning and structural characterization of a lectin from the mussel *Crenomytilus grayanus* with a unique amino acid sequence and antibacterial activity, *Fish Shellfish Immunol.* 35 (2013) 1320–1324. <https://doi.org/10.1016/j.fsi.2013.07.011>.
- [103] X. Zhang, L. Pan, J. Yu, H. Huang, One recombinant C-type lectin (LvLec) from white shrimp *Litopenaeus vannamei* affected the haemocyte immune response in vitro, *Fish Shellfish Immunol.* 89 (2019) 35–42. <https://doi.org/10.1016/j.fsi.2019.03.029>.
- [104] X. Wei, L. Wang, W. Sun, M. Zhang, H. Ma, Y. Zhang, X. Zhang, S. Li, C-type lectin B (SpCTL-B) regulates the expression of antimicrobial peptides and promotes phagocytosis in mud crab *Scylla paramamosain*, *Dev. Comp. Immunol.* 84 (2018) 213–229. <https://doi.org/10.1016/j.dci.2018.02.016>.
- [105] X.K. Jin, S. Li, X.N. Guo, L. Cheng, M.H. Wu, S.J. Tan, Y.T. Zhu, A.Q. Yu, W.W. Li, Q. Wang, Two antibacterial C-type lectins from crustacean, *Eriocheir sinensis*, stimulated cellular encapsulation in vitro, *Dev. Comp. Immunol.* 41 (2013) 544–552. <https://doi.org/10.1016/j.dci.2013.07.016>.
- [106] X.N. Guo, X.K. Jin, S. Li, A.Q. Yu, M.H. Wu, S.J. Tan, Y.T. Zhu, W.W. Li, P. Zhang, Q. Wang, A novel C-type lectin from *Eriocheir sinensis* functions as a pattern recognition receptor with antibacterial activity, *Fish Shellfish Immunol.* 35 (2013) 1554–1565. <https://doi.org/10.1016/j.fsi.2013.08.021>.
- [107] P.B. Armstrong, S. Swarnakar, S. Srimal, S. Misquith, E.A. Hahn, R.T. Aimes, J.P. Quigley, A Cytolytic Function for a Sialic Acid-binding Lectin That Is a Member of the Pentraxin Family of Proteins \*, 271 (1996) 14717–14721.
- [108] X. Wei, D. Yang, H. Li, H. Jiang, X. Liu, Q. Zhang, J. Yang, Sialic acid-binding lectins (SABLs) from *Solen grandis* function as PRRs ensuring immune recognition and bacterial clearance, *Fish Shellfish Immunol.* 72 (2018) 477–483. <https://doi.org/10.1016/j.fsi.2017.11.026>.
- [109] M. Adhya, B. Singha, B.P. Chatterjee, Purification and characterization of an N-

- acetylglucosamine specific lectin from marine bivalve *Macoma birmanica*, *Fish Shellfish Immunol.* 27 (2009) 1–8. <https://doi.org/10.1016/j.fsi.2008.11.001>.
- [110] O. V. Chernikov, V.I. Molchanova, I. V. Chikalovets, A.S. Kondrashina, W. Li, P.A. Lukyanov, Lectins of marine hydrobionts, *Biochem.* 78 (2013) 760–770. <https://doi.org/10.1134/s0006297913070080>.
- [111] D. Zhang, S. Jiang, Y. Hu, S. Cui, H. Guo, K. Wu, Y. Li, T. Su, A multidomain galectin involved in innate immune response of pearl oyster *Pinctada fucata*, *Dev. Comp. Immunol.* 35 (2011) 1–6. <https://doi.org/10.1016/j.dci.2010.08.007>.
- [112] X. Song, H. Zhang, J. Zhao, L. Wang, L. Qiu, C. Mu, X. Liu, L. Qiu, L. Song, An immune responsive multidomain galectin from bay scallop *Argopectens irradians*, *Fish Shellfish Immunol.* 28 (2010) 326–332. <https://doi.org/10.1016/j.fsi.2009.11.016>.
- [113] S. Tasumi, G.R. Vasta, A Galectin of Unique Domain Organization from Hemocytes of the Eastern Oyster ( *Crassostrea virginica* ) Is a Receptor for the Protistan Parasite *Perkinsus marinus*, *J. Immunol.* 179 (2007) 3086–3098. <https://doi.org/10.4049/jimmunol.179.5.3086>.
- [114] M. Huang, T. Zhou, Y. Wu, H. Fei, G. Wang, Z. Li, Y. Lei, Q. Liu, C. Sun, Z. Lv, X.-W. Xu, Characterization and functional comparison of single-CRD and multidomain containing galectins CgGal-2 and CgGal-3 from oyster *Crassostrea gigas*, *Fish Shellfish Immunol.* 78 (2018) 238–247. <https://doi.org/10.1016/j.fsi.2018.04.029>.
- [115] X. Song, H. Zhang, L. Wang, J. Zhao, C. Mu, L. Song, L. Qiu, X. Liu, A galectin with quadruple-domain from bay scallop *Argopecten irradians* is involved in innate immune response, *Dev. Comp. Immunol.* 35 (2011) 592–602. <https://doi.org/10.1016/j.dci.2011.01.006>.
- [116] Changhua Xian, J. Chen, Y. Luo, Y. Lu, Q. Chen, X. Deng, Z. Yu, Identification and Characterization of a Novel Quadruple-Domain Galectin from the Hong Kong Oyster *Crassostrea hongkongensis*, *Russ. J. Mar. Biol.* 45 (2019) 320–329. <https://doi.org/10.1134/S1063074019040114>.
- [117] A. V. Kajava, Tandem repeats in proteins: From sequence to structure, *J. Struct. Biol.* 179 (2012) 279–288. <https://doi.org/10.1016/j.jsb.2011.08.009>.

- [118] B. Kobe, A. V Kajava, When protein folding is simplified to protein coiling: the continuum of solenoid protein structures, *Trends Biochem. Sci.* 25 (2000) 509–515. [https://doi.org/10.1016/S0968-0004\(00\)01667-4](https://doi.org/10.1016/S0968-0004(00)01667-4).
- [119] L. Paladin, M. Bevilacqua, S. Errigo, D. Piovesan, I. Mičetić, M. Necci, A.M. Monzon, M.L. Fabre, J.L. Lopez, J.F. Nilsson, J. Rios, P.L. Menna, M. Cabrera, M.G. Buitron, M.G. Kulik, S. Fernandez-Alberti, M.S. Fornasari, G. Parisi, A. Lagares, L. Hirsh, M.A. Andrade-Navarro, A. V Kajava, S.C.E. Tosatto, RepeatsDB in 2021: improved data and extended classification for protein tandem repeat structures, *Nucleic Acids Res.* 49 (2021) D452–D457. <https://doi.org/10.1093/nar/gkaa1097>.
- [120] S. Notova, F. Bonnardel, F. Lisacek, A. Varrot, A. Imberty, Structure and engineering of tandem repeat lectins, *Curr. Opin. Struct. Biol.* 62 (2020) 39–47. <https://doi.org/10.1016/j.sbi.2019.11.006>.
- [121] F. Bonnardel, J. Mariethoz, S. Salentin, X. Robin, M. Schroeder, S. Perez, F. Lisacek, A. Imberty, UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands, *Nucleic Acids Res.* 47 (2019) D1236–D1244. <https://doi.org/10.1093/nar/gky832>.
- [122] H.G. Beisel, S.I. Kawabata, S. Iwanaga, R. Huber, W. Bode, Tachylectin-2: Crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*, *EMBO J.* 18 (1999) 2313–2322. <https://doi.org/10.1093/emboj/18.9.2313>.
- [123] S. Iwanaga, S.I. Kawabata, T. Muta, New types of clotting factors and defense molecules found in horseshoe crab hemolymph: Their structures and functions, *J. Biochem.* 123 (1998) 1–15. <https://doi.org/10.1093/oxfordjournals.jbchem.a021894>.
- [124] T. Uchida, T. Yamasaki, S. Eto, H. Sugawara, G. Kurisu, A. Nakagawa, M. Kusunoki, T. Hatakeyama, Crystal Structure of the Hemolytic Lectin CEL-III Isolated from the Marine Invertebrate *Cucumaria echinata*, *J. Biol. Chem.* 279 (2004) 37133–37141. <https://doi.org/10.1074/jbc.m404065200>.
- [125] H. Unno, S. Goda, T. Hatakeyama, Hemolytic lectin cel-iii heptamerizes via a

- large structural transition from  $\alpha$ -helices to a  $\beta$ -barrel during the transmembrane pore formation process, *J. Biol. Chem.* 289 (2014) 12805–12812. <https://doi.org/10.1074/jbc.M113.541896>.
- [126] G.A. Rabinovich, Integrating structure and function of tandem-repeat galectins, *Front. Biosci.* S4 (2012) 305. <https://doi.org/10.2741/s305>.
- [127] D. Oficial, Acuacultura Abulón rojo, (2013). <https://www.gob.mx/inapesca/acciones-y-programas/abulon-rojo> (accessed January 10, 2023).
- [128] R.E. Masonbrink, C.M. Purcell, S.E. Boles, A. Whitehead, J.R. Hyde, A.S. Seetharam, A.J. Severin, An Annotated Genome for *Haliotis rufescens* (Red Abalone) and Resequenced Green, Pink, Pinto, Black, and White Abalone Species, *Genome Biol. Evol.* 11 (2019) 431–438. <https://doi.org/10.1093/gbe/evz006>.
- [129] A. Vater, B.A. Byrne, B.C. Marshman, L.W. Ashlock, J.D. Moore, Differing responses of red abalone ( *Haliotis rufescens* ) and white abalone ( *H. sorenseni* ) to infection with phage-associated *Candidatus Xenohaliotis californiensis*, *PeerJ.* 6 (2018) e5104. <https://doi.org/10.7717/peerj.5104>.
- [130] J.S. Griffiths, R.M. Sahasrabudhe, M.P.A. Marimuthu, N. Chumchim, O.H. Nguyen, E. Beraut, M. Escalona, A. Whitehead, A draft reference genome of the red abalone, *Haliotis rufescens* , for conservation genomics, *J. Hered.* 113 (2022) 673–680. <https://doi.org/10.1093/jhered/esac047>.
- [131] N.K. Hernández-Ibarra, R.M. Morelos, J.L. Ramírez, P. Cruz, A.R. Leitch, A.M. Ibarra, Chromosomal and molecular characterization of 5S rRNA genes in the North American abalones *Haliotis rufescens* Swainson (red abalone) and *H. fulgens* Philippi (blue abalone), *Gene.* 695 (2019) 65–74. <https://doi.org/10.1016/j.gene.2019.02.003>.
- [132] I. V. Chikalovets, S.N. Kovalchuk, A.P. Litovchenko, V.I. Molchanova, M. V. Pivkin, O. V. Chernikov, A new Gal/GalNAc-specific lectin from the mussel *Mytilus trossulus*: Structure, tissue specificity, antimicrobial and antifungal activity, *Fish Shellfish Immunol.* 50 (2016) 27–33. <https://doi.org/10.1016/j.fsi.2016.01.020>.

- [133] G.A. Fragkiadakis, Isolation of lectins from hemolymph of decapod crustaceans by adsorption on formalinized erythrocytes, *J. Biochem. Biophys. Methods.* 44 (2000) 109–114. [https://doi.org/10.1016/S0165-022X\(00\)00089-0](https://doi.org/10.1016/S0165-022X(00)00089-0).
- [134] M. Huang, T. Zhou, Y. Wu, H. Fei, G. Wang, Z. Li, Y. Lei, Q. Liu, C. Sun, Z. Lv, X.W. Xu, Characterisation and functional comparison of single-CRD and multidomain containing galectins CgGal-2 and CgGal-3 from oyster *Crassostrea gigas*, *Fish Shellfish Immunol.* 78 (2018) 238–247. <https://doi.org/10.1016/j.fsi.2018.04.029>.
- [135] S. Boivin, S. Kozak, R. Meijers, Optimization of protein purification and characterization using Thermofluor screens, *Protein Expr. Purif.* 91 (2013) 192–206. <https://doi.org/10.1016/j.pep.2013.08.002>.
- [136] U.B. Ericsson, B.M. Hallberg, G.T. DeTitta, N. Dekker, P. Nordlund, Thermofluor-based high-throughput stability optimization of proteins for structural studies, *Anal. Biochem.* 357 (2006) 289–298. <https://doi.org/10.1016/j.ab.2006.07.027>.
- [137] F. Santini, A. Bisio, M. Guerrini, E.A. Yates, Modifications under basic conditions of the minor sequences of heparin containing 2,3 or 2,3,6 sulfated-d-glucosamine residues, *Carbohydr. Res.* 302 (1997) 103–108. [https://doi.org/10.1016/S0008-6215\(97\)00111-0](https://doi.org/10.1016/S0008-6215(97)00111-0).
- [138] Y. Inoue, K. Nagasawa, Selective N-desulfation of heparin with dimethyl sulfoxide containing water or methanol, *Carbohydr. Res.* 46 (1976) 87–95. [https://doi.org/10.1016/S0008-6215\(00\)83533-8](https://doi.org/10.1016/S0008-6215(00)83533-8).
- [139] A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy, *Proc. Natl. Acad. Sci.* 112 (2015). <https://doi.org/10.1073/pnas.1500851112>.
- [140] A.J. Miles, R.W. Janes, B.A. Wallace, Tools and methods for circular dichroism spectroscopy of proteins: a tutorial review, *Chem. Soc. Rev.* 50 (2021) 8400–8413. <https://doi.org/10.1039/D0CS00558D>.
- [141] A. Micsonai, É. Moussong, F. Wien, E. Boros, H. Vadász, N. Murvai, Y.-H. Lee, T. Molnár, M. Réfrégiers, Y. Goto, Á. Tantos, J. Kardos, BeStSel: webserver for

- secondary structure and fold prediction for protein CD spectroscopy, *Nucleic Acids Res.* 50 (2022) W90–W98. <https://doi.org/10.1093/nar/gkac345>.
- [142] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature.* 596 (2021) 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- [143] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291. <https://doi.org/10.1107/S0021889892009944>.
- [144] Y. Yan, H. Tao, J. He, S.-Y. Huang, The HDock server for integrated protein–protein docking, *Nat. Protoc.* 15 (2020) 1829–1852. <https://doi.org/10.1038/s41596-020-0312-x>.
- [145] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera. A visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- [146] V. Solovyev, Statistical Approaches in Eukaryotic Gene Prediction, in: *Handb. Stat. Genet.*, John Wiley & Sons, Ltd, Chichester, UK, 2008: pp. 97–159. <https://doi.org/10.1002/9780470061619.ch4>.
- [147] J. Thompson, The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res.* 25 (1997) 4876–4882. <https://doi.org/10.1093/nar/25.24.4876>.
- [148] T.A. Hall, BIOEDIT: a user friendly biological sequence alignment editor and analysis program for windows 95/98/ NT, *Nucleic Acids Symp.* 41 (n.d.) 95–98. [https://doi.org/https://doi.org/10.14601/Phytopathol\\_Mediterr-14998u1.29](https://doi.org/https://doi.org/10.14601/Phytopathol_Mediterr-14998u1.29).
- [149] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms, *Mol. Biol. Evol.* 35

- (2018) 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- [150] and G.R.V. Richard D. Cummings, Fu-Tong Liu, Gabriel A. Rabinovich, Sean R. Stowell, Galectins, in: E. Varki A, Cummings RD, JD (Eds.), *Essentials Glycobiol.* [Internet], 4th ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; NY, 2022. <https://doi.org/10.1101/9781621824213>.
- [151] F.D. Richard, A. V. Kajava, TRDistiller: A rapid filter for enrichment of sequence datasets with proteins containing tandem repeats, *J. Struct. Biol.* 186 (2014) 386–391. <https://doi.org/10.1016/j.jsb.2014.03.013>.
- [152] V. Perovic, J.Y. Leclercq, N. Sumonja, F.D. Richard, N. Veljkovic, A. V Kajava, Tally-2.0: upgraded validator of tandem repeat detection in protein sequences, *Bioinformatics.* 36 (2020) 3260–3262. <https://doi.org/10.1093/bioinformatics/btaa121>.
- [153] N.C.J. Strynadka, M.N.G. James, Towards an understanding of the effects of calcium on protein structure and function, *Curr. Opin. Struct. Biol.* 1 (1991) 905–914. [https://doi.org/10.1016/0959-440X\(91\)90085-8](https://doi.org/10.1016/0959-440X(91)90085-8).
- [154] C. Olsson, R. Zangana, J. Swenson, Stabilization of proteins embedded in sugars and water as studied by dielectric spectroscopy, *Phys. Chem. Chem. Phys.* 22 (2020) 21197–21207. <https://doi.org/10.1039/D0CP03281F>.
- [155] N.S. Gandhi, R.L. Mancera, The structure of glycosaminoglycans and their interactions with proteins, *Chem. Biol. Drug Des.* 72 (2008) 455–482. <https://doi.org/10.1111/j.1747-0285.2008.00741.x>.
- [156] S. Mohamed, D. Coombe, Heparin Mimetics: Their Therapeutic Potential, *Pharmaceuticals.* 10 (2017) 78. <https://doi.org/10.3390/ph10040078>.
- [157] C.A. de Araújo, M.D. Nosedá, T.R. Cipriani, A.G. Gonçalves, M.E.R. Duarte, D.R.B. Ducatti, Selective sulfation of carrageenans and the influence of sulfate regiochemistry on anticoagulant properties, *Carbohydr. Polym.* 91 (2013) 483–491. <https://doi.org/10.1016/j.carbpol.2012.08.034>.
- [158] V.L. Campo, D.F. Kawano, D.B. da Silva, I. Carvalho, Carrageenans: Biological properties, chemical modifications and structural analysis – A review, *Carbohydr. Polym.* 77 (2009) 167–180. <https://doi.org/10.1016/j.carbpol.2009.01.020>.

- [159] J.D. Esko, J. H. Prestegard, R.J. Linhardt, Proteins That Bind Sulfated Glycosaminoglycans, in: et al. Varki A, Cummings RD, Esko JD (Ed.), Essentials Glycobiol., 3rd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2017. <http://www.ncbi.nlm.nih.gov/pubmed/28876859>.
- [160] I.M.B. Francischetti, D. Ma, J.F. Andersen, J.M.C. Ribeiro, Evidence for a lectin specific for sulfated glycans in the salivary gland of the malaria vector, *Anopheles gambiae*, PLoS One. 9 (2014). <https://doi.org/10.1371/journal.pone.0107295>.
- [161] K. Mitsunaga, J. Harada-Itadani, T. Shikanai, H. Tateno, Y. Ikehara, J. Hirabayashi, H. Narimatsu, T. Angata, Human C21orf63 is a heparin-binding protein., J. Biochem. 146 (2009) 369–373. <https://doi.org/10.1093/jb/mvp079>.
- [162] C.A. Duckworth, S.E. Guimond, P. Sindrewicz, A.J. Hughes, N.S. French, L.Y. Lian, E.A. Yates, D. Mark Pritchard, J.M. Rhodes, J.E. Turnbull, L.G. Yu, Chemically modified, non-anticoagulant heparin derivatives are potent galectin-3 binding inhibitors and inhibit circulating galectin-3-promoted metastasis, Oncotarget. 6 (2015) 23671–23687. <https://doi.org/10.18632/oncotarget.4409>.
- [163] P. Sindrewicz, E.A. Yates, J.E. Turnbull, L.Y. Lian, L.G. Yu, Interaction with the heparin-derived binding inhibitors destabilizes galectin-3 protein structure, Biochem. Biophys. Res. Commun. 523 (2020) 336–341. <https://doi.org/10.1016/j.bbrc.2019.12.054>.
- [164] M.L. Talaga, N. Fan, A.L. Fueri, R.K. Brown, P. Bandyopadhyay, T.K. Dam, Multitasking Human Lectin Galectin-3 Interacts with Sulfated Glycosaminoglycans and Chondroitin Sulfate Proteoglycans, Biochemistry. 55 (2016) 4541–4551. <https://doi.org/10.1021/acs.biochem.6b00504>.
- [165] H. Ideo, The N-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity, Glycobiology. 13 (2003) 713–723. <https://doi.org/10.1093/glycob/cwg094>.
- [166] H. Ideo, T. Matsuzaka, T. Nonaka, A. Seko, K. Yamashita, Galectin-8-N-domain Recognition Mechanism for Sialylated and Sulfated Glycans, J. Biol. Chem. 286 (2011) 11346–11355. <https://doi.org/10.1074/jbc.M110.195925>.
- [167] F. Cabassi, B. Casu, A.S. Perlin, Infrared absorption and raman scattering of



- sulfate groups of heparin and related glycosaminoglycans in aqueous solution, *Carbohydr. Res.* 63 (1978) 1–11. [https://doi.org/10.1016/S0008-6215\(00\)80924-6](https://doi.org/10.1016/S0008-6215(00)80924-6).
- [168] P. Cimperman, L. Baranauskienė, S. Jachimovičiūtė, J. Jachno, J. Torresan, V. Michailovienė, J. Matulienė, J. Sereikaitė, V. Bumelis, D. Matulis, A Quantitative Model of Thermal Stabilization and Destabilization of Proteins by Ligands, *Biophys. J.* 95 (2008) 3222–3231. <https://doi.org/10.1529/biophysj.108.134973>.
- [169] N. Bai, H. Roder, A. Dickson, J. Karanicolas, Isothermal Analysis of ThermoFluor Data can readily provide Quantitative Binding Affinities, *Sci. Rep.* 9 (2019) 2650. <https://doi.org/10.1038/s41598-018-37072-x>.
- [170] M.C. Miller, C. Cai, K. Wichapong, S. Bhaduri, N.L.B. Pohl, R.J. Linhardt, H.-J. Gabius, K.H. Mayo, Structural insight into the binding of human galectins to corneal keratan sulfate, its desulfated form and related saccharides, *Sci. Rep.* 10 (2020) 15708. <https://doi.org/10.1038/s41598-020-72645-9>.
- [171] C.M. Guardia, J.J. Caramelo, M. Trujillo, S.P. Méndez-Huergo, R. Radi, D.A. Estrin, G.A. Rabinovich, Structural basis of redox-dependent modulation of galectin-1 dynamics and function, *Glycobiology.* 24 (2014) 428–441. <https://doi.org/10.1093/glycob/cwu008>.
- [172] A.M. McCormick, N.A. Jarmusik, E.J. Endrizzi, N.D. Leipzig, Expression, Isolation, and Purification of Soluble and Insoluble Biotinylated Proteins for Nerve Tissue Regeneration, *J. Vis. Exp.* (2014). <https://doi.org/10.3791/51295>.
- [173] K. Bum-Erdene, H. Leffler, U.J. Nilsson, H. Blanchard, Structural characterisation of human galectin-4 N-terminal carbohydrate recognition domain in complex with glycerol, lactose, 3'-sulfo-lactose, and 2'-fucosyllactose, *Sci. Rep.* 6 (2016) 1–13. <https://doi.org/10.1038/srep20289>.
- [174] T.G. Krupnova, I. V. Mashkova, A.M. Kostryukova, E.E. Schelkanova, S. V. Gavrilkina, Gastropods as potential biomonitors of contamination caused by heavy metals in South Ural lakes, Russia, *Ecol. Indic.* 95 (2018) 1001–1007. <https://doi.org/10.1016/j.ecolind.2017.12.005>.
- [175] F. Guo, R. Tu, W.X. Wang, Different responses of abalone *Haliotis discus*

- hannai to waterborne and dietary-borne copper and zinc exposure, *Ecotoxicol. Environ. Saf.* 91 (2013) 10–17. <https://doi.org/10.1016/j.ecoenv.2013.01.002>.
- [176] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K.S. Wilson, Overview of the CCP 4 suite and current developments, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 67 (2011) 235–242. <https://doi.org/10.1107/S0907444910045749>.
- [177] G.R. Vasta, H. Ahmed, E.W. Odom, Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates, *Curr. Opin. Struct. Biol.* 14 (2004) 617–630. <https://doi.org/10.1016/j.sbi.2004.09.008>.
- [178] D. Houzelstein, I.R. Gonçalves, A.J. Fadden, S.S. Sidhu, D.N.W. Cooper, K. Drickamer, H. Leffler, F. Poirier, Phylogenetic Analysis of the Vertebrate Galectin Family, *Mol. Biol. Evol.* 21 (2004) 1177–1187. <https://doi.org/10.1093/molbev/msh082>.
- [179] M. Bhagwat, L. Aravind, PSI-BLAST Tutorial, in: *Methods Mol Biol.*, 2007: pp. 177–186. [https://doi.org/10.1007/978-1-59745-514-5\\_10](https://doi.org/10.1007/978-1-59745-514-5_10).

## 11. ANEXOS

**Tabla S1.** Galectinas identificadas en el filo Mollusca. El número de acceso de la proteína, el locus del gen y la longitud de la secuencia de aminoácidos se indican en cada galectina. Solamente se incluyeron los organismos con la secuencia de su genoma completa reportada.

Especie [NCBI RefSeq assembly accession] (Número de genes que codifican proteínas)	Monodominio Galectinas	Bidominio Galectinas	Tridominio Galectinas	Tetradominio Galectinas
<b>Clase Bivalvia</b>				
<i>Crassostrea gigas</i> (Pacific oyster) [GCF_902806645] (31,371)	Galectin 4-like Isoform X1 <a href="#">XP_034304231</a> LOC117680443 225 aa  Galectin 4 Isoform X1 <a href="#">XP_011445426</a> LOC105340877 225 aa  Galectin 5 <a href="#">XP_011455178</a> LOC105347705 CGI_10009817 201 aa  Annexin A7 <a href="#">XP_011435530</a>	Galectin 8 <a href="#">XP_011419929</a> LOC105322757 CGI_10022939 340 aa  Galectin-4-like <a href="#">XP_034326071</a> LOC117689346 398 aa  Galectin 4 <a href="#">XP_034304229</a> LOC105340878 CGI_10023950 314 aa  Galectin-4-like <a href="#">XP_034304230</a> LOC105335002		Galectin 4 <a href="#">XP_011439384</a> LOC105336661 CGI_10009751 553 aa  Galectin 6 <a href="#">XP_034326199</a> LOC105342811 CGI_10014752 555 aa  <a href="#">XP_034326192</a> LOC105326968 554 aa

	<p>LOC105333977 CGI_10005550 263 aa</p> <p><a href="#">XP_034333249</a></p> <p>LOC105328237 157 aa</p>	314 aa		
<p><i>Crassostrea virginica</i> RU13XGHG1-28 (eastern oyster) [GCF_002022765] (34,608)</p>	<p>Galectin 4-like <a href="#">XP_022326944</a> LOC111126530 224 aa</p> <p>Galectin 4-like <a href="#">XP_022287017</a> LOC111099845 248 aa</p> <p><a href="#">XP_022345298</a> LOC111137891 153 aa</p> <p><a href="#">XP_022335811</a> LOC111132306 150 aa</p>	<p>Galectin 4-like <a href="#">XP_022333582</a> LOC111130682 321 aa</p> <p>Galectin 4-like <a href="#">XP_023224298</a> LOC111125115 313 aa</p>		<p>Galectin 8-like <a href="#">XP_022333277</a> LOC111130470 555 aa</p> <p><a href="#">XP_022330659</a> LOC111128968 557 aa</p> <p><a href="#">XP_022330658</a> LOC111128967 557 aa</p>
<p><i>Mizuhopecten</i> <i>yessoensis</i> PY sf001</p>	<p>Galectin 3-like <a href="#">XP_021365057</a></p>	<p><a href="#">OWF53975</a> LOC110445716</p>	<p>Galectin-9-like <a href="#">XP_021378026</a></p>	<p>Galectin 4-like <a href="#">XP_021379982</a></p>

<p>(Yesso scallop) [GCF_002113885] (24,532)</p>	<p>LOC110457919 KP79_PYT1899 5 284 aa</p> <p><a href="#">OWF53976</a></p> <p>LOC110445716 KP79_PYT1534 1 135 aa Alternative mRNA (correspond to the last CRD domain from <a href="#">XP_021346170</a>)</p>	<p>KP79_PYT1534 1 343 aa Alternative mRNA (correspond to the first two CRD domains from <a href="#">XP_021346170</a>)</p>	<p>LOC110466071 KP79_PYT0996 9 535 aa</p>	<p>LOC110467261 KP79_PYT1040 9 550 aa</p> <p>Galectin 4 <a href="#">XP_021346170</a></p> <p>LOC110445716 KP79_PYT1534 1 935 aa (proteína químerica: posee un dominio putative de topoisomerase localizado entre el DRC3 y DRC4)</p> <p>Galectin-6 <a href="#">XP_021379958</a></p> <p>LOC110467247 KP79_PYT1040 8 555 aa</p> <p>Galectin-6-like, partial <a href="#">XP_021344964</a> 380 aa (corresponde a CRD2, CRD3 y CRD4)</p>
<p><i>Pecten maximus</i></p>	<p>Galectin 3-like</p>	<p>Galectin 4-like</p>	<p>Galectin 9-like</p>	<p>Galectin 4-like</p>

(scallop) [GCF_902652985] (26,152)	<a href="#">XP_033761955</a> LOC117343625 298 aa  Galectin 5-like <a href="#">XP_033749982</a> LOC117334457 171 aa	<a href="#">XP_033749980</a> LOC117334456 331 aa	<a href="#">XP_033736262</a> LOC117324490 521 aa	<a href="#">XP_033734568</a> LOC117323441 549 aa  <a href="#">XP_033734567</a> LOC117323440 555 aa
<b>Clase Gastropoda</b>				
<i>Aplysia californica</i> F4 8 (California sea hare) [GCF_000002075] (19405)	<a href="#">XP_012943121</a> LOC106013027 270 aa  Uncertain (not included in the tree)	Galectin 4 <a href="#">XP_005102039</a> LOC101852778 337 aa  Galectin 4 <a href="#">XP_005102053</a> LOC101856671 295 aa		<a href="#">XP_035826037</a> LOC101856665 557 aa  <a href="#">XP_012940845</a> LOC101859875 574 aa
<i>Biomphalaria glabrata</i> BB02 (freshwater snail) [GCF_000457365] (25,552)	Galectin-10-like <a href="#">XP_013086827</a> LOC106071296 154 aa  Galectin-7-like <a href="#">XP_013081931</a> LOC106067305 152 aa  Galectin-6-like <a href="#">XP_013073319</a> LOC106060090 188 aa  Galectin-4-like	Galectin-4-like <a href="#">XP_013073318</a> LOC106060089 357 aa  Galectin-4-like <a href="#">NP_001298212</a> LOC106073720 284 aa  Galectin-4-like <a href="#">XP_013073320</a> LOC106060091 316 aa  Galectin-6-like		32 kDa beta- galactoside- binding lectin- like <a href="#">XP_013067119</a> LOC106055414 565 aa

	<p><a href="#">XP_013061349</a> LOC106050849 144 aa</p> <p>Galectin-7-like</p> <p><a href="#">XP_013061348</a> LOC106050847 122 aa</p> <p><a href="#">XP_013067662</a> LOC106055766 255 aa</p> <p><a href="#">XP_013096013</a> LOC106079398 247 aa</p> <p><a href="#">XP_013093692</a> LOC106077497 176 aa Uncertain (not included in the tree)</p> <p><a href="#">XP_013081522</a> LOC106066965 247 aa Uncertain (not included in the tree)</p> <p><a href="#">XP_013067664</a> LOC106055768 246 aa Uncertain (not included in the tree)</p>	<p><a href="#">XP_013086830</a> LOC106071297 333 aa</p> <p>Galectin-4-like</p> <p><a href="#">XP_013073313</a> LOC106060087 329 aa</p>		
--	---	--	--	--

	<p>LOC106053230  <a href="#">XP_013064209</a>  163 aa  Uncertain (not included in the tree)</p> <p><a href="#">XP_013064207</a>  LOC106053229  195 aa  Uncertain (not included in the tree)</p>			
<p><i>Lottia gigantea</i>  (owl limpet)  [GCF_000327385]  (23,818)</p>		<p>Hypothetical protein  LOTGIDRAFT_129554  <a href="#">XP_009063133</a>  254 aa</p>		<p>Hypothetical protein  LOTGIDRAFT_123780  <a href="#">XP_009059238</a>  554 aa</p>
<p><i>Pomacea canaliculata</i>  (golden apple snail)  [GCF_003073045]  (21,144)</p>	<p><a href="#">XP_025099676</a>  LOC112567263  C0Q70_11389  139 aa</p> <p><a href="#">XP_025098963</a>  LOC112566809  C0Q70_11390  130 aa</p> <p><a href="#">XP_025111571</a>  LOC112574598  C0Q70_16195  313 aa</p>	<p>Galectin-9-like  <a href="#">XP_025103617</a>  LOC112569823  305 aa</p> <p>Galectin-4-like  <a href="#">XP_025103625</a>  LOC112569826  319 aa</p>		<p>Galectin-4-like  <a href="#">XP_025092500</a>  LOC112563027  C0Q70_07797  557 aa</p> <p>Hypothetical  <a href="#">PVD25460</a><sup>1</sup>  C0Q70_13116  791 aa  Alternative mRNA  (resulta de la expression putativa en</p>



	<p><a href="#">XP_025110016</a>  LOC112573662  221 aa</p> <p><a href="#">XP_025105622</a>  LOC112571046  C0Q70_13323  348 aa  Uncertain (not included in the tree)</p> <p><a href="#">XP_025105420</a>  LOC112570918  C0Q70_13319  352 aa  Uncertain (not included in the tree)</p> <p><a href="#">XP_025104486</a>  LOC112570317  353 aa</p> <p><a href="#">A0A2T7PML0</a>  PVD34653  C0Q70_05930  635 aa  Uncertain (not included in the tree)</p> <p>Hypothetical protein  A0A2T7P4R9  <a href="#">PVD28425</a></p>			<p>tandem del loci  LOC112569823  +  LOC112569826)</p>
--	--	--	--	--

	C0Q70_11012 157 aa			
<b>Clase Cephalopoda</b>				
<i>Octopus bimaculoides</i> UCB-OBI-ISO-001 (California two-spot octopus) [GCF_001194135] (15842)	Galectin-4-like <a href="#">XP_014786436</a> LOC106880823 302 aa (probably correspond to N-terminal domain of a bidomain galectin)  Galectin-3-like <a href="#">XP_014784583</a> LOC106879493 109 aa (probably correspond to C-terminal domain of a bidomain galectin)	Galectin-4-like <a href="#">XP_014771267</a> LOC106869861 311 aa		Galectin-4-like <a href="#">XP_014767468</a> LOC106867183 597 aa
<i>Octopus sinensis</i> Ov201803 (East Asian common octopus) [GCF_006345805] (18,882)	Galectin-4-like <a href="#">XP_029643127</a> LOC115217540 229 aa	Galectin-4-like <a href="#">XP_029642873</a> LOC115217348 311 aa  Galectin-9-like <a href="#">XP_029642889</a> LOC115217360 309 aa		Galectin-4-like <a href="#">XP_036357274</a> LOC115209822 586 aa

<sup>1</sup> Esta galectina de tetradominio probablemente se origino a partir de la copia de dos galectinas bidominio de moluscos o puede corresponder a un ARNm bicistrónico alternativo que resulto de la expresión putativa en tandém del loci LOC112569823 + LOC112569826.

**Tabla S2.** Galectinas identificadas en vertebrados e invertebrados. El número de acceso de la proteína y la longitud de la secuencia de aminoácidos se indican en cada galectina.

Espece	Monodominio Galectinas	Bidominio Galectinas	Tridominio Galectinas	Tetradominio Galectinas	Hexadominio Galectinas
<b>Clase Bivalvia</b>					
<i>Ostrea edulis</i>		Galectin <a href="#">ADF80416</a> 335 aa			
<i>Argopecten irradians</i>				Aigal1 <a href="#">ACS72240</a> 549 aa Aigal2 <a href="#">ACS72241</a> 555 aa	
<i>Crassostrea hongkongensis</i>				ChGal <a href="#">ATE62982</a> 555 aa	
<i>Anadara broughtonii</i>				<a href="#">AJW60777</a> 555 aa	
<i>Pinctada fucata</i>		Tandem- repeat galectin <a href="#">ADU60337</a> 348 aa		PfGal <a href="#">ACO36044</a> 555 aa	
<i>Mytilus coruscus</i>		Galectin- 4, Galectin-9 <a href="#">CAC542557</a> <a href="#">5</a>		<a href="#">CAC541854</a> <a href="#">4</a> 556 aa FGENESH	

		724 aa  LGALS9 <a href="#">CAC536360</a> <u>3</u> 337 aa  <a href="#">CAC536360</a> <u>4</u> 315 aa		<a href="#">CAC541854</a> <u>5</u> 549 aa	
<i>Mytilus galloprovincialis</i>		MyGal4 <a href="#">VDI06201</a> 268 aa  Galectin-2 <a href="#">AJQ21509</a> 357 aa  Galectin-3 <a href="#">AJQ21510</a> 314 aa  Galectin-4 <a href="#">AJQ21511</a> 331 aa		Hypothetical predicted protein MgGal <a href="#">VDI75414</a> 556 aa  MyGal1 <a href="#">AJQ21489</a> 555 aa	
<i>Mimachlamys nobilis</i>				MinGal <a href="#">QEF51146</a> 549 aa	
<i>Archivesica packardana</i>		Galectin <a href="#">QDJ94295</a> 311			
<i>Hyriopsis cumingii</i>		Galectin <a href="#">ANQ89716</a> 312 aa			
<i>Sinonovacula</i>		Galectin			

<i>constricta</i>		<a href="#">QHI46338</a> 305 aa			
<i>Tegillarca granosa</i>		Galectin <a href="#">AGH06131</a> 354			
<b>Clase Gastropoda</b>					
<i>Haliotis madaka</i>				HmGal <a href="#">ALU63754</a> 555 aa	
<i>Haliotis rufescens</i>				HrGal <a href="#">AHX26603</a> 555 aa	
<i>Haliotis discus discus</i>		Galectin 4 <a href="#">APW78605</a> 306 aa		HdGal <a href="#">ARR97157</a> 556 aa	
<i>Haliotis discus hanai</i>		Galectin 4- like <a href="#">ABN54798</a> 306 aa			
<i>Littorina littorea</i>		Galectin-1 <a href="#">AJA37890</a> 353 aa  Galectin-2-1 <a href="#">AJA37867</a> 353 aa  Galectin-2-2 <a href="#">AJA37869</a> 307 aa		LIGal3 <a href="#">AJA37899</a> 559 aa	
<b>Clase Cephalopoda</b>					
<i>Sepia pharaonis</i>				LGALS4	

				<a href="#">CAE1330997</a> 448 aa	
<i>Euprymna scolopes</i>				EsGal <a href="#">AHG53998</a> 553 aa	
<b>Filo Brachiopoda</b>					
<i>Lingula anatina</i>		Galectin-4 <a href="#">XP_0134085</a> <u>86</u> 357 aa		Galectin-4 isoform <a href="#">XP_0239336</a> <u>21</u> 576 aa	
<b>Filo Bryozoa</b>					
<i>Bugula neritina</i>			LGALS8 <a href="#">KAF6039441</a> 477 aa		
<b>Filo Arthropoda</b>					
<i>Stegodyphus mimosarum</i>			Beta-galactoside-binding lectin <a href="#">KFM67265</a> 450 aa		
<i>Stegodyphus dumicola</i>			Galectin-9-like <a href="#">XP_03522114</a> <u>3</u> 462 aa		
<i>Centruroides sculpturatus</i>			Galectin-6-like <a href="#">XP_02321629</a> <u>9</u> 465 aa		
<i>Limulus polyphemus</i>			Galectin-6-like <a href="#">XP_01377319</a>		

			<a href="#">5</a> 464 aa		
<i>Rhipicephalus microplus</i>			Galectin-4-like <a href="#">XP_03727484</a> <a href="#">0</a> 490 aa		
<i>Anopheles gambiae</i>	AgGal6 <a href="#">XP_318003</a> 145 aa AgGal7 <a href="#">XP_318001</a> 143 aa				
<i>Anopheles stephensi</i>	AsGal <a href="#">AAO06842</a> 218 aa				
<i>Spodoptera litura</i>			Galectin-6-like <a href="#">XP_02281906</a> <a href="#">7</a> 517 aa		
<i>Spodoptera frugiperda</i>			32 kDa beta-galactoside-binding lectin-like <a href="#">XP_03545403</a> <a href="#">3</a> 447 aa		
<b>Filo Nematoda</b>					
<i>Globodera rostochiensis</i>	Galectin <a href="#">AAB61596</a> 235 aa				
<i>Caenorhabditis elegans</i>	LEC5 <a href="#">BAB11971</a>	LEC1 <a href="#">AAB87718</a>			

	314 aa  LEC6 <a href="#">BAA09794</a> 146 aa  LEC7 <a href="#">Q09605</a> 179 aa  LEC8 <a href="#">BAB11964</a> 180 aa  LEC9 <a href="#">BAB11965</a> 140 aa  LEC10 <a href="#">BAB11966</a> 192 aa  LEC11 <a href="#">BAB11967</a> 232 aa	279 aa  LEC4 <a href="#">BAB11970</a> 283 aa  LEC3 <a href="#">BAB11969</a> 297 aa			
<i>Necator americanus</i>		Galactoside-binding lectin <a href="#">XP_013308382</a> 328 aa			
<i>Toxocara canis</i>			Lec-3 <a href="#">KHN80316</a> 482 aa		
<i>Onchocerca volvulus</i>		Ov87 <a href="#">AAD00843</a> 321 aa			



<i>Haemonchus contortus</i>		HcoGal1 <a href="#">AAD11972</a> 278 aa			
		HcoGal3 <a href="#">AAB88823</a> 283 aa			
		HcoGal4 <a href="#">AAF63404</a> 291 aa			
<i>Brugia malayi</i>		BmGal <a href="#">AAF37721</a> 280 aa			
<i>Dirofilaria immitis</i>		DiGal <a href="#">AAF37720</a> 281 aa			
<b>Filo Annelida</b>					
<i>Helobdella robusta</i>		Hypothetical protein HELRODRA FT_191283 <a href="#">XP_009015119</a> 396 aa			
<b>Filo Porifera</b>					
<i>Geodia cydonium</i>	GcLT1 <a href="#">O18429</a> 190 aa				
<i>Amphimedon queenslandica</i>	AqGal <a href="#">A0A1X7ULM3</a> 220 aa				

Filo Echinodermata					
<i>Asterias rubens</i>		Galectin-8-like <a href="#">XP_0336477</a> <a href="#">79</a> 341 aa		<a href="#">XP_0336255</a> <a href="#">59</a> 645 aa	
<i>Apostichopus japonicus</i>			Galectin-8 <a href="#">QCW05467</a> 392 aa  AjGal <a href="#">PIK56913</a> 456 aa  Galectin-8-like <a href="#">PIK56914</a> 478 aa		
<i>Mesocentrotus nudus</i>		MnGal8 <a href="#">ASV40833</a> 312 aa			
<i>Paracentrotus lividus</i>		PIGal8 <a href="#">CBX45521</a> 311 aa			
Filo Chordata					
<i>Homo sapiens</i>	Galectin-1 <a href="#">NP_002296</a> 135 aa  Galectin-2 <a href="#">NP_006489</a> 132 aa  HSPC159	Galectin-12 <a href="#">NP_149092</a> 314 aa			

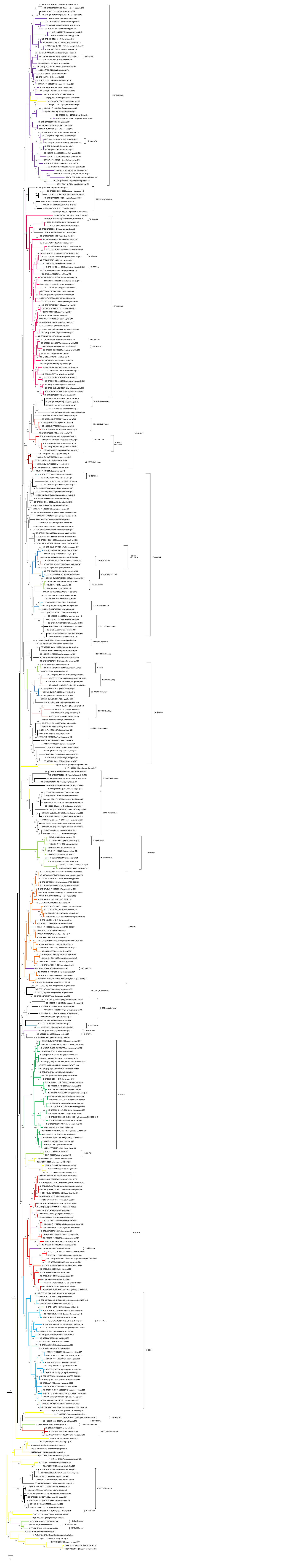
	<p><a href="#">NP_054900</a> 172 aa</p> <p>Galectin-13 <a href="#">NP_037400</a> 139 aa</p> <p>Placental protein 13-like <a href="#">NP_064514</a> 139 aa</p> <p>Galectin-10 <a href="#">NP_001819</a> 142 aa</p> <p>Galectin-7 <a href="#">NP_002298</a> 136 aa</p> <p>LGALS3 <a href="#">P17931</a> 250 aa</p>	<p>Galectin-4 <a href="#">NP_006140</a> 323 aa</p> <p>Galectin-8 <a href="#">NP_006490</a> 359 aa</p> <p>Galectin-9 <a href="#">NP_033665</a> 355 aa</p> <p>Galectin-12 <a href="#">NP_149092</a> 314 aa</p>			
<i>Rattus norvegicus</i>	<p>Grifin <a href="#">NP_476535</a> 144 aa</p> <p>Galectin-1 <a href="#">NP_063969</a> 135 aa</p> <p>Galectin-2 <a href="#">NP_598283</a> 130 aa</p>	<p>Galectin-12 <a href="#">NP_001099803</a> 314 aa</p> <p>Galectin-4 <a href="#">NP_037107</a> 324 aa</p>			

	<p>Galectin-3 <a href="#">NP_114020</a> 262 aa</p> <p>Galectin-7 <a href="#">NP_072104X</a> <a href="#">P_346543</a> 136 aa</p> <p>Galectin-5 <a href="#">NP_037108</a> 145 aa</p>	<p>Galectin-8 <a href="#">NP_446314</a> 316 aa</p> <p>Galectin-9 <a href="#">NP_037109</a> 322 aa</p> <p>Galectin-12 <a href="#">NP_0010998</a> <a href="#">03</a> 314 aa</p>			
<i>Mus musculus</i>	<p><a href="#">BAB32388</a> 144 aa</p> <p>Galectin-1 <a href="#">NP_032521</a> 135 aa</p> <p>Galectin-2 <a href="#">Q9CQW5</a> 130 aa</p> <p>Galectin-7 <a href="#">NP_032522</a> 136 aa</p>	<p>Galectin-12 <a href="#">NP_062389</a> 314 aa</p> <p>Galectin-4 <a href="#">AAC27245</a> 300 aa</p> <p>Galectin-8 <a href="#">NP_061374</a> 316 aa</p> <p>Galectin-12 <a href="#">NP_062389</a> 314 aa</p> <p>Lgals3 <a href="#">P16110</a> 264 aa</p> <p>Galectin-9 <a href="#">NP_034838</a> 353 aa</p>			
<i>Xenopus laevis</i>	Galectin-Ia	Gal-IVa	<a href="#">AAI06496</a>		

	<a href="#">BAB83247</a> 134 aa  Galectin-Ib <a href="#">BAB83256</a> 134 aa  Galectin-Ic <a href="#">AAA49896</a> 135 aa  Galectin-Vb <a href="#">BAC55884</a> 135 aa	<a href="#">BAB83259</a> 332 aa  GalVIa <a href="#">BAC55885</a> 319 aa  Gal-IIa <a href="#">BAB83257</a> 340 aa  GalVIIIa <a href="#">BAC55887</a> 315 aa  Gal-IIIa <a href="#">BAB83258</a> 343 aa	445 aa		
<i>Xenopus tropicalis</i>	Galectin-9 <a href="#">XP_031755300</a> 140 aa		Galectin-9 <a href="#">XP_012808599</a> 446 aa		
<i>Pantherophis guttatus</i>				<a href="#">XP_034294022</a> 829 aa	
<i>Bagarius yarrelli</i>				Galectin 6 <a href="#">TSL75311</a> 518 aa	
<i>Chanos chanos</i>			<a href="#">XP_030621082</a> 452 aa		
<i>Takifugu bimaculatus</i>			Hypothetical protein <a href="#">TNN01185</a> 452 aa		

<i>Takifugu rubripes</i>			<a href="#">XP_01160696</a> <u>3</u> <a href="#">452 aa</a>		
<i>Takifugu flavidus</i>			Galectin-4 <a href="#">TWW76851</a> 417 aa		
<i>Rhinatrema bivittatum</i>				Galectin-8 <a href="#">XP_0294496</a> <u>82</u> 647 aa	
<i>Anguilla anguilla</i>					<a href="#">XP_03524139</a> <u>5</u> 917 aa
<i>Salmo trutta</i>			Galectin-9-like <a href="#">XP_02957143</a> <u>3</u> 546 aa		
<i>Branchiostoma belcheri</i>		Galectin-9-like <a href="#">XP_0196245</a> <u>91</u> 311 aa			
<i>Branchiostoma floridae</i>		Galectin-4-like <a href="#">XP_0356874</a> <u>75</u> 313 aa			
<b>Filo Hemichordata</b>					
<i>Saccoglossus kowalevskii</i>		Galectin-9-like 342 aa			

		Galectin-8- like XP_00273158 9.1 348 aa			
--	--	---	--	--	--





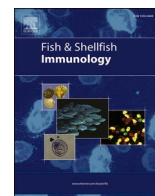
**Figura S1.** Análisis filogenético de galectinas de vertebrados e invertebrados. El árbol filogenético se construyó con el método de máxima verosimilitud en MEGA X utilizando un bootstrap de 500 réplicas. El árbol filogenético contiene los 448 CRD individuales analizados. En el clado 1 de vertebrados, se encuentran galectinas con 2 y 3 dominios. En el clado 2 de vertebrados, se encuentran galectinas con 2, 3 y 4 dominios. En el clado 3 de vertebrados, se encuentran galectinas con 2, 3, 4 y 6 dominios. En el clado 4 de vertebrados, las galectinas humanas monodominio se identifican en verde claro. En el caso de los invertebrados marinos: los clados de galectinas con tetradominios de moluscos se identificaron como CRD1 (azul), CRD2 (rojo), CRD3 (verde) y CRD4 (naranja). El clado con galectinas bidominio de moluscos está marcado en púrpura 2D-CRD1 y en rosa destaca el clado 2D-CRD2. Los clados de otros invertebrados marinos con tres dominios se denominaron Arthropoda y Echinodermata. Los clados de galectinas bidominio de otros invertebrados se denominaron 2D-CRD1 / Nematoda, 2D-CRD2 / Nematoda e invertebrados.

**ARTÍCULO PUBLICADO**



Contents lists available at ScienceDirect

## Fish and Shellfish Immunology

journal homepage: [www.elsevier.com/locate/fsi](http://www.elsevier.com/locate/fsi)

Full length article

Structural and evolutionary insights into the multidomain galectin from the red abalone *Haliotis rufescens* with specificity for sulfated glycans

Angélica Lizeth Toiber-Estrella<sup>a</sup>, Adrián Quintero-Martínez<sup>a</sup>, Adela Rodríguez-Romero<sup>a</sup>, Héctor Riveros-Rosas<sup>b</sup>, Alejandra Hernández-Santoyo<sup>a,\*</sup>

<sup>a</sup> Instituto de Química, Universidad Nacional Autónoma de México, Ciudad de México, Coyoacán, 04510, Mexico

<sup>b</sup> Depto. Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, Coyoacán, 04510, Mexico



## ARTICLE INFO

## Keywords:

Carbohydrate-binding protein

*Haliotis rufescens*

Innate immunity

Sulfated glycans specific lectin

Galectin evolution

Marine mollusks

Tandem-repeat galectin

## ABSTRACT

Galectins are an evolutionarily ancient family of lectins characterized by their affinity for  $\beta$ -galactosides and a conserved binding site in the carbohydrate recognition domain (CRD). These lectins are involved in multiple physiological functions, including the recognition of glycans on the surface of viruses and bacteria. This feature supports their role in innate immune responses in marine mollusks. Here, we identified and characterized a galectin, from the mollusk *Haliotis rufescens* (named HrGal), with four CRDs that belong to the tandem-repeat type. HrGal was purified by affinity chromatography in a galactose-agarose resin and exhibited a molecular mass of 64.11 kDa determined by MALDI-TOF mass spectrometry. The identity of HrGal was verified by sequencing, confirming that it is a 555 amino acid protein with a mass of 63.86 kDa. This protein corresponds to a galectin reported in GenBank with accession number AHX26603. HrGal is stable in the presence of urea, reducing agents, and ions such as  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$ . The recombinant galectin (rHrGal) was purified from inclusion bodies in the presence of these ions. A theoretical model obtained with the AlphaFold server exhibits four non-identical CRDs, with a  $\beta$  sandwich folding and the representative motifs for binding  $\beta$ -galactosides. This allows us to classify HrGal within the tandem repeat galectin family. On the basis of a phylogenetic analysis, we found that the mollusk sequences form a monophyletic group of tetradomain galectins unrelated to vertebrate galectins. HrGal showed specificity for galactosides and glucosides but only the sulfated sugars heparin and  $\iota$ -carrageenan inhibited its hemagglutinating activity with a minimum inhibitory concentration of 4 mM and  $6.25 \times 10^{-5}\%$  respectively. The position of the sulfate groups seemed crucial for binding, both by carrageenans and heparin.

## 1. Introduction

The innate immune response constitutes the first line of defense against infection by pathogens. Soluble and cell-associated lectins play a critical role as recognition/effector factors in the immune responses of both invertebrates and vertebrates. The binding of a lectin to multiple non-reducing terminal carbohydrate ligands on the microbial surface, which are not readily exposed in the host, leads to agglutination, immobilization, and possibly opsonization of the putative pathogen [1, 2]. Based on their structural folds and unique sequence motifs in CRDs, lectins have been classified within several families that differ in their structural and functional aspects and evolutionary history. Among 13 lectin families in animals, several types have been identified in mollusks,

such as: C-type, P-type, F-type, I-type, L-type, R-type, S-type (Galectins), Ficolins, and Chitinase like lectins [3]. However, new lectins isolated from these organisms have proved to be significantly different and are not related to previously known lectin families, such as *Mytilus californianus* lectin [4] and H-type lectins [5].

Among the lectin families in both invertebrate and vertebrate species, galectins are highly conserved through evolution. These lectins are characterized by recognizing a conserved sequence motif in their CRDs and an affinity for  $\beta$ -galactosides; furthermore, they share a common structural fold and at least one conserved carbohydrate recognition domain (CRD) of approximately 130 amino acids that mediates carbohydrate binding. A traditional classification based on structural similarities includes prototype galectins, which have one CRD and exist as

**Abbreviations:** CRD, Carbohydrate Recognition Domain; MCL, *Mytilus californianus* lectin; PSGL-1, P-selectin glycoprotein ligand-1; HrGal, *Haliotis rufescens* galectin; rHrGal, Recombinant *Haliotis rufescens* galectin.

\* Corresponding author.

E-mail address: [hersan@unam.mx](mailto:hersan@unam.mx) (A. Hernández-Santoyo).

<https://doi.org/10.1016/j.fsi.2022.11.015>

Received 29 June 2022; Received in revised form 8 November 2022; Accepted 10 November 2022

Available online 15 November 2022

1050-4648/© 2022 Elsevier Ltd. All rights reserved.

monomers or dimers; tandem repeat-type galectins, which contain two different CRDs separated by a linker of up to 70 amino acids; and chimera-type galectin, which includes a CRD connected to a non-lectin amino-terminal region. Most galectins are bivalent or multivalent regarding their carbohydrate binding activities, which enable recognizing multiple binding partners and the activation of distinct signaling pathways [6,7].

The specificity of some galectins for sulfated glycans has shown great relevance because they appear to be physiologically significant; different cell types and molecules display sulfated glycoconjugates and could modulate parasite-vector-host interactions. For example, a lectin targeting sulfated molecules may interfere with the function of PSGL-1 containing sialyl Lewis X (P-Selectin Glycoprotein Ligand 1), which mediates the interaction of neutrophils with lectin expressed by endothelial cells, and this interaction can interfere with inflammation [8,9]. In molluscs, sulfated polysaccharides such as heparin, heparan sulfate, chondroitin sulfate, dermatan sulfate and keratan sulfate are involved in growth, tissue construction, movement, defense, energy storage, environmental awareness, and self-regulation [10]. In the eastern oyster *Crassostrea virginica*, several N-glycans, including sulfated ones, have been detected in hemocytes as possible ligands for galectin (CvGal1). Contrary to their endogenous functions, glycan-binding proteins are often used by pathogens, which mimic their glycan ligands to hijack host immune systems. This is the case for the devastating parasite *Perkinsus* spp., which takes advantage of CvGal1 specificity [11,12].

Recently, galectins with four domains involved in the innate immune response against various pathogens, including bacteria, fungi, and viruses, have been found in mollusks [13–15]; however, their information is scarce.

To gain a deeper understanding of the properties of these proteins, here we describe the purification and biophysical and structural characterization of a four-domain galectin from red abalone *Haliotis rufescens* (named HrGal) obtained from both natural and recombinant sources.

## 2. Methods

### 2.1. Purification of *Haliotis rufescens* galectin (HrGal)

Red abalones (*Haliotis rufescens*) were obtained from Ensenada, Baja California, México. We homogenized the tissues with PBS buffer (50 mM sodium phosphate buffer, pH 7.4 with 150 mM NaCl) and the homogenate was centrifuged (10,000 g for 30 min) at 4 °C. The supernatant was then loaded onto a galactose-agarose column that had previously been equilibrated with the PBS buffer. The specifically bound fraction was eluted with PBS buffer containing 250 mM D-galactose. The purified protein was dialyzed against PBS using a membrane with a molecular weight cutoff of 3.5 kDa, and the homogeneity of the purified fraction was determined by polyacrylamide gel electrophoresis under denaturing conditions and mass spectrometry. The latter was performed using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (Microflex; Bruker Scientific LLC, Billerica, MA, USA). The matrix used was a saturated solution of sinapinic acid in 30% (v/v) aqueous acetonitrile, 0.1% (v/v) trifluoroacetic acid. Protein concentration was determined according to the BCA method [16], using bovine serum albumin as a protein standard.

### 2.2. Amino acid sequence determination

The amino acid sequence of the isolated lectin was determined by peptide identification using mass spectrometry. Briefly, the purified protein was loaded into a 12% SDS-PAGE stained with Coomassie blue R-250. The single band corresponding to the lectin was excised and faded with 200 µL of NH<sub>4</sub>HCO<sub>3</sub> solution (250 mM in 50% acetonitrile), then reduced with 50 mM TCEP (tris(2-carboxyethyl) phosphine), 10 min at 60 °C and Cys residues alkylated with 100 mM iodoacetamide, 1 h at 25 °C in the dark. The protein was then subjected to in-gel trypsin

digestion (Trypsin Singles, proteomic grade, Sigma-Aldrich) at 37 °C overnight. The peptides were mixed with an α-cyano-4-hydroxycinnamic matrix in 30% acetonitrile, 70% water and 0.1% trifluoroacetic acid and analyzed with a MALDI-TOF Microflex Bruker spectrometer. Bruker Peptide calibration standard II was used for calibrating. The spectra were recorded in positive-ion reflectron mode, and we used FlexAnalysis 3.0 software (Bruker Daltonics) for mass spectra recording and peak detection. The fragments were analyzed with the MASCOT 2.4 program.

### 2.3. Cloning, expression, and purification of recombinant *H. rufescens* galectin (rHrGal)

The gene sequence was synthesized with KpnI and EcoRI restriction sites at the 5' and 3' ends, respectively, by GenScript (Piscataway, New Jersey, USA) and cloned into the pET-32a vector with a His-tag and a TEV protease site at the N-terminus to produce the recombinant protein, which was named rHrGal. Once we verified the insert with restriction enzymes, the pET-32a-HrGal plasmid was transformed into *E. coli* strain BL21 (DE3) for its expression and subsequent purification. The transformed cells were grown in Luria-Bertani media with 100 µg/mL ampicillin at 37 °C with agitation at 180 rpm until an OD 600 of 0.8 was reached. IPTG was added to a final concentration of 0.25 mM and the temperature was lowered to 18 °C for 24 h. Cells were harvested by centrifugation at 15,000 g and suspended in the lysis buffer containing 30 mM Tris-HCl, pH 7.4, 200 mM NaCl, 6 M urea, and 1 mM PMSF (TBS-urea). The cell pellet was lysed by sonication (8 cycles of 10 s) and clarified by centrifuging at 30,000 g for 30 min. The supernatant was loaded onto HisTrap columns (GE Healthcare, Piscataway, NJ, USA) and equilibrated with TBS-urea and 20 mM imidazole. The column was washed with equilibration buffer and the proteins were eluted with 250 mM imidazole. The eluted protein was re-natured by slowly decreasing the urea concentration (6, 4, 2, 0.5 and 0 M in 50 mM TBS pH 7.4, 5 mM DTT, 150 mM NaCl, 10% glycerol) in dialysis for 24 h at 4 °C [17]. The purified fusion proteins were cleaved overnight at 4 °C, with TEV protease (50 mM Tris-HCl, pH 7.4, 200 mM NaCl, 5 mM MgCl<sub>2</sub> 0.25 mM DTT and 5% glycerol), with agitation, and the products were recirculated through the HisTrap column. Finally, we verified the cut in SDS-PAGE 14% and by MALDI-TOF mass spectrometry using a sinapinic acid matrix.

### 2.4. Assays of hemagglutination activity and substrate specificity

The hemagglutination activity of HrGal was tested by using rabbit red blood cells prepared according to the Fragkiadakis protocol [18] in 96-well microtiter U plates (NUNC™). After two-fold dilution in PBS, 50 µL of the protein was mixed with 50 µL of 2% (w/v) of rabbit formalinized erythrocytes. We incubated the mixture at room temperature for 90 min. The hemagglutination titer was reported to be the inverse of the last dilution, exhibiting an agglutinating activity.

The sugar specificity of lectin was tested by inhibiting the hemagglutination activity with different simple sugars, oligosaccharides, and polysaccharides (Table S1). The carbohydrates were two-fold serially diluted in microtiter U plates with PBS buffer. HrGal was incubated at room temperature for 1 h with different concentrations of inhibitors before adding 2% formalinized rabbit erythrocytes in suspension. Plates were incubated at room temperature. The results were expressed as the minimal concentration of carbohydrates that inhibited hemagglutination.

### 2.5. Differential scanning fluorimetry

The thermodynamic stability of HrGal was determined using fluorescent-based thermal shift assays with kits to screen pH, additives, and sugars. The assays were performed on Fast Optical 96-Well plates (Applied Biosystems) using 20 µL of the HrGal containing the

environmentally sensitive extrinsic dye SYPRO Orange (1:1000) (Invitrogen, Life Technologies, USA) and 20  $\mu$ l of kits. The plates were sealed (Greiner VIEWseal, Jena Bioscience), and fluorescence was recorded in a StepOne real-time PCR System (Applied Biosystems) while the temperature continuously increased from 25 to 95 °C [19]. For the analysis of additives, we used the Additive screen HT kit with 96 different conditions (Hampton Research Corp., USA). To analyze pH values, we used the Wizard pH buffer kit with 96 conditions (Rigaku Reagents Inc., USA). The quality of the curves was inspected using StepOne v2.3 software (Life Technologies, USA). The  $T_m$  values were determined by fitting the thermal melting curves to a sigmoidal Boltzmann function using Prism 9 (GraphPad Software Inc., USA). We further investigated the stabilizing conditions resulting from the latter experiments by performing activity assays (see Section 2.4).

## 2.6. Desulfation of heparin

To determine the importance of sulfate groups in the recognition of heparin by rHrGal, we desulfated heparin (Biosynth Ltd., UK) in two different manners. First, it was frozen dried in an alkaline solution, to discharge the sulfated group at position O-3 of GlcN<sub>2,3S</sub> and GlcN<sub>2,3,6S</sub>, as described by Santini et al. [20]. Second, its pyridinium salt was treated with dimethyl sulfoxide containing 5% of methanol for 1.5 h at 50 °C to be de-N-sulfated, following the methodology of Inoue and Nagasawa [21]. In both cases, the separation was performed by cation exchange, monitoring the absorbance at 215 nm. The desulfation of compounds was confirmed by ATR-FTIR analysis of dried samples using a Nicolet iS50 Spectrometer (Thermo Fisher Scientific, USA). Individual adaptive baselines were calculated for all spectra. All spectra were normalized and smoothed by applying a Savitsky-Golay filter of the fifth order. Finally, the spectra of desulfated compounds were scaled subtracted from the original compound spectra and analyzed. Frequency peaks exceeding 50% of visible spectrum ordinates and barycenters corresponding to  $\nu_{as}S = O$  bands were recorded.

## 2.7. Circular dichroism

Circular dichroism (CD) spectra were recorded at 25 °C on a JASCO J-1500 spectropolarimeter (JASCO Inc., Easton, MD, USA) equipped with a Peltier temperature control system. Both HrGal and rHrGal were dialyzed against a 10 mM MES-citrate buffer, 10 mM NaCl at pH 7.4, and filtered using 0.22  $\mu$ m pore size Millipore membrane filters. Far-UV spectra were collected from 190 to 260 nm in a 0.1 cm pathlength cell. Three scans were averaged to obtain the final protein spectra at a concentration of 0.25 mg/mL. The spectra were scaled to molar ellipticities  $[\theta]$  ( $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}$ ). The secondary structure content of the spectra were determined by using the BeStSel (Beta Structure Selection) online server [22].

## 2.8. Structural modeling of HrGal and molecular docking with *l*-carrageenan

The 3D structure of HrGal was obtained using the AlphaFold artificial intelligence program [23]. Once the structure was built, we used the Coot program version 0.8 [24] to correct geometrical issues. Finally, the overall stereochemical quality of the model was assessed using the PROCHECK program [25]. Docking simulations were performed with the HDock server [26] using *l*-carrageenan as the ligand. We identified and analyzed residues potentially involved in ligand binding using CHIMERA [27].

## 2.9. Phylogenetic analysis of HrGal

The amino acid sequences of the tetradomain galectins were recovered by BLASTP searches (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) from the non-redundant protein database at the NCBI site, using the

deposited galectin sequence from *H. rufescens* (AHX26603) as bait. The analysis included TBLASTN searches against cDNA from 10 species of mollusks with a complete sequenced genome to retrieve all putative galectins. Only sequences with an E value lower than  $10^{-5}$  were included. Furthermore, individual CRDs of galectin sequences from 66 invertebrates and 16 vertebrates, including those from human, rat, mouse, and *Xenopus laevis* whose complete genomes have been reported, were aligned. A total of 448 CRDs of 209 galectins were considered for the study. If a retrieved amino acid sequence was incomplete or showed atypical insertions or deletions, the genomic or cDNA sequence for this protein was retrieved and the gene prediction software, Softberry FGENESH+ [28] (<http://www.softberry.com/>) was used to predict the intron/exon gene structure, considering protein homology information from the more similar known complete galectin sequence. The selected species and sequence accession numbers are listed in the supplemental material (Table S3 and Table S4). All amino acid alignments were performed with ClustalX [29] and manually corrected using BIOEDIT [30]. We identified the evolutionary relationship in full-sequence galectins by constructing a rootless phylogenetic tree using the maximum likelihood method, with the Le-Gascuel model, and in the case of individual domains, the Whelan and Goldman model using the MEGA version X program [31]. Branching reliability was estimated using a 500-replica bootstrap.

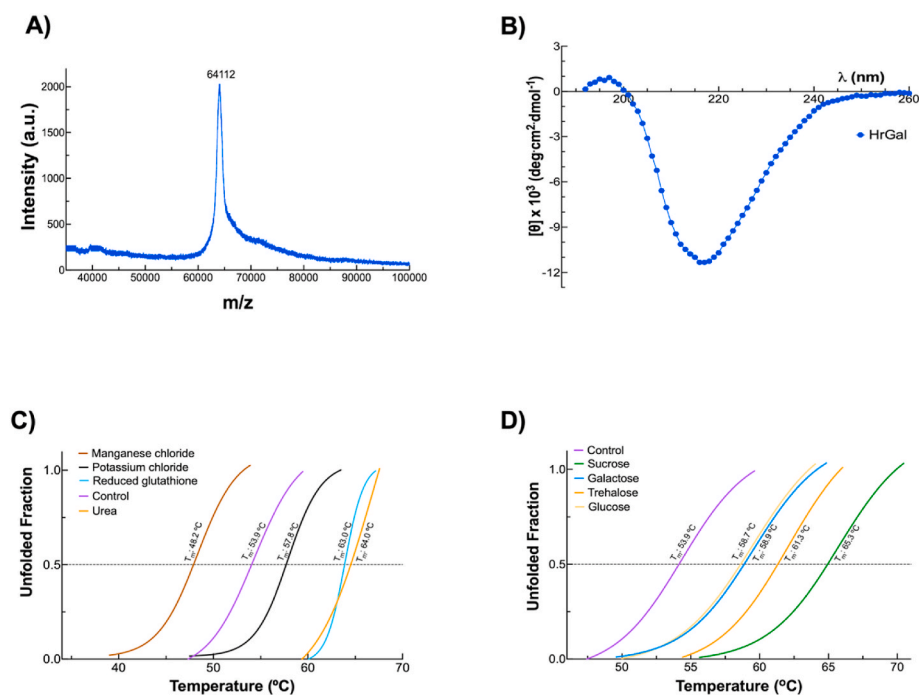
## 3. Results and discussion

### 3.1. Purification, hemagglutinating activity and carbohydrate specificity of *Haliotis rufescens* galectin

Native HrGal purification was achieved by affinity chromatography in a galactose-agarose resin. The purified protein showed a molecular mass of 64.11 kDa as determined by MALDI-TOF mass spectrometry (Fig. 1A) and a specific activity of 1600 U/mg on rabbit erythrocytes. To verify whether it was a calcium-dependent lectin, we passed it through a chelating resin, and it maintained its activity. Therefore, calcium is not essential for lectin activity. However, activity increased four-fold (6,600 U/mg) in the presence of 1 mM CaCl<sub>2</sub>. In some cases, binding of Ca<sup>2+</sup> ions is known to lead to conformational changes that modulate the activity of the target protein. In some extracellular enzymes, the occupation of calcium-binding sites involving surface loops leads to improved protein stability and provides protection against proteolytic digestion [32].

Furthermore, we also performed a carbohydrate binding specificity analysis (Table S1) observing that only heparin and *l*-carrageenan inhibited the hemagglutinating activity of HrGal with a minimum inhibitory concentration of 4 mM for heparin (average MW = 6 kDa) and 6.25 X  $10^{-5}$ g for *l*-carrageenan. It is noteworthy that HrGal has other potential ligands that do not inhibit hemagglutination. The capacity to agglutinate erythrocytes is given by the availability of at least two binding sites. HrGal possesses four nonidentical binding sites with potential different specificities, which means that even if two of them are saturated it might still agglutinate cells. Heparin is an anionic sulfated glycosaminoglycan composed of 1–4 linked repeating disaccharide units comprising a uronic acid (*D*-glucuronic acid or *L*-iduronic acid) and *D*-glucosamine with sulfation at various positions [33]. Carrageenans are sulfated linear polysaccharides of *D*-galactose and 3,6-anhydro-*D*-galactose. The three most essential carrageenans are Kappa ( $\kappa$ ), Iota ( $\iota$ ), and Lambda ( $\lambda$ )-carrageenan; of these, the least sulfated is the first and the most sulfated is the last [34]. Sulfated polysaccharides have been widely recognized for their wide range of biological activities, including anticoagulant properties, and the positions of the sulfate groups are related to their biological activity [34]. Interestingly, kappa and lambda carrageenan showed no inhibitory effects on the HrGal hemagglutinating activity.

Galectins with specificity for sulfated glycans have been reported, such as human galectins 1, 3, and 8. The latter maintains the canonical



**Fig. 1.** Purification of *Haliotis rufescens* galectin (HrGal). A) Matrix-assisted laser desorption ionization time-of-flight mass (MALDI-TOF) spectrum, B) CD spectrum in the far UV-region. Effect of several conditions on the thermal stability of HrGal; C) Effect of different additives. Urea, KCl, and MnCl<sub>2</sub> (50 mM), while reduced glutathione was 5 mM, D) Effect of sugars. They were tested at 125 mM. The melting point values (T<sub>m</sub>) were calculated at the inflection point of the transition temperatures using the Boltzmann sigmoidal fit. Control was prepared in 50 mM sodium phosphate buffer, pH 7.4.

amino acids involved in sugar recognition. This specificity is primarily attributed to N-terminal CRD, where Arg<sup>45</sup>, Gln<sup>47</sup>, Arg<sup>59</sup> (human galectin-8 numbering), and the extended loop region between the S3 and S4 strands interact directly or indirectly with the sulfate contributing to the strong binding of galectin to sulfated oligosaccharides [35, 36].

CD measurements in the far UV region (190 nm - 260 nm) revealed that HrGal has a high content of  $\beta$ -strands (15.9% antiparallel and 23.0% parallel), 12.2% of helices and 3.6% of turns (Fig. 1B). This content agrees with the  $\beta$ -sandwich folding observed in galectins from different sources [7,8].

### 3.2. Differential scanning fluorimetry

During the purification process of HrGal, we observed that it lost its activity after very short periods, so we decided to carry out stability analyses. Only three of the 96 buffers tested in a thermal shift assay gave sigmoidal curves. Interestingly, with CAPSO 50 mM pH 9.0, the T<sub>m</sub> decreased slightly (Fig. S1), but the specific activity increased twofold (13,300 U/mg).

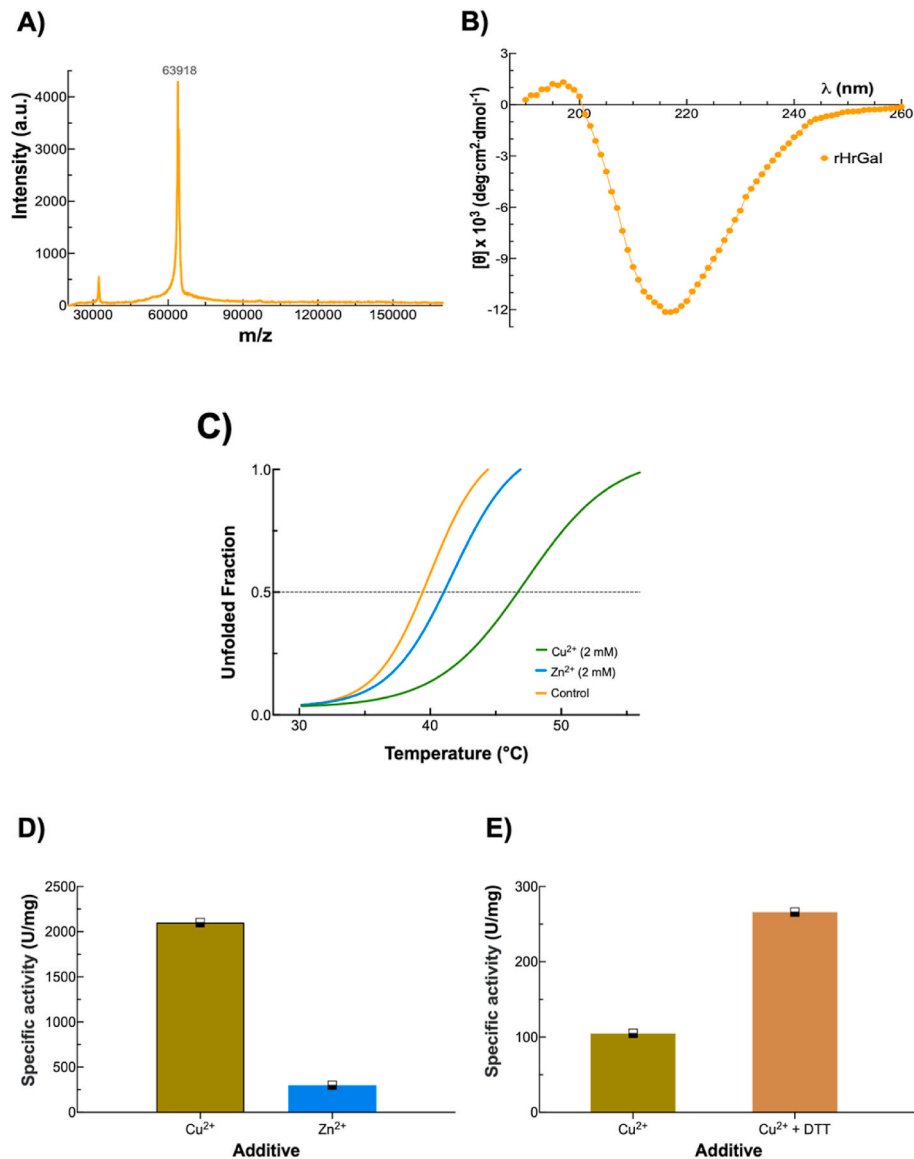
The thermal stability of the HrGal was favored not only by the buffer but also by the presence of some ions, urea, and reducing agents (Fig. 1C). Interestingly, even in the presence of 2 M urea, the lectin maintained its hemmagglutinating activity (13,300 U/mg). Considering the positive effect of glutathione on the stability of HrGal, we tested another reducing agent, TCEP, and a cysteine derivatizing agent that can reversibly transform molecules containing thiol groups into sulfenylate (MMTS, S-methyl methanethiosulfonate). The former increased the activity four times (26,600 U/mg), while the latter increased it eight times (53,000 U/mg). These results suggest the presence of free cysteines in the protein that may cause protein destabilization.

In particular, some glucosides such as sucrose and trehalose, favored HrGal stability (Fig. 1D). The four domains of HrGal have different amino acid sequences, and this characteristic may allow the recognition of these non-reducing disaccharides, increasing their stability. We have observed this behavior in lectins such as MCL from the mussel *M. californianus*. This lectin has three recognition domains, one of them being less specific. Although its preference is for galactosides, this domain was able to bind glucosides in the absence of galactose,

stabilizing the protein [4]. Moreover, disaccharides such as trehalose and sucrose are often used as stabilizing agents; however, the mechanism of how they achieve this effect has been the subject of constant research from different aspects, such as structural and dynamical [37].

### 3.3. Amino acid sequence analysis and characterization of the recombinant lectin

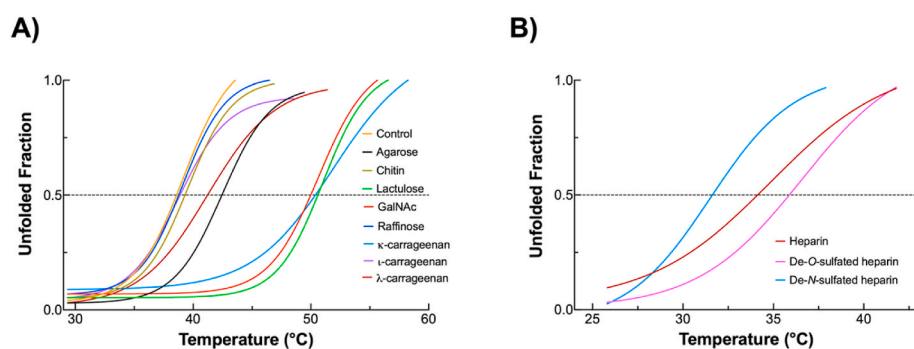
To confirm the HrGal identity, we subjected the protein to sequence analysis using tandem MS/MS of the peptides generated by trypsin digestion. The results revealed a 555 amino acid polypeptide chain with a calculated mass of 63,861 Da and a theoretical pI of 4.73. The sequence corresponded to a previously reported galectin (GenBank code AHX26603) with 100% identity [15]. The difference in mass with that obtained experimentally for the protein purified from the organism may be due to some post-translational modifications that are very common in these organisms, mainly acetylations or glycosylations. With this sequence, we could clone and express the recombinant protein (rHrGal). However, the heterologous expression of HrGal in *E. coli* resulted in the formation of insoluble inclusion bodies. After cell lysis and removal of the cell wall and outer membrane components, the aggregates were solubilized with a strong protein denaturant. We determined that the minimum concentration of urea required to solubilize the protein was 6 M (TBS-Urea 6 M). The protein was then purified using a Ni<sup>2+</sup> affinity column eluting with 250 mM imidazole. To refold rHrGal, we used a urea gradient, as explained above. In the absence of urea, we observed that the protein lost its activity, which recovered when 500 mM imidazole was added. However, high concentrations of imidazole interfere with the techniques used in the characterization of the lectin, and it was necessary to find the ideal conditions to maintain its activity. We first verified that the recombinant protein had the expected mass and was properly folded. As shown in Fig. 2A, the protein matches the theoretical value (63,918 Da) and the CD spectrum is very similar to the spectrum of the endogenous protein (Fig. 2B). Later, we tested the addition of additives and ions in a thermal shift assay and observed that copper and zinc increased the stability of rHrGal (Fig. 2C). In fact, copper increased the stability of the galectin by 7 °C. Under denaturing conditions and in the presence of Cu<sup>2+</sup> and Zn<sup>2+</sup> (2 mM), the lectin maintained its activity (Fig. 2D). However, only in the presence of Cu<sup>2+</sup> and DTT (10 mM) the



**Fig. 2.** Purification and characterization of rHrGal. A) MALDI-TOF spectrum, B) CD spectra in the far UV-region of the recombinant *Haliotis rufescens* galectin; C) Effect of the addition of  $\text{Cu}^{2+}$  and  $\text{Zn}^{2+}$  on the thermal stability of the galectin, D) Activity of unfolded rHrGal in the presence of copper or zinc (2 mM); E) Activity of folded galectin in the presence of DTT(1 mM) and copper (2 mM).

galectin remains active without urea (Fig. 2E) but only for a short time. Finally, we also tested the denatured protein (6 M Urea) in the presence of  $\text{Cu}^{2+}$  (2 mM) and galactose (150 mM). After incubation for 1 h, we

dialyzed the protein against TBS buffer (10 mM Tris-HCl, pH 7.4; 150 mM NaCl, 1 mM DTT) and finally, we obtained a pure and stable rHrGal with a specific activity of 1,650 U/mg. The ions could be essential



**Fig. 3.** Recognition of sugars by rHrGal analyzed by differential scanning fluorimetry A) Effect of different galactosides B) effect of sulfate groups on heparin recognition. The melting point values ( $T_m$ ) were calculated at the inflection point of the transition temperatures using the Boltzmann sigmoidal fit.

because they stabilize the groove where ligand recognition occurs by interacting with residues connecting the S3 and S4 strands. This region is crucial in the recognition of sulfated sugars, as observed in galectins 8 and 4 [36,38].

It is interesting to note that the erythroagglutinating activity of rHrGal on rabbit erythrocytes was inhibited only by heparin and ι-carrageenan, as observed with HrGal. Although still widely used, this technique is often not very sensitive to detect lectin-sugar interactions, so we performed thermal shift assays with all the sugars tested in erythroagglutination assays (Table S1) and observed that in addition to recognizing the glycans already mentioned (Fig. 1D), it also recognized other galactosides such as agarose, chitin, lactulose, GalNAc, raffinose and not only ι-carrageenan but also κ and λ (Fig. 3A).

Neither the highly sulfated λ-carrageenan (3 sulfate groups) nor the low sulfated κ-carrageenan (1 sulfate group) were able to inhibit hemagglutination, contrary to ι-carrageenan (2 sulfate groups), leading us to consider that sulfate groups are important for recognition of HrGal and rHrGal. We tested this phenomenon by differential scanning fluorimetry, resulting in the best stabilization of rHrGal by κ-carrageenan, followed by λ-carrageenan and then ι-carrageenan (Fig. 3A). These results suggest that the degree of sulfation is not the most important variable but probably the group position.

We tested the importance of the position of the sulfate groups by chemically modifying heparin in two different ways, de-N-sulfation and de-O-sulfation (Fig. S2 and Table S2). Heparin is a glycosaminoglycan of which the most common disaccharide unit is composed of a O-sulfated iduronic acid and a glucosamine that is both O-sulfated and N-sulfated. As for carrageenan, the desulfation of heparin modified the way galectin

bound to it. As we observed by isothermal analysis (Fig. S3), the galectin behavior suggested a higher affinity for the de-O-sulfated version, followed by the de-N-sulfated version. We ignore whether this behavior can be associated with the same binding site or other phenomena, given the multidomain nature of rHrGal. Thermal shift analysis showed that at certain concentrations, while de-O-sulfated heparin stabilizes the protein, de-N-sulfated heparin has the opposite effect (Fig. 3B). Further analysis of this effect, including individual domain studies, is needed to be fully understood. Since sulfated glycans have important biological roles in mollusks, their binding by HrGal might play a role in the recognition of self-glycans or in the abalone immune system.

### 3.4. Structural analysis

A sequence alignment of the four HrGal nonidentical carbohydrate-binding domains (CRDs) of HrGal is shown in Fig. 4. The four nonidentical CRDs consisting of 144, 135, 138, and 138 amino acids share an identity from 29.23 to 38.46%. Canonical residues involved in ligand recognition are marked with a green star, and most of them are conserved. In addition, the theoretical model obtained with the AlphaFold server shows that each domain exhibits a typical β sandwich folding of 5 stranded β-sheets that possess the representative motifs to bind β-galactosides; only CRD 1 (N-terminal) has an additional loop and a beta strand (Fig. 5A). Interestingly, each HrGal CRD showed an adequate distance between two cysteines to form a disulfide bond, except for domain 3, which has only one free cysteine (Table 1). Additionally, two Cys were at a suitable distance to form an interdomain disulfide bond between Domain 2 (Cys<sup>272</sup>) and Domain 4 (Cys<sup>512</sup>).

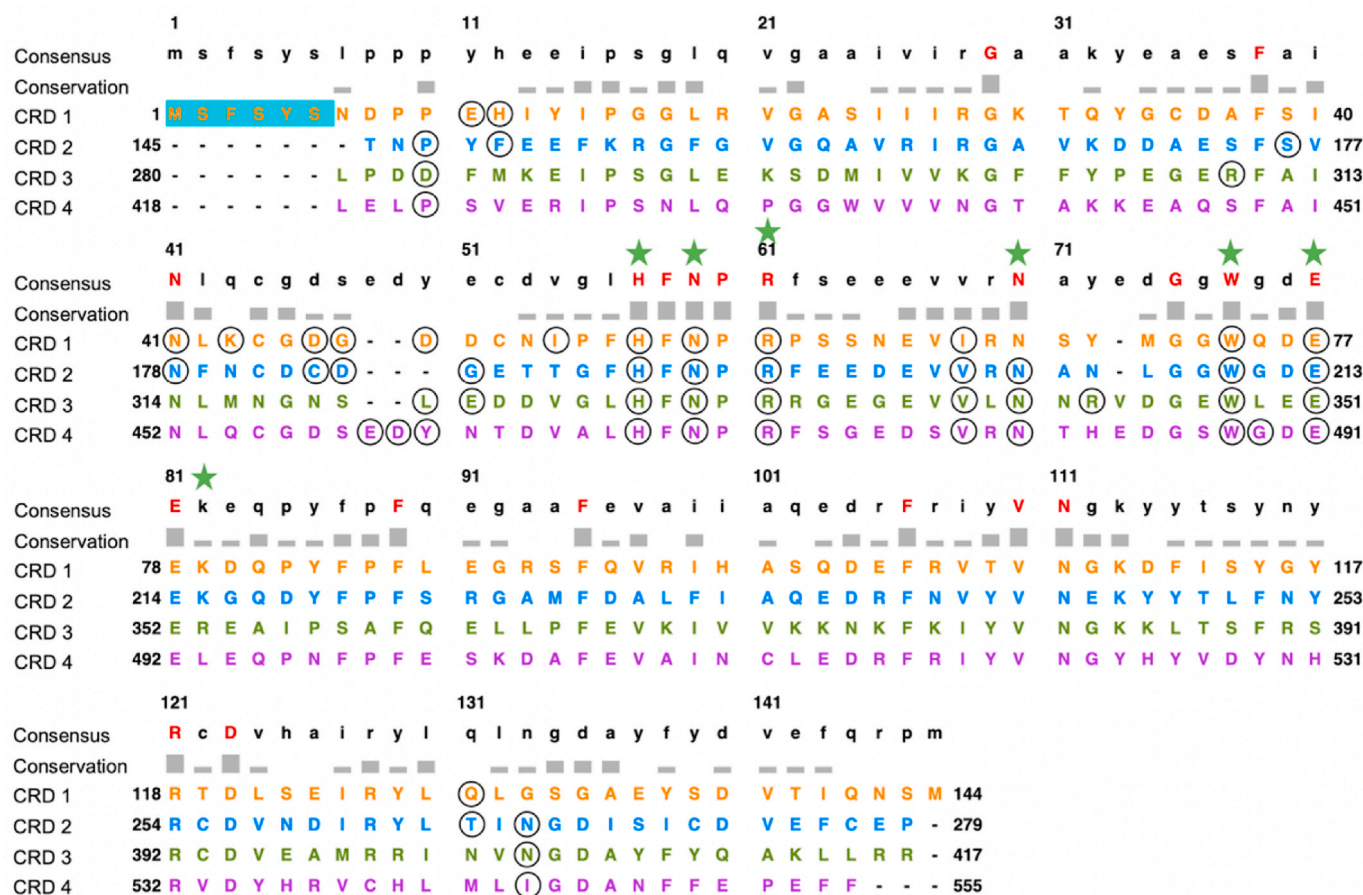
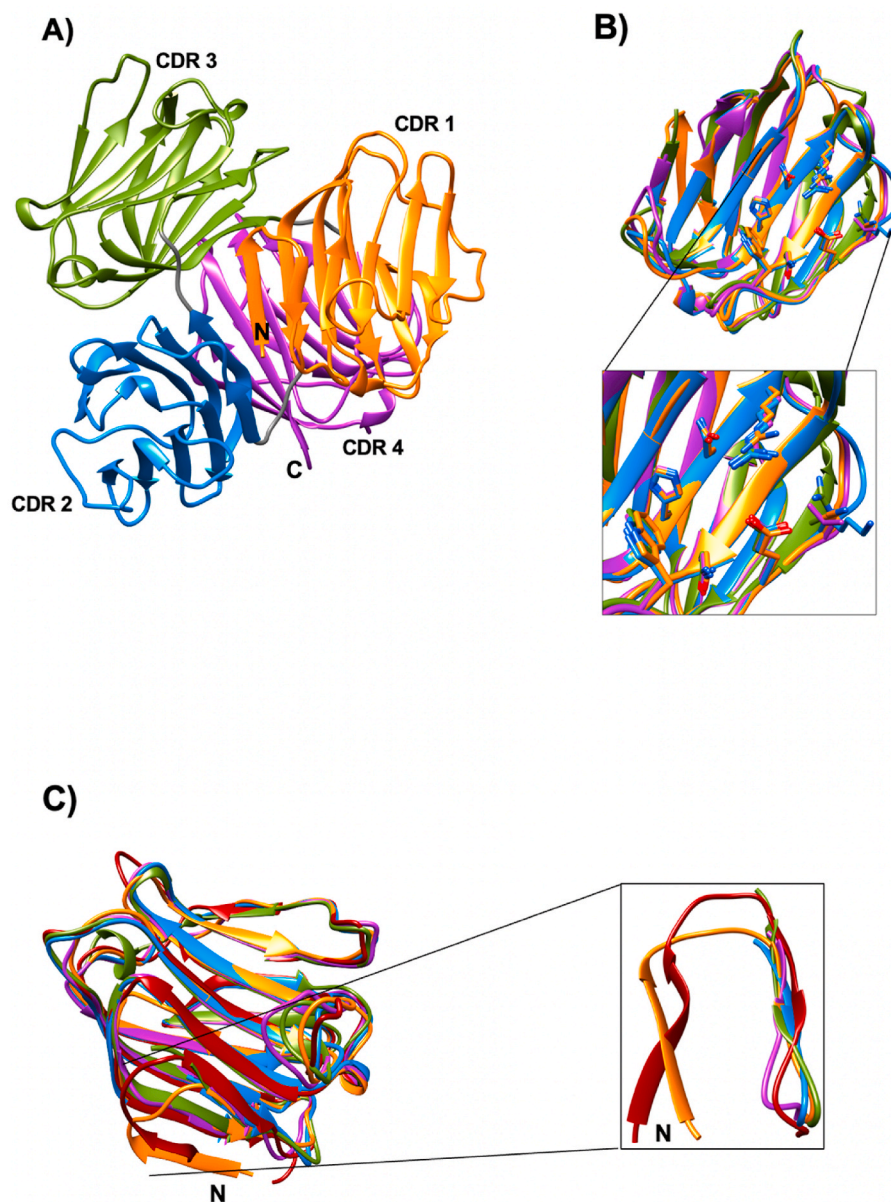


Fig. 4. Sequence alignment of the four nonidentical carbohydrate-binding domains (CRDs) of HrGal. The conserved amino acids are shown in the consensus sequence in red. The canonical residues that interact with the ligand, present in all galectins, are marked with a green star. The additional loop, considered responsible for recognizing sulfated sugars, is shown in cyan. Putative residues interacting with ι-carrageenan are circled (see section 3.5).





**Fig. 5.** The theoretical model of HrGal constructed with the AlphaFold server. A) Ribbon diagram showing the four domains in tandem, B) HrGal domains overlap, each possessing a  $\beta$ -sandwich fold. The conserved amino acids engaged in sugar binding activity are shown in sticks, C) the N-terminal domain of human galectin 8 (red) overlapped with the different domains of HrGal. The close-up shows the additional residues at the N-terminal region in both galectins.

**Table 1**

Free cysteines and disulfide bond in each CRD.

	Disulfide bond	Free Cys
CRD 1	Cys <sup>44</sup> —Cys <sup>50</sup>	Cys <sup>35</sup>
CRD 2	Cys <sup>181</sup> —Cys <sup>255</sup>	Cys <sup>183</sup> and Cys <sup>277</sup>
CRD 3	—	Cys <sup>393</sup>
CRD 4	Cys <sup>539</sup> —Cys <sup>455</sup>	

It is possible that these free cysteines are responsible for the instability of the lectin and that, as we observed in the differential scanning fluorimetry studies (Fig. 1C), the presence of reducing agents is necessary to keep it stable.

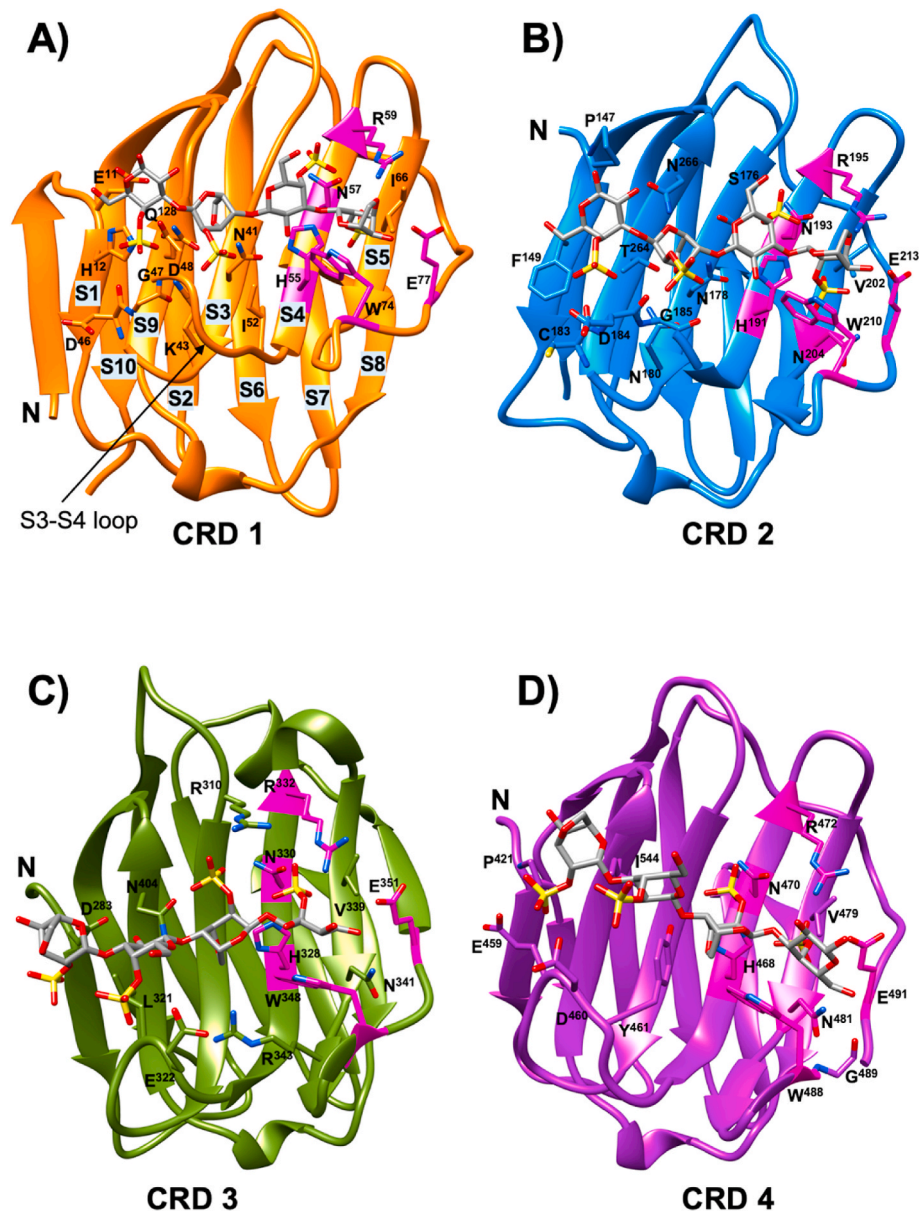
Although the domains are not identical, the residues involved in the recognition are conserved as in all galectins. Fig. 5B shows a superposition of the four CRDs, with the residues responsible for the interaction with the ligand. The main variation is observed in the last residue

of the recognition site Lys<sup>79</sup>, Lys<sup>215</sup>, Arg<sup>353</sup> and Leu<sup>493</sup> in domains 1 to 4, respectively.

Interestingly, the additional loop and a beta strand (Fig. 5C) have also been observed in human galectin 8. The latter has two tandem repeat CRDs, which share 38% amino acid identity [35] and have a higher affinity for 3'-O- sulfated or 3'-O-sialylated lactose. This specificity was primarily attributed to N-terminal CRD (N-domain), in which Gln<sup>47</sup> plays an essential role in the binding sulfated glycans [36].

### 3.5. Molecular docking with *ι*-carrageenan

To further investigate how HrGal recognizes *ι*-carrageenan, we carried out a docking simulation with each CRD. CRD1 showed 15 amino acids interacting with this ligand (Fig. 6A). In contrast, CRD2 showed 16 (Fig. 6B), and in CRDs 3 and 4, the ligand interacts with 13 residues (Fig. 6C and 6D). The hydrogen bonds between CRD and *ι*-carrageenan are 35, 32, 25 and 21 in CRD1 to CRD4, respectively. These variations at



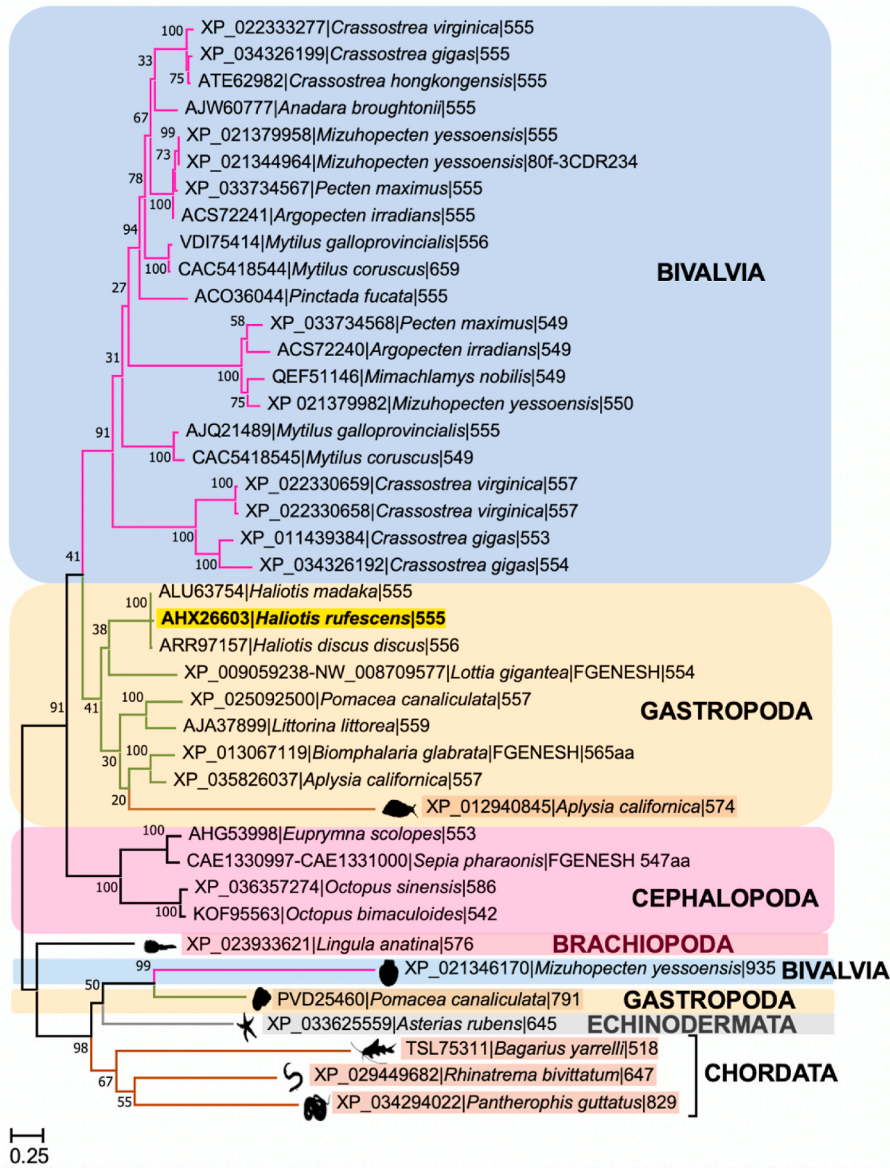
**Fig. 6.** Structure of the galectin binding site in complex with  $\iota$ -carrageenan. Detail of the binding cleft, indicating the amino acid residues that interact with the sulfated glycoside. A), B), C) and D) correspond to CRDs 1, 2, 3, and 4, respectively. The canonical amino acids involved in the interaction of galectins with their ligand are shown in magenta.

the amino acids present in the recognition site and the different interactions that they favor may explain the differences in specificity that each domain should have, as noted in galectin 8, where CRD1 (N-terminal) was responsible for the recognition of the sulfated sugar while the second domain showed specificity for other sugars. The alignment of the different HrGal domains shown in Fig. 4 shows the residues involved in the interaction with the ligand (circled) and the additional loop, which is considered responsible for the recognition of sulfated sugars (cyan). In all CRDs, most canonical amino acids interact with the ligand (Fig. 6 in magenta). Furthermore, some CRD residues are essential to recognize  $\iota$ -carrageenan, such as Glu<sup>11</sup>, His<sup>12</sup>, Asn<sup>41</sup>, Asp<sup>48</sup>, and Gln<sup>128</sup> of CRD1 and the equivalents of domains 2–4.

### 3.6. Phylogenetic analysis of HrGal

Forty-one tetradomain galectin sequences homologous to *Haliotis rufescens* (AHX26603) were retrieved from the NCBI nonredundant

protein database using the BlastP tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Of these sequences, 36 correspond to different molluscan species (bivalves, gastropods, and cephalopods), three belong to chordates, *Pantherophis guttatus* (XP\_034294022), *Bagarius yarrelli* (TSL75311), *Rhinatrema bivittatum* (XP\_029449682), one to the echinoderm *Asterias rubens* (XP\_033625559), and one to the brachiopod *Lingula anatina* (XP\_023933621). Fig. 7 shows a phylogenetic tree constructed from a multiple alignment with the retrieved sequences. All molluscan sequences (except for two) are grouped on the same branch within the tree, suggesting that this is a monophyletic group of tetradomain galectins. It is interesting to observe that the phylogenetic tree topology obtained with tetradomain galectins (Fig. 7), is very similar to the time tree of taxa (species included in Fig. 7), obtained according to the divergence times recovered mainly from TimeTree 5 resource (Fig. S4). Only class bivalvia exhibit a more complex topology because the ancestral tetradomain galectin gene suffered several duplications in this Mollusca class. Indeed, considering that ten Mollusca genomes have



**Fig. 7.** Phylogenetic analysis of tetradomain galectins. Forty-one complete galectin tetradomain sequences are included in the phylogenetic tree. Sequence names are indicated by the GenBank accession number, followed by species and sequence length. The bivalve clades are shaded in blue, the gastropod clade in yellow, and the cephalopod clade in pink. Sequence of *H. rufescens* is highlighted in yellow; sequences illustrated with a silhouette do not belong to the monophyletic tetradomain galectin family of Mollusca (see Supplementary Material; silhouettes were retrieved from PHYLOPIC; <http://phylopic.org/>).

been completely sequenced, it is interesting to observe that only one tetradomain galectin gene can be found in cephalopoda (*Octopus bimaculoides* and *Octopus sinensis*) and gastropoda (*Biomphalaria glabrata*, *Lottia gigantea* and *Pomacea canaliculata*); only *Aplysia californica* has two tetradomain galectin genes, but one of them is divergent (Fig. 7 and Fig. S4). On the contrary, bivalvia possess two (*Pecten maximus*), three (*Crassostrea virginica* and *Crassostrea gigas*) or even four (*Mizuhopecten yessoensis*) tetradomain galectin genes.

To obtain additional evidence for the possible origin of tetradomain galectins in mollusks, the individual domains of each galectin were aligned with each other (Fig. S5). For this last analysis, we also retrieved from the NCBI non-redundant protein database all galectin domains found in mollusks and the individual galectin domains present in the 100 sequences with the highest identity to galectin from *Haliotis rufescens* (AHX26603). For reference, all individual galectin domains identified by Dr. Denis Houzelstein's group (2004) [39], present in humans, rats, and mice, and those of *Xenopus laevis* and *X. tropicalis*, were also included. Thus, an alignment with 448 individual galectin domains belonging to animals was obtained, comprising monodomain, bidomain, tridomain or tetradomain galectins. Fig. S5 shows the phylogenetic tree constructed from the individual domains. It is interesting to note that, in

this tree, each of the individual domains of the Mollusca tetradomain galectins (named CRD1 to CRD4) are also grouped into independent branches: CRD1 (blue), CRD2 (red), CRD3 (green), and CRD4 (orange), strongly supporting that the four domains of the molluscan tetradomain galectins comprise a monophyletic group. The tetradomain galectins that are not part of this monophyletic group in Mollusca are indicated in Fig. 7, with a small image illustrating the silhouette of the species to which the sequence belongs. In Fig. S5, these same silhouettes show the location of these divergent sequences, where we can observe that they are in separate branches, showing that their origin is independent of the rest of the tetradomain galectins of Mollusca. On the other hand, since the *Lingula anatina* genome was completely sequenced previously and the *L. anatina* 4-CRD galectin found in this work does not belong to the Mollusca 4-CRD galectin family. This suggests that the Mollusca 4-CRD galectin family originated after Brachiopoda-Mollusca divergence.

In the phylogenetic tree of the supplementary material (Fig. S6), each of the 448 sequences studied can be observed in more detail. It is interesting to note in this figure that the bidomain galectins of mollusks (2D-CRD1/mollusk, 2D-CRD2/mollusk) and nematodes (2D-CRD1/Nematoda and 2D-CRD2/Nematoda) formed separate monophyletic clades suggesting an independent origin in each of their respective

clades.

#### 4. Conclusions

We purified a galectin with four tandem domains and a molecular mass of 64 kDa from the red abalone *Haliotis rufescens* (HrGal) and established the optimal conditions to maintain its stability. Since this galectin has at least one free cysteine per domain, it requires the presence of reducing agents such as DTT to maintain its activity. This galectin showed specificity for galactosides and glucosides, but only the sulfated sugars heparin and  $\iota$ -carrageenan inhibited its hemagglutinating activity. The important roles and abundance of sulfated glycans in mollusks point to a possible vital role of HrGal in either self or pathogen recognition. The position of the sulfate groups seemed crucial for binding, both by carrageenans and heparin, but further studies are required to explain this behavior.

HrGal exhibits a terminal domain with an additional loop, such as in human galectin 8, that is considered responsible for the recognition of sulfated sugars. Variations in the different domains can lead to changes in specificity for other ligands. In CRD1 (a domain that is believed to be responsible for recognizing sulfated sugars), the amino acids responsible for  $\iota$ -carrageenan recognition, in addition to canonical ones, are Glu<sup>11</sup>, His<sup>12</sup>, Asn<sup>41</sup>, Asp<sup>48</sup>, and Gln<sup>128</sup> of CRD1 and the equivalents of domains 2–4.

Based on the phylogenetic analysis, we found that in the different taxa, other domains, duplications, and fusions occurred that gave rise to proteins with more than one domain. The distinct 2-, 3-, and 4-domain galectins of mollusks probably originated independently of the multi-domain galectins found in vertebrates. We note that the mollusk sequences comprise a monophyletic group of tetradomain galectins that are not related to vertebrate galectins.

#### CRedit authorship contribution statement

**Angélica Lizeth Toiber-Estrella:** Investigation, Methodology. **Adrián Quintero-Martínez:** Methodology, Investigation, Writing – review & editing. **Adela Rodríguez-Romero:** Investigation, Formal analysis, Writing – review & editing. **Héctor Riveros-Rosas:** Investigation, Formal analysis, Writing – review & editing. **Alejandra Hernández-Santoyo:** Conceptualization, Supervision, Writing – original draft, Formal analysis, Writing – review & editing.

#### Declaration of competing interest

The authors declare that they have no conflict of interest.

#### Data availability

No data was used for the research described in the article.

#### Acknowledgements

This work was supported by DGAPA, UNAM grants PAPIIT-IN213022 (AHS) and PAPIIT-IN219022 (HRR). Angélica Lizeth Toiber Estrella is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) who received fellowship 492615 from CONACyT.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fsi.2022.11.015>.

#### References

- [1] N. Sharon, Lectins: past, present and future, *Biochem. Soc. Trans.* 36 (2008) 1457–1460, <https://doi.org/10.1042/BST0361457>.
- [2] K. Kubelkova, A. Macela, Innate immune recognition: an issue more complex than expected, *Front. Cell. Infect. Microbiol.* 9 (2019) 241, <https://doi.org/10.3389/fcimb.2019.00241>.
- [3] M. Chellappaakialakshmi, C. Ravi, Investigation on mollusc lectins, in: *Aquat. Lectins*, Springer Nature Singapore, Singapore, 2022, pp. 81–95, [https://doi.org/10.1007/978-981-19-0432-5\\_5](https://doi.org/10.1007/978-981-19-0432-5_5).
- [4] E. García-Maldonado, P. Cano-Sánchez, A. Hernández-Santoyo, Molecular and functional characterization of a glycosylated Galactose-Binding lectin from *Mytilus californianus*, *Fish Shellfish Immunol.* 66 (2017) 564–574, <https://doi.org/10.1016/j.fsi.2017.05.057>.
- [5] A.J. Pietrzyk-Brzezinska, A. Bujacz, H-type lectins – structural characteristics and their applications in diagnostics, analytics and drug delivery, *Int. J. Biol. Macromol.* 152 (2020) 735–747, <https://doi.org/10.1016/j.ijbiomac.2020.02.320>.
- [6] G.R. Vasta, Galectins in host–pathogen interactions: structural, functional and evolutionary aspects, 169–196, [https://doi.org/10.1007/978-981-15-1580-4\\_7](https://doi.org/10.1007/978-981-15-1580-4_7), 2020.
- [7] G.R. Vasta, C. Feng, M.A. Bianchet, T.R. Bachvaroff, S. Tasumi, Structural, functional, and evolutionary aspects of galectins in aquatic mollusks: from a sweet tooth to the Trojan horse, *Fish Shellfish Immunol.* 46 (2015) 94–106, <https://doi.org/10.1016/j.fsi.2015.05.012>.
- [8] M.C. Miller, C. Cai, K. Wichapong, S. Bhaduri, N.L.B. Pohl, R.J. Linhardt, H.-J. Gabius, K.H. Mayo, Structural insight into the binding of human galectins to corneal keratan sulfate, its desulfated form and related saccharides, *Sci. Rep.* 10 (2020), 15708, <https://doi.org/10.1038/s41598-020-72645-9>.
- [9] I.M.B. Francischetti, D. Ma, J.F. Andersen, J.M.C. Ribeiro, Evidence for a lectin specific for sulfated glycans in the salivary gland of the malaria vector, *Anopheles gambiae*, *PLoS One* 9 (2014), e107295, <https://doi.org/10.1371/journal.pone.0107295>.
- [10] L.C. Wang, L.Q. Di, J.S. Li, L.H. Hu, J.M. Cheng, H. Wu, Elaboration in type, primary structure, and bioactivity of polysaccharides derived from mollusks, *Crit. Rev. Food Sci. Nutr.* 59 (2019) 1091–1114, <https://doi.org/10.1080/10408398.2017.1392289>.
- [11] S. Kurz, C. Jin, 金春生, A. Hykollari, D. Gregorich, B. Giomarelli, G.R. Vasta, I.B. H. Wilson, K. Paschinger, Hemocytes and plasma of the eastern oyster (*Crassostrea virginica*) display a diverse repertoire of sulfated and blood group A-modified N-glycans, *J. Biol. Chem.* 288 (2013) 24410–24428, <https://doi.org/10.1074/jbc.M113.478933>.
- [12] S. Tasumi, G.R. Vasta, A galectin of unique domain organization from hemocytes of the eastern oyster (*Crassostrea virginica*) is a receptor for the protistan parasite *Perkinsus marinus*, *J. Immunol.* 179 (2007) 3086–3098, <https://doi.org/10.4049/jimmunol.179.5.3086>.
- [13] W.M. Gayashani Sandamalika, J. Lee, Quadruple domain-containing galectin from marine invertebrate disk abalone (*Haliotis discus discus*): molecular perspectives in early development, immune expression, and potent antiviral responses, *Fish Shellfish Immunol.* 106 (2020) 920–929, <https://doi.org/10.1016/j.fsi.2020.09.017>.
- [14] X. Song, H. Zhang, L. Wang, J. Zhao, C. Mu, L. Song, L. Qiu, X. Liu, A galectin with quadruple-domain from bay scallop *Argopecten irradians* is involved in innate immune response, *Dev. Comp. Immunol.* 35 (2011) 592–602, <https://doi.org/10.1016/j.dci.2011.01.006>.
- [15] W. Maldonado-Aguayo, J. Teneb, C. Gallardo-Escárate, A galectin with quadruple-domain from red abalone *Haliotis rufescens* involved in the immune innate response against *Vibrio anguillarum*, *Fish Shellfish Immunol.* 40 (2014) 1–8, <https://doi.org/10.1016/j.fsi.2014.06.013>.
- [16] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M. D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, *Anal. Biochem.* 150 (1985) 76–85, [https://doi.org/10.1016/0003-2697\(85\)90442-7](https://doi.org/10.1016/0003-2697(85)90442-7).
- [17] M. Wang, L. Wang, M. Huang, Q. Yi, Y. Guo, Y. Gai, H. Wang, H. Zhang, L. Song, A galectin from *Eriocheir sinensis* functions as pattern recognition receptor enhancing microbe agglutination and haemocytes encapsulation, *Fish Shellfish Immunol.* 55 (2016) 10–20, <https://doi.org/10.1016/j.fsi.2016.04.019>.
- [18] G.A. Fragkiadakis, Isolation of lectins from hemolymph of decapod crustaceans by adsorption on formalinized erythrocytes, *J. Biochem. Biophys. Methods* 44 (2000) 109–114, [https://doi.org/10.1016/S0165-022X\(00\)00089-0](https://doi.org/10.1016/S0165-022X(00)00089-0).
- [19] S. Boivin, S. Kozak, R. Meijers, Optimization of protein purification and characterization using ThermoFluor screens, *Protein Expr. Purif.* 91 (2013) 192–206, <https://doi.org/10.1016/j.pep.2013.08.002>.
- [20] F. Santini, A. Bisio, M. Guerrini, E.A. Yates, Modifications under basic conditions of the minor sequences of heparin containing 2,3 or 2,3,6 sulfated-d-glucosamine residues, *Carbohydr. Res.* 302 (1997) 103–108, [https://doi.org/10.1016/S0008-6215\(97\)00111-0](https://doi.org/10.1016/S0008-6215(97)00111-0).
- [21] Y. Inoue, K. Nagasawa, Selective N-desulfation of heparin with dimethyl sulfoxide containing water or methanol, *Carbohydr. Res.* 46 (1976) 87–95, [https://doi.org/10.1016/S0008-6215\(00\)83533-8](https://doi.org/10.1016/S0008-6215(00)83533-8).
- [22] A. Micsonai, F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra, *Nucleic Acids Res.* 46 (2018) W315–W322, <https://doi.org/10.1093/nar/gky497>.
- [23] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler,

- T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstern, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature* 596 (2021) 583–589, <https://doi.org/10.1038/s41586-021-03819-2>.
- [24] P. Emsley, K. Cowtan, Coot : model-building tools for molecular graphics, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60 (2004) 2126–2132, <https://doi.org/10.1107/S0907444904019158>.
- [25] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291, <https://doi.org/10.1107/S0021889892009944>.
- [26] Y. Yan, H. Tao, J. He, S.-Y. Huang, The HDock server for integrated protein–protein docking, *Nat. Protoc.* 15 (2020) 1829–1852, <https://doi.org/10.1038/s41596-020-0312-x>.
- [27] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera?A visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612, <https://doi.org/10.1002/jcc.20084>.
- [28] V. Solovyev, Statistical approaches in eukaryotic gene prediction, in: *Handb. Stat. Genet.*, John Wiley & Sons, Ltd, Chichester, UK, 2008, pp. 97–159, <https://doi.org/10.1002/9780470061619.ch4>.
- [29] J. Thompson, The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res.* 25 (1997) 4876–4882, <https://doi.org/10.1093/nar/25.24.4876>.
- [30] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symp. Ser.* (1999) 95–98, [https://doi.org/10.14601/Phytopathol\\_Mediterr-14998u1.29](https://doi.org/10.14601/Phytopathol_Mediterr-14998u1.29).
- [31] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, X. Mega, Molecular evolutionary genetics analysis across computing platforms, *Mol. Biol. Evol.* 35 (2018) 1547–1549, <https://doi.org/10.1093/molbev/msy096>.
- [32] N.C.J. Strynadka, M.N.G. James, Towards an understanding of the effects of calcium on protein structure and function, *Curr. Opin. Struct. Biol.* 1 (1991) 905–914, [https://doi.org/10.1016/0959-440X\(91\)90085-8](https://doi.org/10.1016/0959-440X(91)90085-8).
- [33] S. Mohamed, D. Coombe, Heparin mimetics: their therapeutic potential, *Pharmaceuticals* 10 (2017) 78, <https://doi.org/10.3390/ph10040078>.
- [34] V.L. Campo, D.F. Kawano, D.B. da Silva, I. Carvalho, Carrageenans, Biological properties, chemical modifications and structural analysis – a review, *Carbohydr. Polym.* 77 (2009) 167–180, <https://doi.org/10.1016/j.carbpol.2009.01.020>.
- [35] H. Ideo, The N-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity, *Glycobiology* 13 (2003) 713–723, <https://doi.org/10.1093/glycob/cwg094>.
- [36] H. Ideo, T. Matsuzaka, T. Nonaka, A. Seko, K. Yamashita, Galectin-8-N-domain recognition mechanism for sialylated and sulfated glycans, *J. Biol. Chem.* 286 (2011) 11346–11355, <https://doi.org/10.1074/jbc.M110.195925>.
- [37] C. Olsson, R. Zangana, J. Swenson, Stabilization of proteins embedded in sugars and water as studied by dielectric spectroscopy, *Phys. Chem. Chem. Phys.* 22 (2020) 21197–21207, <https://doi.org/10.1039/D0CP03281F>.
- [38] K. Bum-Erdene, H. Leffler, U.J. Nilsson, H. Blanchard, Structural characterisation of human galectin-4 N-terminal carbohydrate recognition domain in complex with glycerol, lactose, 3'-sulfo-lactose and 2'-fucosyllactose, *Sci. Rep.* 6 (2016), 20289, <https://doi.org/10.1038/srep20289>.
- [39] D. Houzelstein, I.R. Gonçalves, A.J. Fadden, S.S. Sidhu, D.N.W. Cooper, K. Drickamer, H. Leffler, F. Poirier, Phylogenetic analysis of the vertebrate galectin family, *Mol. Biol. Evol.* 21 (2004) 1177–1187, <https://doi.org/10.1093/molbev/msh082>.