



UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE BOGOTÁ

ELUCIDACIÓN DE LA ESTRUCTURA PRIMARIA DE LA LECTINA DE *Salvia bogotensis* (LSBo-I) ESPECÍFICA PARA LA DETECCIÓN DE LOS ANTÍGENOS T y Tn

Jessica Natalia Espinosa Velandia

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Química
Bogotá, Colombia
2022

ELUCIDACIÓN DE LA ESTRUCTURA PRIMARIA DE LA LECTINA DE *Salvia bogotensis* (LSBo-I) ESPECÍFICA PARA LA DETECCIÓN DE LOS ANTÍGENOS T y Tn

Jessica Natalia Espinosa Velandia

Tesis de investigación presentada como requisito parcial para optar al
título de: **Magister en Ciencias - Bioquímica**

Directora:
NOHORA ANGÉLICA VEGA CASTRO
Doctora en Ciencias Químicas

Línea de Investigación:
LECTINAS DE LAMIACEA
Grupo de Investigación en Proteínas (GRIP)

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Química
Bogotá, Colombia
2022

Agradecimientos

A la Universidad Nacional de Colombia, Facultad de Ciencias, Departamento de química, por la oportunidad de pertenecer al programa de Maestría en Ciencias-Bioquímica.

A mi directora de tesis la Dra. Nohora Angélica Vega Castro, por su orientación, colaboración, seguimiento y confianza en el desarrollo de este trabajo.

Al grupo de investigación en proteínas, GRIP, de la Universidad Nacional de Colombia, al Dr. Edgar Reyes por su apoyo, colaboración y orientación en el desarrollo del proyecto, igual que a todos mis compañeros en el grupo de investigación por sus enseñanzas, consejos y apoyo.

Al Doctor Gerardo Corzo Burguete, profesor del Instituto de Biotecnología – UNAM por su apoyo y los valiosos aportes durante el desarrollo de este trabajo.

Al Ministerio de ambiente y desarrollo (MADS), por otorgar el Contrato de acceso a recursos genéticos No. RGE 248 y a la Vicerrectoría de Investigación.

A mis padres Ana Velandia y Armando Espinosa por sus consejos, paciencia y apoyo incondicional y a mi hermano Oscar Javier Espinosa por su apoyo y colaboración en todo momento.

A mi esposo Ángel por su apoyo incondicional, por impulsarme cada día a continuar, y no dejar que me rindiera, porque aquí también se encuentra gran parte de él. A mi hija Emma, por llegar y hacer que todo esfuerzo valga mucho más la pena.

A todos mis amigos, por estar siempre presentes, escucharme y saber que decir en cada momento.

Resumen

Elucidación de la estructura primaria de la lectina de *Salvia bogotensis* (LSBo-I) específica para la detección de los antígenos T y Tn

El estudio de las lectinas vegetales ha sido de gran importancia debido a su capacidad de reconocer residuos de carbohidratos, los cuales han sido llamados glicotopes; alguno de ellos se encuentra enmascarados en la superficie celular y están asociados con una gran variedad de carcinomas. Algunas de las lectinas que más se han estudiado son las de origen vegetal de la familia *Leguminosae*, sin embargo, también se han aislado de otras familias botánicas, tales como Lamiácea.

Las lectinas de Lamiáceas se caracterizan porque reconocen los glicotopes T y/o Tn (1). La información disponible acerca de estas proteínas es muy escasa debido a que los procesos de purificación no permiten obtener la cantidad necesaria para realizar estudios de caracterización estructural y de interacción con el antígeno Tn/T; por otra parte, hasta el momento solo hay dos secuencias de especies relacionadas obtenidas a partir del cDNA, la lectina de *Salvia miltorriza* (LSM) (2) y de *Gleheda* (3) y solo se han caracterizado cuatro lectinas de la zona de Eurasia y dos lectinas de especies endémicas de Colombia, la lectina de *Salvia bogotensis* (LSBo-I) (4,5) y de *Lepechinia bullata* (LLb-I) (5)

En el presente trabajo, se realizó la purificación de la lectina de *Salvia bogotensis* I (LSBo-I), empleando cromatografía de intercambio iónico y afinidad. Las fracciones

obtenidas presentaron actividad aglutinante por eritrocitos Tn y se observó por SDS PAGE, una banda alrededor de 36 kDa por SDS-PAGE que es similar a lo obtenido en estudios previos (4). La sepharosa 4B, fue acoplada a una IgG de conejo (antiLSBo-I), y la elución se llevó a cabo mediante un gradiente de pH. En cada etapa de purificación se evaluó la actividad aglutinante con eritrocitos humanos A⁺, los cuales fueron tratados enzimáticamente con neuraminidasa y β galactosidasa para exponer el Tn. Se obtuvo el 71% de la secuencia y por predicción se determinó que pertenece a las lectinas tipo Leguminosa (L). También se llevaron a cabo ensayos para la obtención del ADNc y la secuencia de nucleótidos de la LSBo-I.

Palabras Clave: Lectina vegetal, *Salvia bogotensis*, Familia Lamiaceae, Glicotopes T y Tn, Lectinas tipo L.

Abstract

Elucidation of the primary structure of the *Salvia bogotensis* lectin (LSBo-I) specific for the detection of T and Tn antigen

Plant lectins studies have been of great importance due to their ability to recognize carbohydrate residues on surface cells, epitopes-like glycotopes are called. Some of them are masked on the cell surface and are associated with a wide variety of carcinomas. The most studied lectins belong to the Leguminosae family however, they have also been isolated from other botanical families, such as Lamiaceae. Their lectins are characterized by recognizing T and/or Tn glycotypes (1). The information available about these proteins is very scarce owing to purification processes do not allow obtaining the amount to carry out structural studies and interaction with Tn/T glycans. Secondly, there are only two sequences related to Lamiacea lectins, *Salvia miltorriza* (SML) (2) and *Gleheda* (3) and four lectins have been studied from Eurasia zone and two endemic lectins from Colombia, *Salvia bogotensis* (SBoL-I) (4,5) and *Lepechinia bullata* (LBL-I) (5)

In this work, Ion-exchange and affinity chromatography's led to getting SBoL-I that showed only a band by SDS PAGE, around 36 kDa, according to previous studies (4). Main purification step was carried out using affinity chromatography, a Sepharose 4B was coupled to IgG (antiSBoL-I) and pH gradient elution. The agglutinating activity was tested using A+ human erythrocytes, which were enzymatically treated with neuraminidase and β -galactosidase, to expose Tn

structures. Finally, a sequence with 71% coverage was obtained and three-dimensional model type of Leguminous was predicted by homology. Additionally, cDNA assays were carried out to get nucleotide sequence from SBoL-I.

Keywords: Vegetal lectin, *Salvia bogotensis*, Lamiaceae family, T and Tn glycotope, L-type lectins.

Contenido

Resumen	V
Abstract	VII
Lista de figuras	XII
Lista de tablas	XV
Lista de símbolos y abreviaturas	XVI
1. Introducción	1
2. Marco teórico	2
2.1 Lectinas.....	2
2.2 Clasificación de las lectinas	3
2.2.1. De acuerdo con el Sitio de reconocimiento.....	3
2.2.1.1. Merolectinas:.....	3
2.2.1.2. Hololectinas:.....	3
2.2.1.3. Superlectinas.....	3
2.2.1.4. Quimerolectinas:.....	3
2.2.2. De acuerdo con la especificidad del monosacárido que inhibe su actividad hemoaglutinante	4
2.2.3. De acuerdo con las diferentes familias de proteínas (Pfam):.....	5
2.3 Lectinas de origen Animal.....	7
2.4 Lectinas de origen Vegetal.....	10
2.4.1. Familia lectinas tipo ABA	11
2.4.2. Familia lectinas tipo heveína.....	12
2.4.3. Familia lectinas tipo amaranto.	12
2.4.4. Familia lectinas tipo Jacalina	14
2.4.5. Lectinas tipo Leguminosas (Tipo L)	15
2.5. Familia Lamiaceae – <i>Salvia bogotensis</i>	24
3. Justificación	26
4. Objetivos	28
4.1. Objetivo General	28
4.2. Objetivos específicos	28
5. Metodología	29
5.1. Purificación de lectina de <i>Salvia bogotensis</i> (LSBo-I)	29
5.1.1. Métodos generales	29
5.1.1.1. Obtención del material vegetal.....	29
5.1.1.2. Cuantificación de proteínas.....	30
5.1.1.3. Electroforesis unidimensional en geles de poliacrilamida	30
5.1.1.4. Diálisis.....	30
5.1.1.5. Ensayos de Aglutinación.....	30
5.1.2. Purificación de inmunoglobulinas g (IgG) anti-LSBo.....	31
5.1.2.1. Precipitación de proteínas con PEG 8000.....	31

5.1.2.2.	<i>Cromatografía de intercambio iónico</i>	32
5.1.2.3.	<i>Preparación del soporte de Afinidad Sepharosa 4B-IgG</i>	32
5.1.2.4.	<i>Ensayo de ELISA</i>	33
5.1.3.	Extracción y purificación de la lectina de <i>Salvia bogotensis</i> (LSBo)	34
5.1.3.1.	<i>Obtención del extracto de proteínas</i>	34
5.1.3.2.	<i>Precitación de proteínas</i>	35
5.1.3.3.	<i>Cromatografía de intercambio iónico</i>	35
5.1.3.4.	<i>Tratamiento con Pectinex®</i>	36
5.1.3.5.	<i>Cromatografía de inmunoafinidad sobre Sepharosa 4B IgG (antiLSBo-I)</i>	36
5.1.3.6.	<i>Cromatografía de afinidad sobre Sepharose 4B - Concanavalina A (ConA)</i>	36
5.1.4.	Determinación de la estructura primaria con la proteína pura	37
5.1.4.1.	<i>Preparación del digerido tróptico de la LSBo-I</i>	37
5.1.4.2.	<i>Análisis de los péptidos por nano LC-MS/MS</i>	37
5.1.4.3.	<i>Búsqueda en bases de datos e identificación de péptidos</i>	38
5.1.4.4.	<i>Predicción de la estructura terciaria</i>	39
5.1.4.5.	<i>Refinamiento modelos 3D</i>	39
5.1.4.6.	<i>Acople molecular</i>	40
5.2.	Extracción de ARN total de las semillas de <i>Salvia bogotensis</i>	40
5.2.1.	Extracción con Bromuro de hexadeciltrimetilamonio (CTAB).	40
5.2.2.	Extracción de ARN total con Trizol	41
5.2.3.	Síntesis de cDNA y amplificación de fragmentos de ADN	41
5.2.4.	Amplificación de fragmentos de ADN	42
6.	Resultados y Análisis	43
6.1.	Purificación de lectina de <i>Salvia bogotensis</i> (LSBo-I)	43
6.1.1.	Purificación del anticuerpo (IgG) contra LSBo-I.....	43
6.1.2.	Extracción y purificación de la lectina (LSBo-I).....	47
6.1.3.	Análisis de péptidos trópticos obtenidos de la LSBo-I por nano LC-MS/MS para determinar la secuencia de aminoácidos.....	55
6.1.3.1.	<i>Modelamiento de la estructura terciaria</i>	67
6.1.3.2.	<i>Acoplamiento molecular</i>	71
6.2.	Ensayos preliminares para la obtención de LSBo-I recombinante.....	76
6.2.1.	Extracción ARN por los métodos de CTAB y Trizol	77
6.2.2.	Diseño de primers.....	81
6.2.2.1.	<i>Traducción Inversa</i>	81
6.2.2.2.	<i>Diseño de primers</i>	85
6.2.3.	PCR	86
7.	Conclusiones y Recomendaciones	89
7.1	Conclusiones	89
7.2	Recomendaciones	90
	Referencias	90
	Anexos	101
A.	Anexo. Esquema inmunización conejos para obtener anti LSbo-i	102
B.	Anexo. Purificación por sepharosa 4b- cona.....	103
C.	Anexo. Secuencia de péptidos obtenidos por espectrometría de ms/	104
D.	Anexo. Alineamiento múltiple de la secuencia parcial de LSbo-i con los 12 péptidos obtenidos de ms/ms.....	107
E.	Anexo. Validación inicial de los modelos con mejor score de i-tasser y phyre ²	109

F.	Anexo. Cuantificación extracción de rna de semillas de <i>salvia bogotensis</i>	111
G.	Anexo. Tabla de codones de <i>salvia miltiorrhiza</i>	112
H.	Anexo: secuencias de transcripción inversa.....	114
I.	Anexo: pares de primers obtenidos mediante la herramienta bioinformática primerblast..	116
J.	Anexo. Validación de todos los primers diseñados con el programa pcr primer stats.....	117

Lista de figuras

Figura 1.	Esquema de clasificación de las lectinas vegetales de acuerdo con el sitio de reconocimiento.....	4
Figura 2.	Monosacáridos reconocidos por las lectinas	5
Figura 3.	Clasificación de las Lectinas.....	6
Figura 4.	Lectina Tipo C del veneno de <i>Crotalus atrox</i>	8
Figura 5.	Galectinas.....	9
Figura 6.	Patrón de expresión de las Siglecs dentro del sistema hematopoyético.. ..	9
Figura 7.	Lectina de <i>Agaricus bisporus</i> (ABL) con glicotipe T.....	11
Figura 8.	Familia tipo Heveina	12
Figura 9.	Lectina de <i>Amaranthus caudatus</i>	13
Figura 10.	Estructura de AAT4.....	14
Figura 11.	Lectinas tipo Jacalina	14
Figura 12.	Estructura de las lectinas de la familia de las monocotiledóneas	15
Figura 13.	Lectina tipo L que correspondiente a Concanavalina A (Con A).....	16
Figura 14.	Lectina de <i>Glechoma hederácea</i> (Gleheda)	19
Figura 15.	Lectina de <i>S. miltiorrhiza</i> (SmL).....	20
Figura 16.	Secuencia parcial de LSBo-I.....	22
Figura 17.	Predicción de la estructura tridimensional de LSBo-I.....	23
Figura 18.	<i>Salvia bogotensis</i> , foto tomada en el sector en Mondoñedo, vía la Mesa-Mosquera.....	24
Figura 19.	Algunos géneros de Lamiaceae representativos	25
Figura 20.	Sitio de colecta de las semillas de <i>S. bogotensis</i>	29
Figura 21.	Degradación enzimática con neuraminidasa y β galactosidasa en eritrocitos tipo A para exponer los glicotipes T y Tn para ensayos de eritroaglutinación.	31
Figura 22.	Esquema de purificación de IgGs anti LSBo-I con precipitación empleando polietilenglicol 8000 (PEG 8000).	32
Figura 23.	Reacción de acople molecular de un ligando (IgG anti LSBo-I) a una resina de sepharosa 4B pre activada con bromuro de cianógeno (CNBr)	33
Figura 24.	Esquema de extracción de proteínas totales con buffer PBS 1x – tiourea 5mM a partir de harina de semillas de <i>Salvia bogotensis</i>	35
Figura 25.	Precipitación fraccionada de proteínas con etanol 96% en una concentración de 0-60%.....	35
Figura 26.	(A). Cromatografía de intercambio iónico sobre DEAE Sephacel (B) SDS PAGE Tris- tricina (12,5 %), Carril 1: Antisuero, Carril 2: Precipitado PEG 10%, Carril 3: Fracción no retenida I (FNR) de la cromatografía DEAE Sephacel (II).....	45
Figura 27.	Representación de la inmunoglobulina G (IgG), en rojo, las cadenas pesadas y en verde las cadenas ligeras, también se pueden ver en líneas punteadas los puentes disulfuro intracatenarios e intercatenarios	45

Figura 28.	Ensayo de ELISA para la detección de la interacción LSBo-I vs IgGs.....	46
Figura 29.	Extracto de proteínas precipitado con etanol 0-60 % y solubilizado en PBS 1X	48
Figura 30.	SDS PAGE Tris Glicina 12,5 %. Extracto proteico de semillas de <i>S. bogotensis</i> (Sbo), PM (patrones de peso molecular). Carril 1, Extracto I. Carril 2 Extracto II, Carril 3. Extracto III. Carril 4. Pool de extractos, Carril 5. Precipitado de proteínas con etanol 0-60 %. Carril 6. Extracto precipitado con etanol 0-60 % dializado.....	48
Figura 31.	(A) Cromatografía DEAE Sephadex A-50® (I) Fracción no retenida FNR, (II) Fracción retenida FR, elución con alta fuerza iónica (PBS 1X - NaCl 500 mM). (B) Fracción no retenida DEAE - SDS PAGE Tris – Glicina 12,5 %.....	49
Figura 32.	Cromatografía de Inmunoadinidad usando un soporte de Sepharosa-4B inmunoglobulina G (IgG:Anti- LSBo-I).....	51
Figura 33.	(A). SDS PAGE 12,5 % LSBo-I obtenida por inmunoadinidad sobre Sepharosa 4B IgG. mw: patrones de peso molecular, Carril 1. Fracción retenida (FR), pH 8,3. Carril 2. Fracción retenida (FR), pH 2,8. Carril 3. Fracción no retenida (FNR). Se colocaron 30 µg de proteína total en cada carril. (B)(C). SDS PAGE 12,5 %, para la LSBo-I obtenida por Cromatografía de afinidad sobre Sepharosa 4B-aMSB.....	52
Figura 34.	(A) Cromatografía de afinidad sobre Sepharose 4B- ConA. (I) FNR, (II) FR Glc, (III), FR 2,8. (B) Fracciones de cromatografía de afinidad - Sepharosa 4B ConA, SDS PAGE 12,5 %: carril 1: FR eluida con NaCl 1 % - Glucosa 0,2 M (5µg de proteína), Carril 2: FR eluida con Gly HCl pH 2,8 (10 µg de proteína), Carril 3: FNR.....	53
Figura 35.	Aglutinación de eritrocitos Tn al 2.5 % con la LSBo-I ^b (A) ^a FNR DEAE (4,5 µg de proteína) (+2), (B) FR pH 2,8 (0,18 µg de proteína) (+4), (C) FR pH 8,3 (0,64 µg de proteína) (+4), (D) FNR -Con A, (0,56 µg de proteína) (+3) (E) Blanco (eritrocitos sin lectina).....	54
Figura 36.	SDS-PAGE 15 % LSBo-I. Se observa la banda de proteína que fue cortada y sometida a digestión con tripsina.....	56
Figura 37.	Alineamientos pareados de Secuencia parcial de LSBo-I con los péptidos (A) Péptido 1, (B) Péptido 2, (C) Péptido 3, (D) Péptido 11, (E) Péptido 12. Un * (asterisco) indica residuo conservado. Dos puntos (:), indican conservación entre grupos de propiedades muy similares. Un punto (.), indica la conservación entre grupos de propiedades débilmente similares.....	60
Figura 38.	Alineamiento múltiple de los 5 péptidos seleccionados (1,2,3,11 y,12) con la lectina parcial de LSBo-I y la secuencia nueva (propuesta) de 278 aa.....	61
Figura 39.	Estructura primaria parcial LSBo-I propuesta.....	62
Figura 40.	Alineamiento múltiple de LSBo-I (<i>S. bogotensis</i> I) (en rojo), con 10 lectinas específicas por el glicótopo Tn pertenecientes a diferentes familias.....	63
Figura 41.	(A) Porcentaje de identidad entre las lectinas alineadas con LSBo-I. (B) árbol Filogenético. 64	64
Figura 42.	Estructura obtenida con I-Tasser sin minimización de energías.....	68
Figura 43.	Validación predicha del modelo obtenido por Phyre2 refinado por GalaxyWEB empleando MolProbity.....	69
Figura 44.	Validación predicha del modelo obtenido por Phyre ² refinado por GalaxyWEB empleando Ramachandran.....	69
Figura 45.	Estructura predicha por Phyre ² con minimización de energías por GalaxyWEB para LSBo-I.....	70
Figura 46.	Predicción de la estructura 3D de LSBo por Phyre ² en la que se resalta en rojo el bolsillo de unión a ligando a GalNac determinado por Discovery studio.....	70

Figura 47.	Predicción de la estructura tridimensional de LSBo-I predicha por Phyre ² con distancias entre las hojas frontal y posterior determinadas con Pymol.	71
Figura 48.	Modelo de acoplamiento entre LSBo-I y el glicotipe T obtenido por Autodock Vina.	73
Figura 49.	Modelo de acoplamiento entre LSBo-I y glicotipe Tn, obtenido por Autodock Vina.....	74
Figura 50.	Modelo de acoplamiento entre LSBo-I y GalNAC obtenido por Autodock Vina.....	75
Figura 51.	Alineamiento tridimensional de las lectinas de <i>S. bogotensis</i> -I propuesta en este trabajo con la propuesta en trabajos previos.....	76
Figura 52.	Semillas de <i>S. bogotensis</i> en estado inmaduro.	77
Figura 53.	Gel de agarosa 2 %. Extracción de RNA total de semillas de <i>S. bogotensis</i> a partir de (A) Método de CTAB (B) Método de Trizol..	79
Figura 54.	Alineamiento múltiple de las 5 secuencias obtenidas de traducción Inversa RT con codones ambiguos, RT estándar, RT con codones de <i>A. thaliana</i> , RT con codones de <i>S. miltiorrhiza</i> y RT con codones de <i>O. sativa</i>	84
Figura 55.	Secuencia consenso propuesta de 822 pb, para la obtención de cDNA de la LSBo-I. 84	
Figura 56.	Alineamiento de las secuencias traducida obtenida de la secuencia consenso (Traslate), con la secuencia parcial de LSBo-I (5) (Parcial)	85
Figura 57.	Conjunto de Primers diseñados para la obtención del gen de la LSBo-I.	86
Figura 58.	PCR con los primers diseñados. Gel de agarosa 2 %	87
Figura A-1.	Determinación del título del antisuero anti-lectina LSBo-I.....	101
Figura D-1.	Múltiple alineamiento realizado en clustal O (1.2.4) de la secuencia parcial de LSBo-I con los 12 péptidos obtenidos de MS/MS.....	106
Figura E-1.	Validación predicha del modelo obtenido por I-Tasser empleando molprobity.....	108
Figura E-2.	Validación predicha del modelo obtenido por I-Iasser empleando ramachandran.....	108
Figura E-3.	Validación predicha del modelo obtenido por Phyre2 empleando molprobity.....	108
Figura E-4.	Validación predicha del modelo obtenido por Phyre2 empleando ramachandran.....	109
Figura H-1.	LSBo-I secuencia de nt más probable usando codones para <i>S. miltiorrhiza</i>	112
Figura H-2.	LSBo-I secuencia ambigua derivada de todos los codones posibles para cada aminoácido.....	112
Figura H-3.	Secuencia empleando la tabla de codones para <i>arabidopsis thaliana</i>	112
Figura H-4.	Secuencia empleando la tabla de codones para <i>oryza sativa</i>	112
Figura H-5.	Protein sequence back-translation con el algoritmo backtranambig.....	113
Figura J-1.	Resultados obtenidos de la validación de todos los Primers diseñados con el programa PCR primer stats	115

Lista de tablas

Tabla 1.	Clasificación de acuerdo con el monosacárido que reconoce las lectinas.....	4
Tabla 2.	Principales características de las lectinas aisladas de Lamiaceas T/Tn.....	17
Tabla 3.	Características moleculares para lectinas aisladas de las semillas de <i>S. bogotensis</i> ..	21
Tabla 4.	Purificación de IgGs totales (anti- LSBo-I)	45
Tabla 5.	Acoplamiento de IgGs a Sepharose 4B	47
Tabla 6.	Etapas de purificación de la lectina de <i>Salvia bogotensis</i> (LSBo-I)	55
Tabla 7.	Péptidos obtenidos por digestión trípica y secuenciación (MS/MS).....	56
Tabla 8.	Análisis de aminoácidos obtenidos de la LSBo-I	66
Tabla 9.	Afinidad que presenta el modelo a diferentes ligandos.....	72
Tabla 10.	Descripción de las interacciones de los aminoácidos en el sitio de reconocimiento con el carbohidrato T.	73
Tabla 11.	Descripción de las interacciones de los aa en el sitio de reconocimiento con el Ligando con epitope T.....	74
Tabla 12.	Descripción de las interacciones de los aminoácidos en el sitio de reconocimiento con los carbohidratos GalNAC.....	75
Tabla 13.	Aminoácidos identificados en el sitio de unión de LSBo-I, que presentan interacción con los ligandos ensayados.	75
Tabla 14.	Extracción de RNA total de semillas de <i>Salvia bogotensis</i> (SBo)	79
Tabla 15.	Primers seleccionados para obtención del gen de LSBo-I.....	86
Tabla A-1.	Protocolo de Inmunización con lectina de <i>S. bogotensis</i>	101
Tabla A-2.	Obtención de antisueros para evaluar la respuesta al antígeno	101
Tabla B-1.	Cuantificación de purificación por Sepharosa 4B- ConA	102
Tabla C-1.	Secuencia de péptidos obtenidos por espectrometría de MS/MS y % de cobertura con proteína con dominio tipo Leguminosa	103
Tabla F-1.	Extracción de RNA de semillas de <i>Salvia bogotensis</i>	110
Tabla G-1.	Tabla de codones de <i>Salvia miltiorrhiza</i> , empleada para la traducción inversa de la secuencia de la LSBo-I	111
Tabla I-1.	Pares de primers obtenidos mediante la herramienta bioinformática PrimerBlast.....	114

Lista de símbolos y abreviaturas

ABREVIATURA	TERMINO
AAT4	Dominio Tipo amarantina (AA) Dominio de toxina (T)
ABA	Lectina <i>Agaricus bisporus</i>
ABTS	Ácido 2,2'-azino-bis(3-etilbenzotiazolina-6-sulfónico)
Ala	Alanina
ADN	Ácido Desoxiribonucleico
ALL	Lectina <i>Amaranthus leucocarpus</i>
Ara	Arabinosa
ARN	Ácido Ribonucleico
Asn	Asparagina
Asp	Ácido aspártico
Arg	Arginina
aMSB	Asialo Mucina Submaxilar Bovina
aMSO	Asialo Mucina Submaxilar Ovina
antiLSBo-I	IgG Antilectina de <i>Salvia bogotensis</i> -I
BCA	Ácido bicinconínico
BLAST	Herramienta básica para búsqueda de alineamientos locales
BSA	Albúmina Bovina de Suero
CA	Cromatografía de afinidad
cDNA	ADN complementario
CDR	Domino de reconocimiento a carbohidrato
CHS	Carbohidratos
Con A	Concanavalina A
CTAB	Métodode extracción de bromuro de hexadeciltrimetilamonio
Cys	Cisteína
DSA	<i>Datura stramonium</i> agglutinin
DEAE	Dietilaminoetilo
dNTP	Desoxinucleótido trifosfatado
DTT	Ditiotreitol
ELISA	Enzyme Linked Inmuno-Sorbent Assay
EM	Espectrometría de Masas
ESI-EM	Espectroscopía de Masas – Ionización por Electrospray
FNR	Fracción no retenida
FR	Fracción retenida
Fuc	Fucosa
Gal	Galactosa

ABREVIATURA	TERMINO
Glc	Glucosa
GalNAc	N-acetilgalactosamina
GlcNAc	N-acetilglucosamina
Gln	Glutamina
Glu	Ácido glutámico
Gly	Glicina
GRIP	Grupo de Investigación en Proteínas
HeLa	Linea celular de cancer de cervix
His	Histidina
Ile	Isoleucina
HPLC	Cromatografía líquida de alta resolución
IgG	Inmunoglobulina G
kDa	Kilo Dalton
LCA	Lectina Lens culinaris
LEL	<i>Lectina Solanum lycopersicum</i>
LSBo-I	<i>Lectina I de Salvia bogotensis</i>
LSBo-II	<i>Lectina II de Salvia bogotensis</i>
LLb-I	<i>Lectina I de Lepechinia bullata</i>
LPA	<i>Lectina Limulus polyphemus</i>
LLb-II	<i>Lectina II de Lepechinia bullata</i>
LSM	<i>Lectina de Salvia miltiorrhiza</i>
LSS	<i>Lectinas de Salvia sclarea</i>
LTA	<i>Lectina Lotus tetragonolobus</i>
Lys	Lisina
M	Molaridad
mA	Miliamperios
MALDI-MS	Mass Spectroscopy
Man	Manosa
mARN	ARN mensajero
MCF-7	Linea celular de cancer de seno
MES	Ácido 4-morfolinetanosulfónico sal hemisódica
Met	Metionina
mL	Mililitros
mM	Milimolar
MOA	Marasmius oreades
Mr	Masa relativa
MS	Espectroscopía de Masas
MSB	Mucina Submaxilar Bovina
MSO	Mucina Submaxilar Ovina
MUC-2	Mucina tipo 2

ABREVIATURA	TERMINO
NCBI	Centro Nacional para la Información Biotecnológica
NeuAca	Ácido siálico
ng	Nanogramos
NK	Natural killer
nL	Nanolitros
nM	Nanomolar
PAGE	Electroforesis en gel de Poliacrilamida
PB	Pares de bases
PBS	Buffer Fosfato Salino
PCR	Reacción en Cadena de la Polimerasa
PEG	polietilenglicol
Pfam	Familias de proteínas
PFT	Toxina formadora de poros
Phe	Fenilalanina
PM	Peso Molecular
PMN	polimorfonucleares
PNA	Lectina <i>Arachis hypogaea</i>
Ppm	Partes por millón
Pro	Prolina
PVDF	Fluoruro de Polivinilideno
Ram	Ramnosa
RCA	Ricinus communis
RMN	Resonancia magnetica nuclear
RSL	Lectina de <i>bi vknjig tk</i>
RT	retro- transcripción
PCR	Reacción en cadena de la polimerasa
SDS	Dodecil sulfato de Sodio
Ser	Serina
Siglecs	lectinas que unen ácido siálico
sTn	Antígeno sialil Tn
STL	Lectina <i>Solanum tuberosum</i>
T	Antígeno T (Gal β 1,3 GalNAc-O-Ser/Thr)
Talfa	Antígeno (Gal β 1,3GalNAc α 1-R)
Taq	Thermus aquaticus
TBS	Buffer Tris Salino
TFA	Ácido trifluoroacético
Thr	Treonina
Tipo L	Lectinas tipo Leguminosa
Tm	Temperatura de melting
Tn	Antígeno Tn (GalNAc-O-Ser/Thr)

ABREVIATURA	TERMINO
TNF	Factor de necrosis tumoral alfa
Trp	Triptófano
Tyr	Tirosina
UDA	<i>Aglutinina Urtica dioica</i>
Val	Valina
VML	<i>Lectina Vatairea macrocarpa</i>
VVLB ₄	<i>Isolectina B₄ de Vicia villosa</i>
WBAI	<i>Lectina Psophocarpus tetragonolobus</i>
WGA	<i>Lectina Triticum vulgaris</i>

1.Introducción

Por su capacidad de interacción con carbohidratos específicos como los glicotopes T, Tn y sTn, los cuales son expuestos en células tumorales, las lectinas vegetales han sido usadas como moléculas para el diagnóstico y/o pronóstico de patologías donde se encuentran glicosilaciones aberrantes (6–9). Las lectinas aisladas de la familia Fabácea han sido ampliamente estudiadas, se ha encontrado que aproximadamente diecinueve lectinas reconocen específicamente el glicotope Tn (10) y únicamente han sido caracterizadas las estructuras tridimensionales de los complejos con el Tn en tres de ellas, la lectina de *Psophocarpus tetragonolobus* Linneo (WBAI), la isolectina de *Vicia villosa* Roth B4 (VVLB₄) y *Vatairea macrocarpa* Ducke (VML)(11).

La información disponible con respecto a las lectinas en la familia Lamiacea es muy escasa, puesto que únicamente se han reportado las estructuras terciarias *in silico* de las lectinas de *Salvia miltiorrhiza* Bunge (2) y de *Glechoma hederácea* P.Foum (3), que fueron obtenidas a partir de la secuenciación del gen y no de la proteína nativa como las lectinas de *Salvia sclarea* Linneo (LSS) (12–14) y *Moluccella laevis* Linneo (15), (16), las cuales se estudiaron a nivel bioquímico.

En el Grupo de Investigación en Proteínas (GRIP) se han caracterizado las lectinas de *Salvia palifolia*, *Hyptis mutabilis* Rich (17), *Lepechinia bullata* Kunth (LLb-I) (5), y *Salvia bogotensis* Benth (LSBo-I) (4,5). Fueron purificadas a partir de extractos salinos de semillas, empleando métodos de precipitación fraccionada y diferentes técnicas cromatográficas; además, se estableció el 62 % de la secuencia de la LSBo-I (5). No obstante, es importante alcanzar una cobertura mayor en su secuencia y de esta forma llevar a cabo estudios de estructura terciaria para entender su mecanismo de interacción con glicotopes expuestos en tejidos tumorales como el Tn. La obtención de las lectinas de Lamiáceas en suficiente cantidad y de buena calidad, ha sido una limitante para continuar con los estudios de estructura, actividad funcional y biológica; en este aspecto es importante buscar nuevas alternativas de purificación de la proteína nativa u obtener el cDNA y por lo tanto la proteína recombinante. En este estudio se avanzó en la obtención de la LSBo-I, por cromatografía de afinidad empleando anticuerpos policlonales y se llevaron a cabo ensayos para obtener el cDNA de la lectina.

2. Marco teórico

2.1 Lectinas

Se encuentran ampliamente distribuidas en la naturaleza, por lo tanto, pueden ser aisladas de diversas fuentes biológicas como mamíferos, plantas, hongos, invertebrados, bacterias, entre otros. Inicialmente conocidas como “hemaglutininas” al ser detectadas por la aglutinación de eritrocitos con el veneno de la serpiente *Crotalus durissus* en 1880, y en 1886 con la detección de aglutinación de una fracción de proteína tóxica de ricino (denominado ricina) proveniente de extractos de frijoles *Ricinus communis* (18). El término “lectina” fue propuesto por Boyd y Shapleigh en 1954 a partir de la palabra latina “*legere*”, que significa escoger o seleccionar (19). La principal característica de estas proteínas o glicoproteínas, es su capacidad de unión a carbohidratos de manera específica y reversible, propiedad que les da un enorme potencial para estudiar diversas glicoproteínas y los mecanismos dados por interacciones proteína – carbohidrato (20). Por lo tanto, reconocen de manera específica carbohidratos de la superficie celular o de células en suspensión, que permiten que se les detecten por aglutinación de células o precipitación de glicoconjugados. El reconocimiento de la lectina por el carbohidrato se da principalmente por interacciones débiles que incluyen puentes de hidrógeno e interacciones hidrofóbicas, de tal forma que es reversible. Estas proteínas son heterogéneas en su estructura, usualmente multivalentes, y por lo general poseen más de un dominio de reconocimiento a carbohidrato, lo que les confiere su capacidad aglutinante. Son diferentes a las inmunoglobulinas que reconocen epitopes de antígenos o a las enzimas que tienen carbohidratos como sustratos y en este caso son modificados (21).

La gran versatilidad de estas proteínas ha permitido que hoy en día se postulen como agentes insecticidas (22,23), antifúngicos (24,25), antivirales (26) y antitumorales (21,22,23), entre otros. Desde otro punto de vista, se emplean como ligandos en cromatografía de afinidad (29), en estudios de histoquímica y citoquímica (25,26). Más recientemente, han sido incluidas en técnicas de alto rendimiento como los microarreglos para evaluar los glicotopes presentes en tejidos, células u otras muestras; un ejemplo es el estudio de enfermedades como diabetes (31) o tejidos tumorales (32). Otras aplicaciones incluyen los biosensores basados en las interacciones proteína – carbohidrato que se transforman en señales cuantificables (33).

2.2 Clasificación de las lectinas

La gran diversidad de estructuras y por lo tanto de funciones biológicas que se les atribuyen a las lectinas presentes en los diferentes reinos de la naturaleza, al igual que su capacidad de reconocer diferentes estructuras de carbohidratos, han permitido proponer diferentes formas de clasificación, las cuales se resumen en la Figura 3. En forma muy breve se explican algunas de ellas:

2.2.1. De acuerdo con el Sitio de reconocimiento

Esta clasificación se ha propuesto principalmente para lectinas de origen vegetal de acuerdo con los sitios de reconocimiento y otras funciones diferentes a la interacción con residuos carbohidratos así (Figura 1):

- 2.2.1.1. **Merolectinas:** Son proteínas que tienen un único dominio de unión a carbohidrato, es decir son monovalentes y no precipitan glicoconjugados o células (34). Un ejemplo es la heveína, como su nombre lo indica es aislada de la *Hevea brasiliensis*, con actividad antifúngica (35).
- 2.2.1.2. **Hololectinas:** presentan dos o más dominios de unión a carbohidratos que reconocen carbohidratos idénticos u homólogos; de modo que pueden aglutinar células o precipitar glicoconjugados. A este grupo pertenecen una gran cantidad de las lectinas vegetales (34).
- 2.2.1.3. **Superlectinas:** poseen dos o más dominios de unión a carbohidrato que reconocen carbohidratos diferentes, tales como residuos de Glucosa (Glc) y Galactosa (Gal). Un ejemplo es la lectina de *Agaricus bisporus* (ABA), que detecta el disacárido T (Gal β 1-3GalNAc α) y estructuras relacionadas de glicoproteínas O-glicosídicas. Por difracción de rayos X se encontró un sitio de unión por N-acetil glucosamina (GlcNAc) (36).
- 2.2.1.4. **Quimerolectinas:** se caracterizan por tener un dominio de unión a carbohidrato y un dominio no relacionado que cuenta con actividad biológica o catalítica bien definida (37). A esta clase de lectinas pertenece la del hongo *Marasmius oreades* (MOA), en el N-terminal tiene un dominio de reconocimiento por los determinantes del grupo sanguíneo B (Galactosa) y en el c-terminal un dominio con actividad catalítica de cisteína proteasa dependiente de calcio (38).

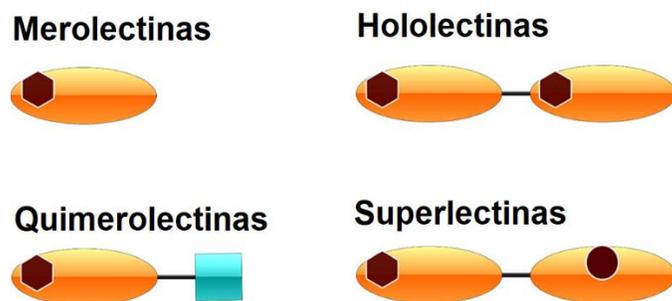


Figura 1. Esquema de clasificación de las lectinas vegetales de acuerdo con el sitio de reconocimiento. Se presenta el dominio de lectina (ovalo naranja), otro dominio con diferente actividad (cuadrado azul). Los diferentes carbohidratos reconocidos son representados como un hexágono o un círculo. Figura tomada y modificada (34).

2.2.2. De acuerdo con la especificidad del monosacárido que inhibe su actividad hemo aglutinante

Teniendo en cuenta esta característica, se divide en cinco grupos como se muestra en la Tabla 1:

Tabla 1. Clasificación de acuerdo con el monosacárido que reconoce las lectinas.

Grupo	Monosacárido	Abreviatura	Lectina
I	α -D-manosa α -D-glucosa	Man, Glc	Concanavalina A (ConA) <i>Lens culinaris</i> (LCA).
II	β -galactosa N-acetil- α -D-galactosamina	Gal, GalNAc	<i>Ricinus communis</i> (RCA) <i>Arachis hypogaea</i> (PNA) <i>Amaranthus leucocarpus</i> (ALL)
III	N-acetil- β -D-glucosamina	GlcNAc	<i>Triticum vulgare</i> (WGA)
IV	α -D-fucosa	Fuc	<i>Lotus tetragonolobus</i> (LTA)
V	α -N-acetilneuraminico	NeuAc α (SiA)	<i>Limulus polyphemus</i> (LPA)

De acuerdo con la base de datos UniLectin3D <https://www.unilectin.eu/unilectin3D>, la Galactosa (Gal) es el monosacárido principalmente reconocido por el 53 % de las lectinas; le siguen N-acetilglucosamina (GlcNAc) (27 %), glucosa (Glc) (24 %), manosa (Man) (22 %), fucosa (Fuc) (16 %), N-acetilgalactosamina (GalNAc) (13 %) y ácido siálico (NeuAc α) (12 %). No obstante, en la naturaleza se pueden encontrar azúcares en menor proporción como la ramnosa (Ram) y arabinosa (Ara), que también son reconocidos por algunas lectinas (Figura 2) (39).

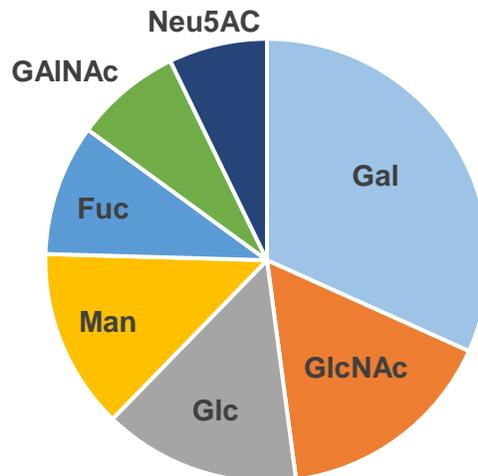


Figura 2. Monosacáridos reconocidos por las lectinas. Información obtenida de la base de datos UniLectin3D. Galactosa (Gal) el más reconocido por aproximadamente el 53 % de las Lectinas, seguido de (GlcNAc), glucosa (Glc), manosa (Man), fucosa (Fuc), N-acetilgalactosamina (GalNAc) y ácido siálico (NeuAc α).

2.2.3. De acuerdo con las diferentes familias de proteínas (Pfam):

Según la estructura tridimensional determinada por cristalografía de rayos X o análisis de resonancia magnética nuclear bidimensional (RMN), las lectinas se han clasificado en 48 familias topológicamente diferentes, encontrándose que los motivos estructurales de hojas β son predominantes en la estructura. Dos hojas β pueden ensamblarse para formar un β sándwich, una estructura muy común en las familias de lectinas, aunque suelen mostrar diferencias en la ubicación de los sitios de unión. Por ejemplo, el plegamiento tipo inmunoglobulina de las sialoadhesinas animales difiere notablemente del plegamiento característico de las lectinas de leguminosas (33).

En la base de datos Pfam se encuentran las familias de lectinas de acuerdo con los dominios de reconocimiento. Las lectinas tipo C y las galectinas, son dos grandes familias que se encuentran en animales mientras que en plantas están las tipo R (41) y las tipo L que se distinguen por su estructura tridimensional (42).

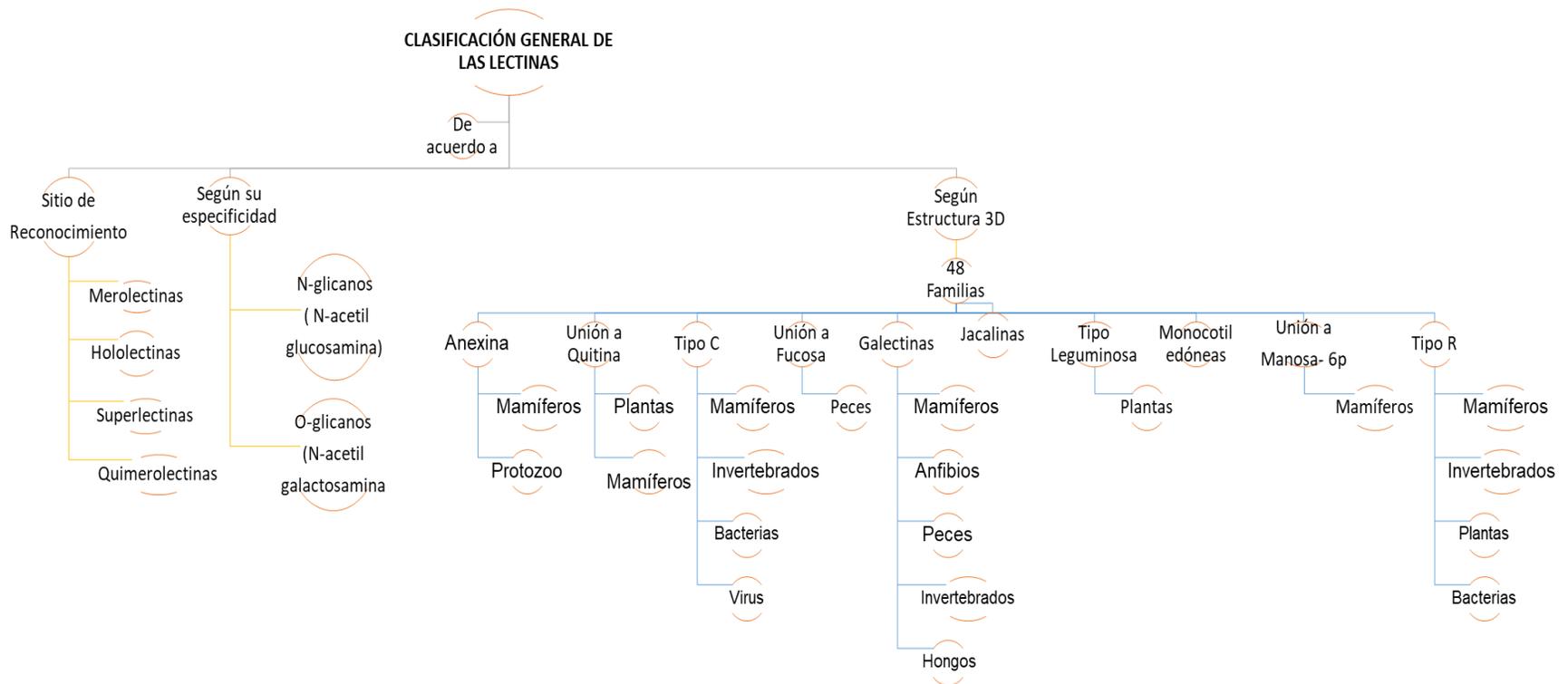


Figura 3. Clasificación de las Lectinas: Se han descrito 48 familias proteínas de acuerdo con su dominio de reconocimiento (se incluye solo algunas).

2.3 Lectinas de origen Animal

Los estudios referentes con estas proteínas en animales inician con la detección de actividad de aglutinación de eritrocitos por parte del veneno, de *Crotalus durissus* en 1902, no obstante, dos años antes del descubrimiento de ricina, 1886, había un reporte de la presencia de estas proteínas en el veneno. Muchos años después, la trombolectina del veneno de *Bothrops atrox* fue la primera lectina que se aisló y se caracterizó por tener monómeros de 15 kDa además de requerir iones Ca^{++} para su actividad; muchas no fueron reconocidas como proteínas de unión a carbohidratos (43). La función de las lectinas animales se asocia con la gran diversidad de estructuras tridimensionales, puesto que, al menos doce (12) familias se unen a distintas estructuras de carbohidratos mediante interacciones proteína-proteína, proteína-lípido o proteína-ácido nucleico (44), lo que conduce a que se activen diferentes vías de señalización en la célula, tales como sistema complemento, adhesión, migración, señalización, tráfico celular, regulación inmunitaria, autoinmunidad y morfogénesis, entre muchas más (45,46).

Las lectinas de tipo C, están en un amplio rango de animales, entre ellas están las selectinas, las cuales están asociadas a membrana plasmática, otras son solubles como las collectinas o las que se encuentran en los venenos. Un ejemplo bien documentado, corresponde a la lectina del veneno de la serpiente de cascabel occidental (RSL) (*Crotalus atrox*). Es una lectina tipo C, oligomérica y específica a galactosa. Su estructura cristalina en complejo con lactosa y/o tiodigalactósido, a una resolución de 2,2 y 2,3 Å, respectivamente, está formada por dos pentámeros simétricos dispuestos en una orientación escalonada, espalda con espalda. Cada monómero corresponde a un único dominio canónico de reconocimiento de carbohidratos (CDR) y está unido por puentes disulfuro a un monómero en el otro pentámero, de esta forma se generan 10 sitios de unión, que se encuentran al borde del decámero (Figura 4-A). Esta estructura permite que se den interacciones multivalentes; un mecanismo donde RSL promueve la reticulación del receptor y la agregación celular. Cada módulo (CDR), tiene aproximadamente 110 aminoácidos con especificidad por galactosa donde tres aminoácidos, Gln-Pro-Asp, se consideran determinantes en la interacción mediada por Ca^{++} (Figura 4-A) (47).

Cada monómero (Figura 4-B) posee un dominio de unión canónico y, se compone de dos hojas beta (β), cada una con 4 hebras antiparalelas flanqueadas por dos hélices cortas 3_{10} , giros y regiones desordenadas. La interfase entre pentámeros está mediada por interacciones débiles y la reducción de puentes disulfuro disocia los dímeros. La capacidad de mediar interacciones multivalentes explica el hecho de que RSL es un potente mediador de la agregación plaquetaria, un proceso que involucra al receptor en entrecruzamiento, reticulación y señalización posterior (47).

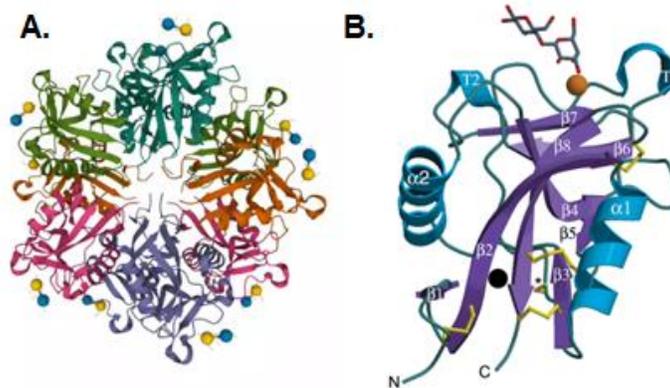


Figura 4. Lectina Tipo C del veneno de *Crotalus atrox*. (A) Dos pentámeros unidos simétricamente, donde los monómeros de cada pentámero se unen mediante puentes disulfuro. Se observan los sitios de unión al carbohidrato (CDR) o dominios canónicos, en este caso tiene 10. (B). Monómero formado dos hojas β y 8 hebras antiparalelas (morado), hélices α (azul), giros γ (azul) y regiones desordenadas. Ion calcio (anaranjado). Modificado (47).

Otras lectinas de origen animal son las galectinas, las cuales han sido principalmente descritas en mamíferos, aves y esponjas. Se encuentran en la superficie celular y citoplasma, mediando procesos como proliferación y adhesión, y migración celular entre otros (48). Inicialmente, se les dio el nombre de lectinas tipo S, dado que el grupo tiol (cisteína) está involucrado en el reconocimiento de β galactósidos, presentes en glicoconjugados. Un ejemplo son las Congerinas (Con) aisladas del mucus del pez *Conger myriaster* (Figura 5-A), componentes del sistema de defensa con propiedades de opsonización y con actividad citotóxica. Las dos lectinas, Con-I y Con-II, son homodiméricas, pero se diferencian en la forma de unión entre las hojas β que conforman cada monómero; en Con-I la interacción es paralela (Figura 5-A), lo que confiere mayor estabilidad y reconocimiento de los carbohidratos de parásitos u otros glicanos (49). La Galectina de pollo, adopta el típico pliegue de galectina, con dos hojas β antiparalelas de seis y cinco hebras conectadas por varios bucles, formando una estructura de β sándwich (Figura 5-B) (50).

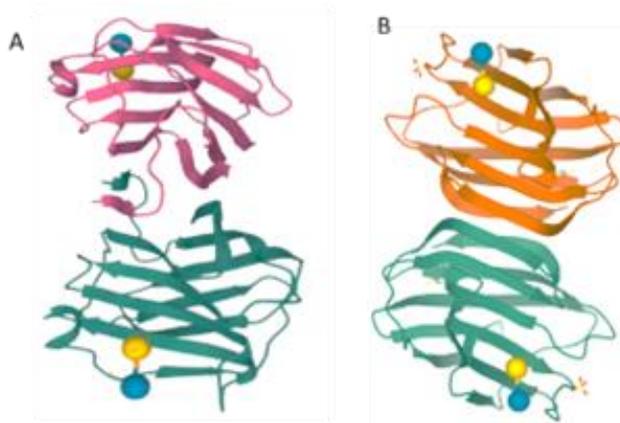


Figura 5. Galectinas. Se observan las estructuras diméricas, donde en cada monómero, hay dos hojas β antiparalelas, de seis y cinco hebras conectadas por bucles, generando una estructura β sándwich. La interacción entre los monómeros difiere para estas lectinas como se observa para la Galectina de *Conger myriaster* Tipo I (Con-I) (**A**) (49) y la Galectina de pollo (**B**) (50).

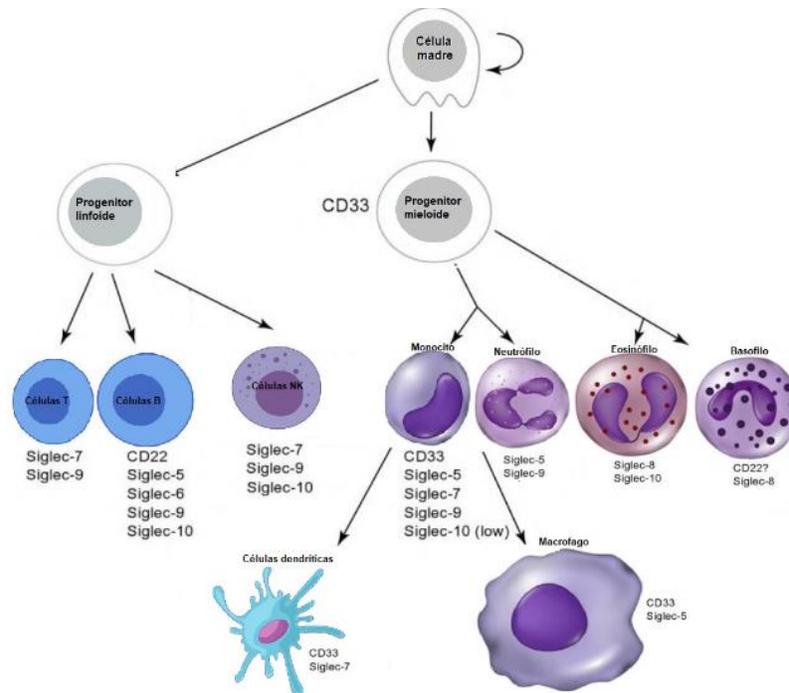


Figura 6. Patrón de expresión de las Siglecs dentro del sistema hematopoyético. Imagen modificada de (51).

Otra gran familia de lectinas animales son las Selectinas, las cuales están en la membrana de las células endoteliales, polimorfonucleares (PMN), linfocitos T y B circulantes, y natural killer (NK); su expresión depende de la etapa de desarrollo de la célula. Por ejemplo, la expresión de la selectina E es estimulada por la

interleuquina 1 (IL-1) y el factor de necrosis tumoral alfa (TNF α), lo que muestra su papel en procesos de inflamación agudos. Por su parte, las lectinas que se unen al ácido siálico y que son similares a inmunoglobulinas (Siglecs), también están presentes en leucocitos, oligodendrocitos y células de Schwann; se expresan luego del proceso de diferenciación hematopoyética a partir de los precursores linfoides y mieloides (Figura 6) (52,53).

2.4 Lectinas de origen Vegetal

En las plantas, la mayoría de estas moléculas están en los cotiledones y endospermos de las semillas constituyendo del 2 al 10 % del total de proteína (54). Algunas de las funciones que pueden tener estas proteínas están relacionadas con procesos de regulación fisiológica y defensa contra el ataque de microorganismos (55), almacenamiento de proteínas y transporte de carbohidratos entre otras (56). Otras lectinas se expresan en bajas concentraciones cuando las plantas se cultivan en condiciones normales de crecimiento, sin embargo, cuando la planta se expone a diferentes condiciones de estrés biótico se induce o aumenta su expresión, por lo tanto, hay lectinas que son inducibles (57).

Las lectinas de almacenamiento y las inducibles no muestran diferencias en sus estructuras tridimensionales, pero se localizan en diferentes compartimentos de la célula vegetal. Las primeras, se sintetizan con un péptido señal y, por lo tanto, siguen la ruta secretora desde el retículo endoplásmico hasta el aparato de Golgi, ubicándose en vacuola, membrana plasmática o siendo secretadas a la pared celular y al medio extracelular (34). Las lectinas inducibles se sintetizan sin un péptido señal y se ubican en el citoplasma o en el núcleo de la célula, por lo que están involucradas en interacciones endógenas específicas que intervienen en procesos de señalización y regulación (58). Además, participan en mecanismos de defensa de las plantas contra el ataque de patógenos y predadores (59), se expresan en muy bajas cantidades y algunas veces no se pueden detectar; contrario a lo que sucede con las lectinas de almacenamiento, que han sido de utilidad en el estudio de la estructura de glicanos, detección de nuevos grupos sanguíneos, como mitógenos de linfocitos, (60), purificación de glicoproteínas, marcadores tumorales, reparto de drogas y en el estudio de glicoproteomas de los organismos (61,62). Adicionalmente, las lectinas de *Artocarpus heterophyllus*, concanavalina A, *Galanthus nivalis*, *Musa acuminata*, *Myrianthus holstii*, *Narcissus pseudonarcissus* y *Urtica dioica* se han postulado como agentes antivirales (63).

Se han clasificado y agrupado en doce familias de acuerdo con el dominio de reconocimiento del carbohidrato (CDR) (34,37,64), algunas de ellas se describen a continuación:

2.4.1. Familia de Lectinas tipo ABA

La lectina de *Agaricus bisporus* (ABA) se caracteriza por tener un plegamiento único compuesto por dos hojas β conectadas por un motivo hélice-bucle-hélice. El monómero tiene dos sitios de reconocimiento con especificidades por dos azúcares diferentes, N-acetil-glucosamina (GlcNAc) y N-acetil-galactosamina (GalNAc), localizado en sitios opuestos (Figura 7), se asocia en forma tetramérica y reconoce glicotipe T α (Gal β 1,3GalNAc α 1-R) en O-glicanos, el cual se expone en células tumorales y en queratinocitos humanos con efectos anti-proliferativos, además estimulan células del sistema inmune (65,66).

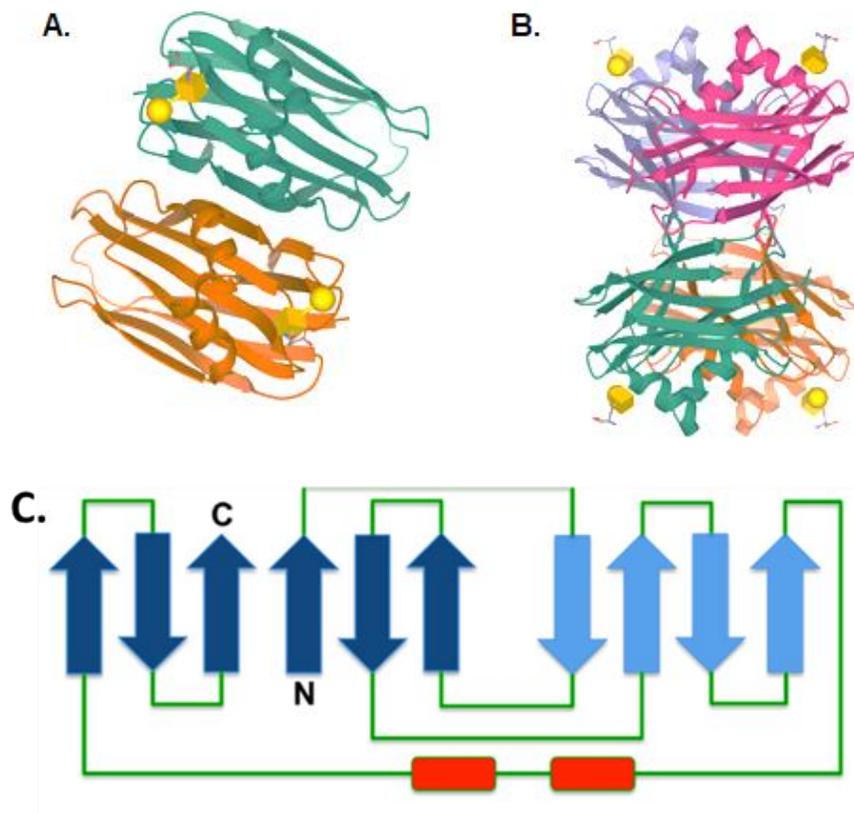


Figura 7. Lectina de *Agaricus bisporus* (ABL) con glicotipe T, tomado de PDB: <https://www.rcsb.org/structure/1Y2V>. La proteína es un tetramero (B), y cada monómero (A) presenta una estructura formada con dos láminas beta conectadas por un motivo hélice-bucle-hélice (67), (C) Topología estructural hélices α representadas en color rojo

2.4.2. Familia de Lectinas tipo heveína

El dominio Heveina es característico de la lectina aislada del árbol de latex (*Hevea brasiliensis*). Hay merolectinas que se caracterizan por ser monoméricas y presentar afinidad a oligómeros de GlcNAc, (quitina). Además tener actividad antifúngica (68), pueden presentarse de forma individual o como un conjunto de hasta 7 dominios dispuestos en tándem tipo quimerolectina con un dominio quitinasa hacia el extremo C-terminal (69,70). El plegamiento del dominio esta soportado por cuatro puentes disulfuro (69), donde su estructura tridimensional se caracteriza por contener tres hebras de hoja β y dos α hélices cortas (Figura 8). Una subunidad generalmente está constituida por cuatro dominios que comprenden 30 a 43 aminoácidos con residuos de glicina y cisteína conservados. Los puentes disulfuros son intracatenarios, resultando en estructuras secundarias que no son regulares, a excepción de una pequeña α hélice de cinco residuos, cada dominio presenta un sitio de reconocimiento a carbohidrato, y requiere iones metálicos (71).

Se han reportado lectinas con dominios tipo heveina en plantas y hongos (34). De igual modo, se resalta especificidad en el reconocimiento por N-acetil lactosamina, quitopentósidos o quitotetrosas para lectinas como *Datura stramonium* (DSA), *Solanum lycopersicum* (LEL); *Solanum tuberosum* (STL), aglutinina *Urtica dioica* (UDA), y *Triticum vulgare* (WGA) (72).

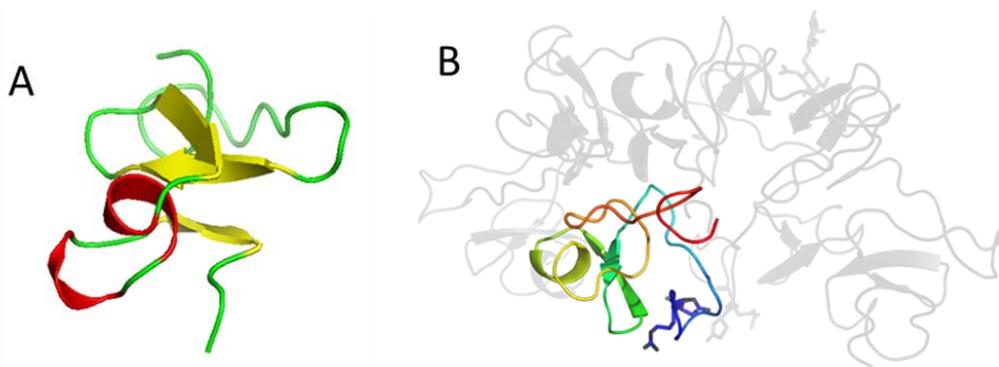


Figura 8. Familia tipo Heveina. **(A)** Plegamiento tipo Heveina caracterizado por presenta tres hebras plegadas β (amarillo), dos hélices α cortas (rojo), en verde regiones desordenadas. Código PDB: 1HEV. **(B)** Lectinas de *Triticum vulgare* (WGA). <http://scop.berkeley.edu/thumbs/1.61/k7/d1k7ta1/sl.png>.

2.4.3. Familia de Lectinas tipo Amaranto.

Las lectinas de tipo Amaranto fueron aisladas inicialmente a partir de semillas de *Amaranthus caudatus*, y se caracterizan por ser homodiméricas, con un peso molecular de 33 kDa y con afinidad por el antígeno $T_{\alpha}(\text{Gal}(\beta 1 \rightarrow 3)\text{GalNAc}\alpha 1 \rightarrow \text{Ser/Thr})$ monomérico o polivalente. Las subunidades de amarantina

comprenden dos dominios homólogos dispuestos en tándem de alrededor 150 aminoácidos. Cada dominio tiene un plegamiento de β -trefoil formado por dos hojas β antiparalelas, cubiertas por tres orquillas que se conectan a través de una α hélice 3_{10} en un barril corto (71); hay dos sitios de unión localizados en cada una de las dos subunidades hacia el extremo N-terminal y en la interfase (Figura 9). Las lectinas con dominio tipo amaranto se han reportado principalmente como hololectinas, sin embargo, en los últimos años, se han identificado secuencias homólogas a la amarantina fuera de la familia *Amaranthaceae*, lo que sugiere que la aparición de proteínas similares a la amarantina es mucho más amplia y que aún quedan por descubrir más proteínas similares en el reino vegetal.

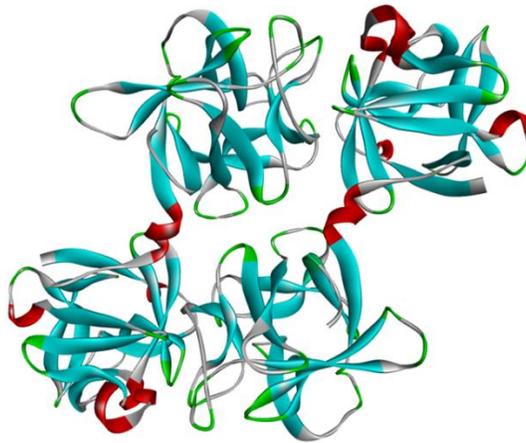


Figura 9. Lectina de *Amaranthus caudatus*. (PDB 1JLY, <https://www.rcsb.org/3d-view/ngl/1jly>). Hay dos sitios de unión específicos por el glicotipe T (Gal- β 1, 3-GalNAc- α -O-Ser/Thr). Cada dominio está constituido por dos hojas β plegadas formada por 6 hebras antiparalelas. En rojo, una región corta en forma de hélice 3_{10} que une los dos dominios de la subunidad. Estructura visualizada en Discovery studio.

Un estudio reciente, reveló 16 secuencias que contienen estos dominios en el genoma del pepino (*Cucumis sativus L.*), homólogos de amarantina del pepino poseen una disposición de dominio AAT, que consta de dos dominios de amarantina (el dominio AA) unidos a un dominio de toxina (el dominio T), con una secuencia similar a la aerolisina, la toxina formadora de poros (PFT) producida por *Aeromonas sp.* Una característica importante de estas proteínas formadoras de poros es que se sintetizan como proteínas solubles que posteriormente se oligomerizan y se convierten en poros transmembrana en la membrana diana (73) (Figura 10).

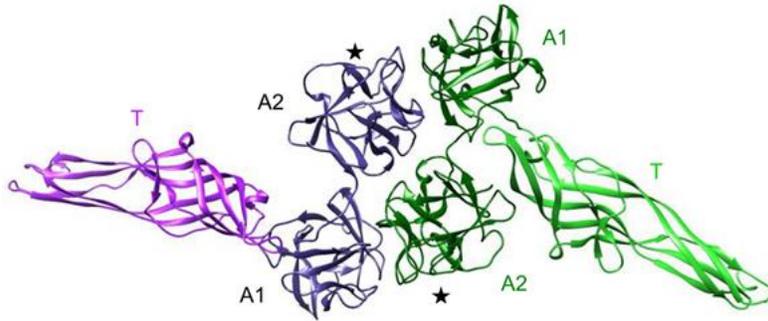


Figura 10. Estructura de AAT4. Monómeros, de color violeta y verde (dominios de amarantina, A1, A2), unidos a un dominio similar a la aerolisina (T). Los dominios A2 contienen un sitio de unión a carbohidratos (★) (57).

2.4.4. Familia de Lectinas tipo Jacalina

De las semillas de la fruta de jack (*Artocarpus integrifolia*) se extrajo la primera lectina de este tipo y es la proteína modelo para la extensa familia de lectinas relacionadas con Jacalina. Esta familia de lectinas se puede dividir en dos grupos en función de su especificidad de unión a carbohidratos. Las lectinas de unión a galactosa se ubican en la vacuola; mientras que las lectinas de unión a manosa, residen en el citoplasma y el núcleo. Son tetraméricas, glicosiladas, y cada subunidad tiene una cadena pesada (α) y una cadena ligera (β); constituidas por tres hojas β antiparalelas arregladas a manera de un prisma triangular (Figura 11) (57).

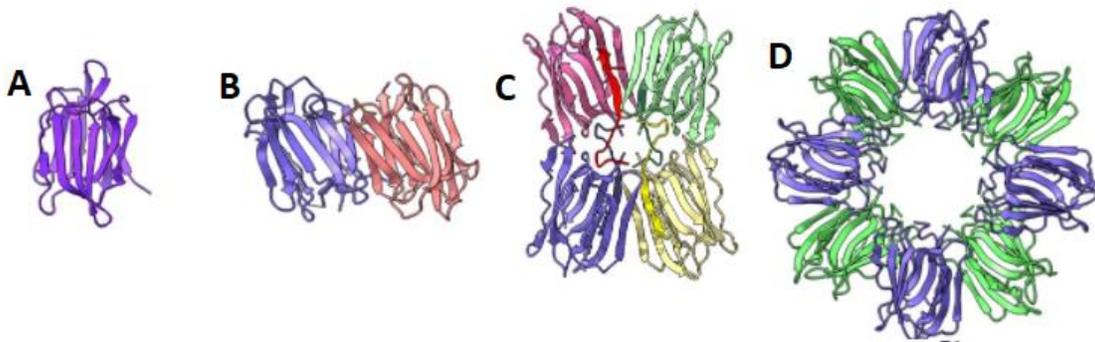


Figura 11. Lectinas tipo Jacalina Cada subunidad tiene un pliegue de prisma beta simétrico triple formado por tres láminas beta de cuatro hebras (A). El 62 % de la estructura cuaternaria de la lectina es estabilizada por interacciones hidrofóbicas. (A) monómero Jacalina (PDB:1JAC), (B) dímero Calsepa (PDB:1OUW), (C) tetramero Jacalina (PDB:1UGW) (D). octámero Heltuba (PDB:1C3K) (55).

Otras lectinas pertenecen a la familia de las monocotiledóneas (Figura 12), específicas por manosa, provenientes de orquídeas, ajo y amarilis, con secuencias de aminoácidos altamente conservadas. Son tetraméricas, cada monómero tiene un peso molecular de 12 kDa, constituidas por cuatro hojas β antiparalelas unidas entre sí por giros (71). Las lectinas de la familia de ricina (RIP), están constituidas por dos cadenas; A y B, unidas por dos puentes disulfuro. La cadena A es la responsable de la toxicidad, dada su actividad de N-glicosidasa sobre el ribosoma que inactiva la traducción, mientras que la cadena B posee actividad de lectina y reconoce una variedad de monosacáridos (GalNAc, GlcNAc, Gal y Man) (61, 62,63). Una lectina tipo RIP-2, extraída del látex de *Euphorbia tirucalli* es Eutirucallin, con subunidades de 32 kDa (74) y actividad antitumoral.

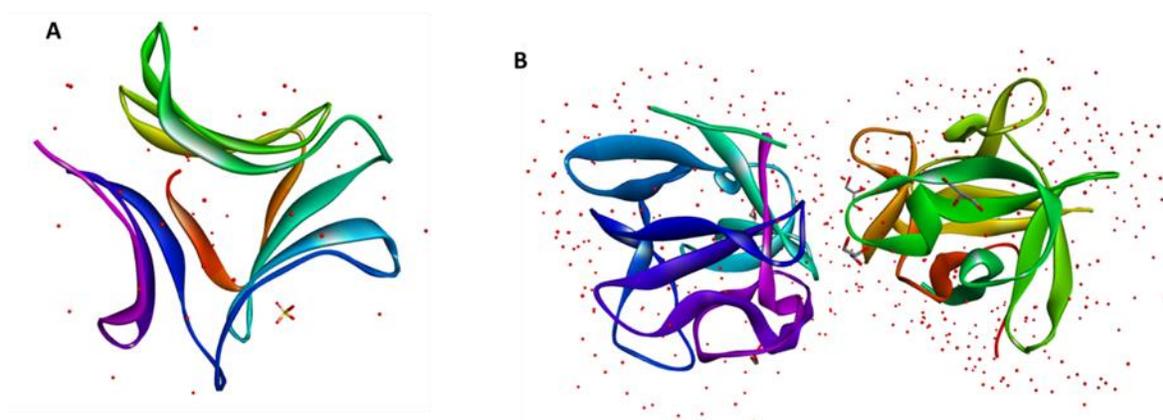


Figura 12. Estructura de las lectinas de la familia de las monocotiledóneas. **(A).** 1XD6 estructura cristalina de la lectina de unión a manosa de Gastrodia PDB DOI: 10.2210/pdb1XD6/pdb. Lectinas de la familia de ricina (RIP). **(B).** 4ION estructura cristalina *Macrolepiota procera* ricin B-like lectin (MPL), PDB DOI: 10.2210/pdb4ION/pdb. Estructuras visualizadas en *Discover studio* visualizer.

2.4.5. Lectinas tipo Leguminosas (Tipo L)

Este motivo se ha encontrado en lectinas de la familia Fabácea (Leguminosae), aunque se puede encontrar en otras familias botánicas, como las Lamiáceas. Generalmente se encuentran constituidas por dos o cuatro subunidades idénticas de 25 a 30 kDa, cada una contiene un sitio de unión (CDR) y algunas requieren iones Ca^{2+} , Mn^{2+} y Mg^{2+} para el reconocimiento de los carbohidratos. El plegamiento típico de los monómeros es llamado “Jelly roll” que se forma a partir de tres hojas β antiparalelas. En la figura 13 (A-B), se observa para la lectina de concanavalina A (ConA), una hoja β trasera plana formada por seis (6) hebras antiparalelas (color rojo), una curva delantera de siete (7) hebras (color verde) y una hoja β superior de cinco (5) hebras que las mantiene unidas. Se genera una estructura en forma de

domo con bucles localizados en la parte superior del monómero formando el dominio de reconocimiento de carbohidratos (CDR) (flecha) (75).

El número de hebras β que conforman las hojas β puede variar entre las lectinas; la hoja β frontal curva puede tener de 5-7 hebras, la plana posterior (trasera) de 4-6 hebras β ; conectadas en uno de los lados por otra hoja antiparalela superior de 5 hebras β . Las hojas frontal y posterior son casi paralelas y se pueden encontrar a una distancia cercana a 13 Å, según se ha determinado en alrededor de 300 estructuras de proteínas que contienen este tipo de plegamiento. Esta distancia es importante para la formación de un cluster de residuos hidrofóbicos cuyas cadenas laterales están posicionadas entre las hojas frontal y posterior dando un plegamiento estable. En conjunto, estas estructuras generan diversidad y especificidad por diferentes estructuras de carbohidratos; igualmente, algunas lectinas son tetraméricas (Figura 13 C-D), dada la asociación de dos dímeros alineados paralelamente por la hoja posterior (color rojo) (76).

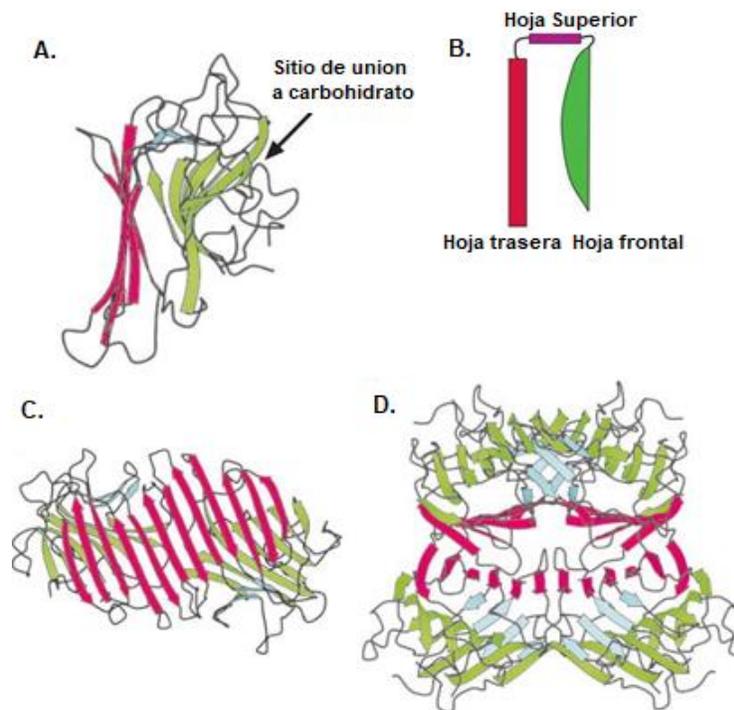


Figura 13. Lectina tipo L que correspondiente a Concanavalina A (Con A) (56), estructura clásica de lectinas de la familia Fabácea. Esta lectina es extraída de *Canavalia ensiformis*, y fue una de las primeras en ser caracterizada y conocida su estructura tridimensional. Reconoce específicamente α -D-glucosa y α -D-manosa, así como oligosacáridos ricos en manosa, uniendo glicoproteínas y glicolípidos (**A**) Monómero (**B**) Representación del plegamiento tipo Jelly roll (**C**) Forma dimerica, alineación paralela de dos monómeros, 12 hebras antiparalelas (**D**) Tetramero, se observa la asociación de dos dímeros que interactúan a partir de las hebras posteriores y hacia el exterior, están las hebras frontales y los sitios de unión a carbohidrato (CDR) (75).

Las lectinas aisladas de Fabácea están en mayor proporción en las semillas, sin embargo, se encuentran en raíz, tallo y hojas. Durante el reconocimiento del carbohidrato participan cuatro o cinco residuos de aminoácidos invariables, tales como ácido aspártico (Asp), asparagina (Asn) y un aminoácido aromático o leucina, incluso el Asp, reconoce el ion calcio (Ca⁺⁺), sin embargo, no todas las lectinas requieren calcio (Ca⁺⁺) para su actividad (77). Por su parte, para las lectinas aisladas de Lamiacea, no se han reportado estudios que describan su estructura y función. Hasta el momento, se han caracterizado las lectinas de *S. sclarea* (14), *Moluccella laevis* (78), *S. palifolia* (17), *Hyptis mutabilis* (17), *Lepechinia bullata* (5), *S. bogotensis* (4), *Gleheda* (3) y *S. miltiorrhiza* (56). Sus características bioquímicas difieren al igual que su especificidad, además se han estudiado muy pocas especies, aunque se destacan por su especificidad de reconocimiento de los glicotopes T α y Tn (Gal β 1,3GalNAc α 1-O-Ser/Thr GalNAc α 1-O-Ser/Thr) asociados a superficies celulares en tejidos tumorales (79). Estas lectinas son glicoproteínas con altos porcentajes de carbohidrato, básicas, se asocian preferentemente en formas diméricas y tienen en su estructura puentes disulfuro intra e intercatenarios (Tabla 2). Igualmente, se ha establecido la estructura primaria de las lectinas de *Gleheda* (Figura14-A) (3) (AAN05097.1 en GenBank) y de *S. miltiorrhiza* (ABU87404.1 en GenBank) (56) (Figura 15-A), a partir del ARNm. Las dos especies están clasificadas entre las “lectinas tipo Leguminosa”, y un alineamiento de estas dos secuencias mostró un 42 % de identidad (5).

Tabla 2. Principales características de las lectinas aisladas de Lamiaceas T/Tn

Lectina	Aglutinación	Peso molecular (kDa)		CHS (%)	pI	Secuencia N-terminal	Referencia
		Nativa	Subunidades				
<i>Lepechinia bullata</i>	T, Tn, conejo	--	93 */** 48*, 28*/**	13.4	8.5	N.D	(80)
<i>Hyptis mutabilis</i>	Tn	N.D	55,51	27,7	7.5 y 6,7	N.D	(13,14,17)
<i>Glechoma hederacea</i>	T, Tn	85 \pm 80- 93 $\pm\pm$	26, 28	10	N.D.	--KTTH-- FAVPPALT ALFQGDAF -----	(3)
<i>Molucella laevis</i>	Tn	130	26, 28, 42	10	N.D.	N.D.	(78)
<i>Salvia miltiorrhiza</i>	Conejo y ratón	N.D	26.3	N.D	4.76	MAKLLQNL IPLLSAIVLL LA--	(2)
<i>Salvia sclarea</i>	Tn	50	35	15	8.8 - 8.0	ND	(13,14)
<i>Salvia palifolia</i>	Tn	N.D	58, 64	23,8	7,4 y 6,8	N.D	(13,14,17)
<i>Clerodendrum trichotomum</i>	ND	56	28	16	N.D	N.D	(81)

N.D. No determinado. *. **condiciones no reductoras. \pm Filtración en gel superosa. $\pm\pm$ Velocidad de sedimentación. $\pm\pm$ Banda menor, glicofoma. CHS: Carbohidratos

La lectina recombinante de *Gleheda* mostró 30 % de identidad y 50 % de similitud con lectinas de Leguminosa como *Erythrina corallodendron*, *Robinia pseudoacacia*,

Glycine max, *Pisium sativu*, *Lathyrus ochrus* y *Cytisus sessilifolius* (3), además, presentó actividad insectistática sobre las larvas del escarabajo de la patata (*Leptinotarsa decemlineata*) (82), lo que indica que estas lectinas pueden tener funciones similares a las descritas para las de la familia fabácea (3). Esta lectina reconoce el antígeno Tn en una forma similar a la establecida en *Vicia villosa* *isolectina* B4 (VVLB4); el reconocimiento se da a través de una red de puentes de hidrógeno con Asp, Gly y Asn. La lectina de Gleheda reconoce α GalNAc mediante los grupos (OH) en posiciones 3, 4, 5, 6 y el grupo N-acetil con los aminoácidos Asp73, Gly93, Asn117, Tyr201 y Asn202, señalados en rojo en la Figura 14-A. His115 presenta interacciones hidrofóbicas con el anillo piranosa del azúcar (3). De igual forma, en la lectina de *S. miltiorrhiza* (SML) son importantes Asp107, Ala139, Asn151, Gly234, Gln235 e His236 en la interacción con el carbohidrato, señalados en rojo en la Figura 15-A (2); correspondiendo con la secuencia de LSBo-I donde los aminoácidos Asn149 y Gly 236 están en posiciones cercanas, y podría sugerir que son residuos importantes en el reconocimiento del antígeno Tn (5).

En general, el reconocimiento del glicotopo Tn (GalNAc α 1-O-Ser / Thr) por las lectinas vegetales no difiere del reconocimiento de GalNAc, adoptando una orientación similar a la determinada para otras lectinas específicas de Gal/GalNAc. Una red similar de puentes de hidrógeno, O3, O4, O6, y el grupo acetilo interactúa con los residuos que forman el bolsillo de unión a carbohidratos. Sin embargo, la especificidad por Tn se da por puentes de hidrogeno adicionales que se establecen con un residuo aromático cercano, una molécula de agua ubicada en la proximidad del sitio de unión, y/o el residuo de serina o treonina. A su vez, dependiendo de las lectinas, el bolsillo de unión a carbohidratos acomoda preferentemente el residuo de GalNAc, o los residuos de Gal, tras la unión del glicotopo T (Gal β 1,3GalNAc α 1-O-Ser/Thr); la otra unidad de azúcar sobresale del bolsillo de unión a carbohidratos, pero generalmente interactúa con otros residuos de aminoácidos donde la especificidad de unión depende de las interacciones adicionales, dadas también con la molécula de agua que potencia la unión.

De igual forma, las lectinas se diferencian en el tamaño del sitio de unión, por ejemplo, la amarantina acomoda el disacárido. Sin embargo, es importante resaltar que *in vivo* las lectinas muestran una afinidad alta por glicanos complejos que contienen 3 a 7 residuos de monosacáridos y que incluyen estructuras ramificadas o modificaciones, tales como O-acetilación, sulfatación o fosforilación. Por ello, la visión apoyada en la inhibición de lectinas usando monosacáridos y glicanos simples ha evolucionado hacia la elucidación de las interacciones biológicas y fisiológicas entre lectinas y glicanos complejos existentes en la naturaleza. Dichas interacciones están influenciadas por la forma como se presenta el glicotopo del glicano (de forma monovalente, polivalente, en cluster); además, la presentación de

glicanos asociados a membrana es más restringida que la de los glicanos en solución, debido a las restricciones de orientación impuestas al glicoconjugado a través de sus interacciones laterales con otras moléculas, como lípidos y proteínas de la membrana (83). Por consiguiente, la forma de interacción de cada lectina es única y se destaca la importancia de las moléculas de agua e interacciones con aminoácidos en la vecindad del sitio de unión que contribuyen en la discriminación de la diversidad de glicanos (84).



Figura 14. Lectina de *Glechoma hederácea* (Gleheda) (A) Estructura primaria (B) Predicción de la estructura terciaria de la secuencia de aminoácidos (1) Vista posterior, hoja β aplanada de seis hebras. (2) Vista frontal, hoja β curva de siete hebras, (3) Vista lateral, espacio entre las hojas posterior y frontal. (★) indica los posibles sitios de unión a carbohidratos Modificado de (3)

Respecto a la estructura tridimensional, solo se han propuesto predicciones con diferentes herramientas bioinformáticas; tanto para las lectinas de Gleheda, como de *S. miltiorrhiza*, muestran el plegamiento típico de las lectinas de tipo leguminosa. Por ejemplo, para la lectina de *S. miltiorrhiza*, el modelo tridimensional usando SWISS-MODEL (<http://swissmodel.expasy.org/>), presentó una hoja β frontal curva de siete hebras antiparalelas, una hoja β posterior de seis hebras, unidas por bucles; además, una tercera hoja β de cuatro hebras (S) ubicada en la parte superior del monómero entre las dos hojas. Adicionalmente, se observó la unión de iones de calcio (Ca^{++}) y manganeso (Mn^{++}), que ayudan a la interacción con residuos del carbohidrato (Figura 15-B). El modelo concuerda con el mencionado previamente para lectinas de leguminosa, donde hay solamente hojas β y ausencia total de α -hélices, así como bucles y regiones al azar; dando como resultado una estructura de β sándwich (2). En la figura 14-B también se presenta la estructura terciaria

predicha para la lectina Gleheda; en su parte posterior, muestra una lámina formada de seis hebras y la lámina frontal siete hebras; las dos laminas interconectadas por bucles y vueltas, en la parte superior del monómero la hoja S formada por cuatro hojas β que ayuda a mantener juntas las hojas delantera y trasera (3).



Figura 15. Lectina de *S. miltiorrhiza* (SmL) (A) Estructura primaria (B) Predicción de la estructura terciaria de la secuencia de aminoácidos (1) Hoja β aplanada de seis hebras. (2) Hoja β de siete hebras, (3) CDR, sitio de unión a carbohidratos. En rojo se muestran los posibles sitios de unión a carbohidratos. Modificado de (2)

En las especies endémicas, *L. bullata* y *S. bogotensis*, se llevó a cabo la caracterización bioquímica completa de las lectinas específicas por T/Tn, a partir de la proteína nativa. A su vez, se detectó otra lectina específica por residuos de manosa/glucosa (5). En la tabla 3, se comparan las dos lectinas encontradas para *S. bogotensis*, designadas como LSBo-I y LSBo-II. Las lectinas se diferenciaron en sus propiedades bioquímicas, actividad y especificidad, siendo LSBo-II similar con aquellas encontradas en varias especies de plantas de la tribu Diocleae, que muestran especificidad hacia manosa/glucosa (Man/Glc) y han sido purificadas por afinidad usando resinas basadas en dextrano, y su actividad aglutinante sobre eritrocitos puede ser inhibida por manósidos o glucósidos. Estas lectinas muestran entre sí altos porcentajes de identidad de sus secuencias (80-99%), en comparación con lectinas Man/Glc de otras plantas Fabaceae de diferente tribu, como aquella de *P. vulgaris* de la tribu Phaseolinae.

Tabla 3. Características moleculares para lectinas aisladas de las semillas de *S. bogotensis* (5).

Propiedades moleculares	LSBo-I	LSBo-II
Monómero	38702 Da*	30 – 33 kDa**
Bandas en SDS – PAGE (kDa)	72.6; 38.8; 35.7**	30,60**
% Azucres	16%***	Glicoproteína*, ND
PI	8.6 – 8.8	6.5
N-terminal	AIVTSFXYTSFSSSSDIKLQ	ADTIVAVELD

*Determinado por ES-MS, **por SDS PAGE, *** Cuantificado por el método de Dubois. # Aislada por Sepharosa ConA, demostrado presencia de oligosacáridos de Tipo N altos en manosa reconocidos por esta lectina.

Para la LSBo-I se determinó previamente el 62 % de la secuencia (5) (Figura 16), no obstante, se debe alcanzar el 100 %, y de esta forma estudiar sus mecanismos de acción a nivel molecular y el estudio de las interacciones con los glicanos de células tumorales. Igualmente, se han realizado estudios de histoquímica e inmunohistoquímica en células tumorales y tejidos para la detección de glicotopes T/Tn, se han preparado anticuerpos policlonales en conejo (85) y en gallina (86,87) que reconocen la LSBo-I, con los cuales, mediante ensayo de ELISA se evidenció que reconocen algunas lectinas de Leguminosas, lo que probablemente muestra la presencia de epítopes de secuencia y conformación similares (87) en lectinas lamiácea y fabáceas.

Otros trabajos realizados por cromatografía de afinidad con LSBo-I permitieron aislar glicoproteínas de citoplasma con alta densidad de glicotope Tn a partir de las líneas celulares de HeLa (cáncer de cérvix) y MCF-7 (cáncer de seno). Estas proteínas fueron reconocidas mediante Western blotting por anticuerpos policlonales dirigidos contra la mucina de tipo 2 (MUC-2) (88), componente principal del moco en el colon, que sufre alteraciones en los adenocarcinomas (89). Recientemente se postuló la detección y cuantificación de MUC2 en suero de pacientes con cáncer de seno como un método diagnóstico, pero sin correlación alguna como método pronóstico o de supervivencia (89). En cáncer colorrectal la no detección de MUC2 en tejidos se correlaciona con un mal pronóstico, mientras que la expresión de MUC5AC y MUC6 es asociada con un pronóstico favorable de la enfermedad en etapas II y III (90). Además, son de utilidad en la identificación de biomarcadores basados en glicofomas de mucinas, como el STn- MUC2 que se encuentra preferentemente en cáncer de mama mientras que en ovario STn-MUC-I (91). Los cambios en la síntesis normal de oligosacáridos conducen a la aparición de diferentes patologías, como desórdenes tumorales (92–97). Algunas de estas glicoproteínas son empleadas como marcadores histológicos o tumorales y permiten monitorear procesos metastáticos (98–101), por lo que LSBo-I tiene un gran potencial como alternativa para la detección de glicanos presentes en células

tumorales y podría ser de gran utilidad en el diagnóstico o en posibles terapias contra el cáncer (102,103), de ahí la importancia de su estudio.

10	20	30	40	50	60
MATLLQNL	LT	LDNTRxSxTL	LTHPNFN	GHGQHGS	QSPSPSRFx
70	80	90	100	110	120
LTGALAQK	TSVHPN	LEGGDAFE	GPKxFxITxx	xxxxxADGxA	xFxxPVGxxx
130	140	150	160	170	180
PSGSTGGY	AVNPKGxGxx	xQGVDA	xGSYGS	NHGFCGS	SNSELVQSSG
190	200	210	220	230	240
LGHQESR	YxxxxxYEEL	QVTAGxRKxx	xxSYEAELDD	FVTLQVT	SAATSGD
250	260	270	274		
KDIxxWx	xMxxxxxxx	xxxxxxGxxI	RQxV		

Figura 16. Secuencia parcial de LSBo-I, los residuos resaltados en rojo corresponden a aminoácidos importantes para el reconocimiento de los antígenos T y Tn. (X) indica la posición en la que no se identificó aminoácido. La secuencia propuesta previamente tiene 274 residuos. 216 aminoácidos fueron ubicados o asignados en los alineamientos, lo que corresponde al 62 % del total de los 347 residuos reportados para LSBo-I (4,5).

Con la secuencia parcial de LSBo-I se realizó una aproximación a la estructura terciaria mediante análisis bioinformático, a pesar de que aún faltan aminoácidos por asignar a la secuencia, el modelo mostró homología con la estructura de lectinas de tipo leguminosa al presentar nueve hebras antiparalelas (Figura 17-A) unidas entre sí por estructuras random coil. Igualmente, presentó dos regiones con tendencia a ser hélice α (flechas rojas): la primera, en el amino terminal (azul) y la segunda, ubicada entre las hebras 8 y 9 (anaranjado). Se encontró similitud estructural con la lectina de *V. villosa* isolectina B4 (VVB4) y *S. miltiorrhiza*, específicas por el antígeno Tn, lo que permitió observar la disposición de las hebras β en la estructura, así como la localización del sitio de reconocimiento a carbohidrato (Figura 17-B). Al completar la asignación de aminoácidos a la estructura primaria, se podría alcanzar un mejor ajuste de los modelos tridimensionales (5).

Los cambios en la estructura de los carbohidratos expuestos en la superficie celular que conforman las glicoproteínas presentes en diferentes tejidos usualmente están relacionados con el desarrollo de diferentes enfermedades. Como se ha mencionado, un interés especial se ha dado a los glicotopes T, su precursor Tn y sus formas sialiladas, así como al estudio de las lectinas que los reconocen de forma específica como las presentes en especies de la familia Lamiácea. A pesar de la importancia que tienen las lectinas de Lamiaceae por el reconocimiento los

antígenos T/Tn, la información estructural es muy escasa para entender su mecanismo de interacción con los carbohidratos (104).

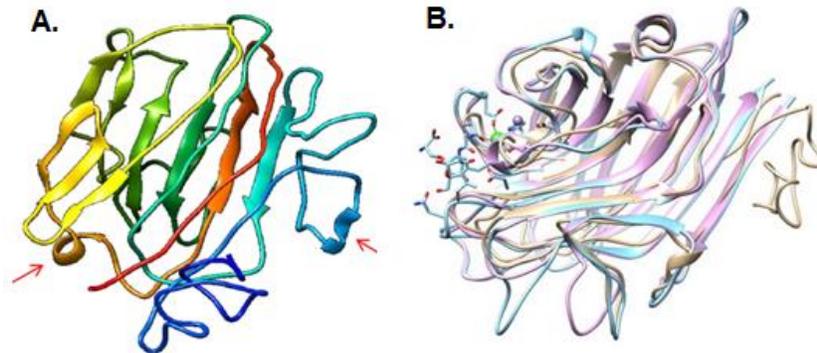


Figura 17. Predicción de la estructura tridimensional de LSB0-I. (A) monómero del LSB0-I (B) Superposición de lectinas específicas por T/Tn. Para LSB0-I (color beige) se observaron regiones con estructura secundaria aleatoria en donde SML-II (azul claro) y VVB4 (Magenta) muestran hebras β , el análisis se realizó sobre la secuencia con una cobertura del 62% (5).

Actualmente, se conocen 19 lectinas de diferentes familias que reconocen Tn, (Amaranthaceae, Fabácea, Moraceae, Lamiácea, Orchidaceae) (11,104); así como los estudios de secuencia y estructura tridimensional. En este sentido, hay mayor número de lectinas de especies estudiadas en las demás familias con respecto a las lectinas de la familia de las Lamiáceas.

Para profundizar en el conocimiento de las lectinas de lamiáceas de especies endémicas, se requiere buscar nuevas alternativas de purificación de las proteínas nativas o la determinación del cDNA para establecer su secuencia y obtener lectinas recombinantes. La cantidad de lectina que se encuentra en las semillas es baja comparada con especies de leguminosas; igualmente la obtención de las lectinas a partir de los extractos crudos es muy dispendiosa y en la mayoría de los casos se obtiene la lectina denaturada. El aislamiento en una cantidad suficiente para su caracterización molecular y estructural presenta varias dificultades, por lo tanto muy pocas lectinas de especies endémicas han sido purificadas y caracterizadas bioquímicamente (1,4,14). Sin embargo, la secuencia de las lectinas de *Gleheda* (3) y *S. miltiorrhiza* (2) fueron obtenidas mediante clonación molecular a partir de semillas y hojas respectivamente. En conformidad con lo descrito, es importante buscar alternativas de aislamiento o enriquecer las secuencias del ADN de estas especies y de los genes que codifican las lectinas.

La clonación y expresión de lectinas se han convertido en una alternativa para llevar a cabo estudios de estructura y función, debido a que permite tener moléculas puras “homogéneas” y en cantidad suficiente, con secuencias de aminoácidos definidas y

propiedades más precisas. Adicional a esto, la expresión en huéspedes heterólogos ha permitido evaluar la influencia de diversas modificaciones postraduccionales en el plegamiento y el direccionamiento de esta clase de proteínas de unión a carbohidratos (105). Las lectinas de *V. macrocarpa* (105) y *S. miltiorrhiza* (2) específicas por el glicotipe Tn han sido obtenidas por clonación molecular.

2.5. Familia Lamiaceae – *Salvia bogotensis*

Es una planta endémica de Colombia, que se encuentra ampliamente distribuida en toda la cordillera oriental. Son arbustos de tamaño pequeño o mediano, con hojas típicamente estrechas, inflorescencias moderadamente cortas y flores con color azul intenso (Figura 18). La distribución de esta especie abarca una amplia serie de territorios áridos o semiáridos en los departamentos de Boyacá y Cundinamarca, desde el Sur de la Sabana de Bogotá (Mosquera), hasta la región de Chita y Cocuy en Boyacá, donde asciende a mayor altitud, presentando formas con indumento foliar más denso en ambas caras. Los sectores intermedios donde también se encuentra esta especie son la región de Samacá y Tunja-Cucaita, el sector Tibasosa, Paz del Río, Santa Rosa, Floresta y la región del lago de Tota. Ocasionalmente esta entre las plantas medicinales que se venden en las plazas de mercado, con el nombre genérico de Salvia (106).

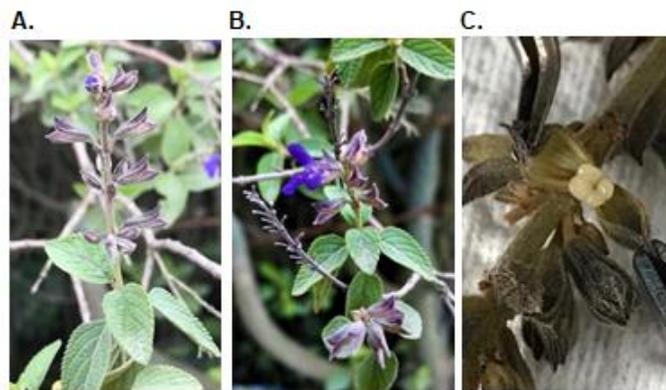


Figura 18. *Salvia bogotensis*, foto tomada en el sector en Mondoñedo, vía la Mesa-Mosquera, donde está ampliamente distribuida. **A.** Flores cerradas donde se almacenan las semillas y presentan la forma característica de labios que le dan su nombre característico. **B.** Flor abierta. **C.** Semillas verdes de la planta dentro de la flor. Imagen propia

Por su parte, la familia *Lamiaceae* está distribuida en todo el mundo, abarcando cerca de 7.000 especies (18). Por la disposición de sus flores son plantas ornamentales (Figura 19). Se han descrito alrededor de 190 especies de *Lamiaceae* en Colombia, siendo *Salvia* el género más diverso, con 75 especies (107). Se

pueden encontrar en casi todos los ecosistemas terrestres, aunque son más abundantes en bosques de clima templado y en zonas montañosas. En el mundo existen seis regiones de alta diversidad, Mediterráneo y Asia Central, África y Madagascar, China, Australia, Sudamérica y Norteamérica. Se han estudiado principalmente los aceites esenciales de romero, tomillo y salvia como conservantes de alimentos. Desde otra perspectiva, plantas como romero, tomillo, orégano y salvia son remedios tradicionales que se emplean para el tratamiento de gastritis, infecciones, dermatitis, bronquitis e inflamación (108,109). Estudios en especies de salvia, demuestran que los efectos farmacológicos de sus aceites esenciales se basan en la presencia de más de 100 compuestos activos, que pueden clasificarse en hidrocarburos monoterpénicos, monoterpenos oxigenados, hidrocarburos sesquiterpénicos, diterpenos, compuestos no isoprenoides y sesquiterpenos oxigenados (110).

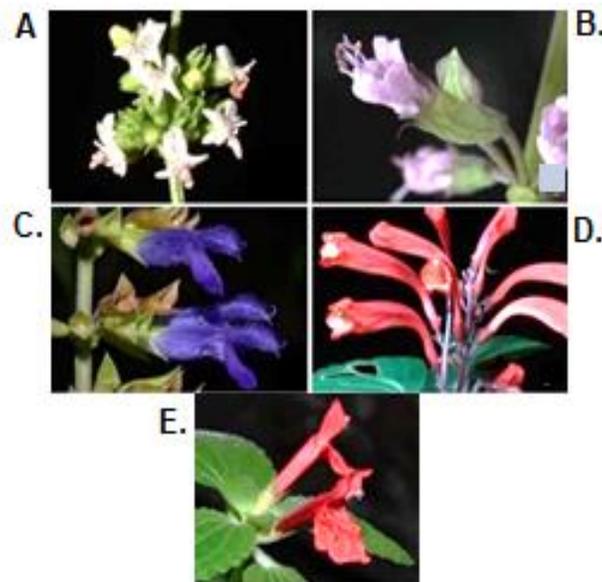


Figura 19. Algunos géneros de Lamiaceae representativos. A) Hyptis (*H. verticillata* Jacq.), B) Ocimum (*O. carnosum* (Spreng.) Link & Otto ex Benth.), C.) Salvia (*S. mexicana* L.), D) Scutellaria (*S. formosa* Leonard), E) Stachys (*S. coccinea* Ortega). Modificado de (111).

3. Justificación

Por analogía con el genoma, proteoma y transcriptoma, el glicoma es el grupo completo de glicanos y glicoconjugados que son sintetizados por una célula u organismo bajo unas condiciones fisiológicas específicas (112,113). Por lo tanto, es importante establecer la función de los carbohidratos que codifican la información conducente a las interacciones receptor-ligando, macromolécula-célula, célula-célula y célula-matriz extracelular, que se manifiestan en el establecimiento intracelular de asociaciones multiméricas, así como en la adhesión y reconocimiento intercelular. La diversidad de carbohidratos que participan en dichas interacciones es evidente dado que están presentes en glicoproteínas, glicolípidos y proteoglicanos de la superficie celular o matriz extracelular que actúan como receptores de moléculas que desencadenan diferentes procesos biológicos. El patrón de glicosilación de las proteínas es dependiente del tipo de célula en que se produce y puede modificarse ante diversas situaciones biológicas o ante la presencia de enfermedades; estas observaciones han llevado al desarrollo de la glicopatología, y en este contexto las lectinas han sido de gran utilidad.

Los cambios en la estructura de los carbohidratos expuestos en la superficie celular que conforman las glicoproteínas presentes en diferentes tejidos usualmente están relacionados con el desarrollo de diferentes enfermedades. Un interés especial se ha dado a los glicotopes T, su precursor Tn y sus formas sialiladas (sTn), así como al estudio de las lectinas que los reconocen de forma específica como las presentes en especies de la familia Lamiácea. A pesar de la importancia que tienen las lectinas de Lamiaceae por el reconocimiento de los glicotopes T/Tn, la información estructural es muy escasa para entender su mecanismo de interacción con los carbohidratos (104), comprender cómo podrían interactuar con otras glicoproteínas ricas en glicotopes Tn, T, sTn, como son las mucinas epiteliales o glicoproteínas.

Actualmente se conocen muy pocas lectinas que reconocen solo estas estructuras, así como los estudios de secuencia y estructura tridimensional, en ese sentido hay mayor número de lectinas de otras especies estudiadas respecto a las lectinas de la familia de las Lamiáceas. Para profundizar en el conocimiento de las lectinas de lamiáceas de especies endémicas, se requiere buscar nuevas alternativas de purificación de las proteínas nativas o la determinación del cDNA para establecer su secuencia y obtener lectinas recombinantes. La cantidad de lectina que se encuentra en las semillas es baja comparada con las semillas de leguminosas; igualmente la obtención de las lectinas a partir de los extractos crudos es muy dispendiosa y en la mayoría de los casos se obtiene la lectina denaturada. El

aislamiento en cantidad suficiente para su caracterización molecular y estructural presenta varias dificultades; por lo tanto, muy pocas lectinas de especies endémicas han sido purificadas y caracterizadas bioquímicamente (1,4,14). Sin embargo, la secuencia de las lectinas de *Gleheda* (3) y *S. miltiorrhiza* (2) fueron obtenidas mediante clonación molecular. En conformidad con lo descrito, es importante buscar nuevas alternativas de aislamiento o enriquecer las secuencias del ADN de estas especies y de los genes que codifican las lectinas.

De esta forma se puede establecer la estructura terciaria para entender el mecanismo de interacción y de selección de glicotopes específicos presentes en células transformadas, además de comprender su actividad biológica. Por lo tanto, este trabajo se encaminó en la obtención y estudio de la estructura primaria de la lectina de *S. bogotensis* (LSBo-I) a partir del material vegetal, por técnicas moleculares, que permitieran llevar a cabo estudios *in silico* de estructura tridimensional e interacción con el determinante Tn y estructuras relacionadas.

4. Objetivos

4.1. Objetivo General

Establecer la estructura primaria de la lectina de *Salvia bogotensis* (LSBo-I)

4.2. Objetivos específicos

- Obtener la proteína nativa por cromatografía de afinidad utilizando el anticuerpo (anti LSBo-I) dirigido contra la lectina de *Salvia bogotensis* (LSBo-I).
- Evaluar las mejores condiciones de extracción de ARN total a partir de material vegetal (semillas u hojas) para obtener el ADNc.

5. Metodología

5.1. Purificación de lectina de *Salvia bogotensis* (LSBo-I)

5.1.1. Métodos generales

5.1.1.1. *Obtención del material vegetal*

Previamente se hizo la recolección del material que contenía las semillas de *S. bogotensis* inmaduras (color verde) y maduras (color café) en la región de Mondoñedo (Cundinamarca) Vía Mosquera – Soacha (Figura 20). La especie fue identificada por el Dr. José Luis Fernández, y se dejaron testigos en el instituto de Ciencias Naturales (ICN) quedando registrados con los voucher COL-422764, COL-368928 y COL-420421. El estudio de la especie está enmarcado en el contrato de acceso a recurso genético otorgado por el Ministerio de Ambiente y Desarrollo Sostenible número 246, expediente RGE023.

Las semillas se aislaron de las flores y se seleccionaron aquellas que presentaban coloración café oscura (semillas maduras), así como las que tenían color verde (semillas inmaduras). Estas últimas fueron colectadas y almacenadas en nitrógeno líquido para evitar la oxidación. Posteriormente, las semillas maduras fueron molidas para obtener su harina, que fue almacenada a temperatura ambiente para posterior extracción de proteínas.

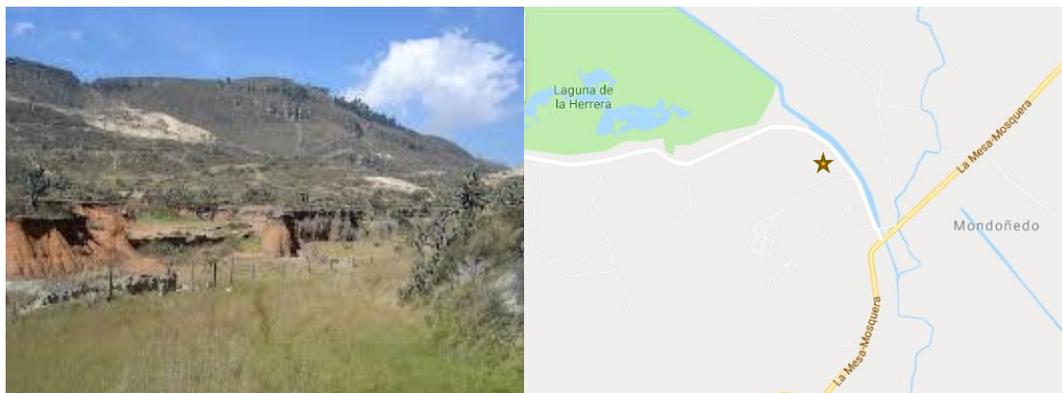


Figura 20. Sitio de colecta de las semillas de *S. bogotensis*, Mondoñedo, vía la Mesa-Mosquera coordenadas 4°41'05.6"N 74°15'13.2"W, el terreno es una montaña árida y seca que está al costado de la carretera.

5.1.1.2. Cuantificación de proteínas

Se determinó la concentración por el micrométodo del ácido bicinconínico (BCA) descrito por Smith (1985) (98,99), como patrón se empleó una solución de Albúmina Sérica Bovina (BSA).

5.1.1.3. Electroforesis unidimensional en geles de poliacrilamida

Se realizó electroforesis según los métodos descritos por Schägger y Von Jagow (115) en geles del 12,5 % o el método Laemmli empleando geles del 12.5% (116). Las muestras fueron preparadas con calentamiento (Δ) durante 10 minutos a 98 °C. Los geles fueron fijados y teñidos con azul de Coomassie R-250 o tinción de plata (117) siguiendo los protocolos disponibles en el laboratorio del grupo de investigación

5.1.1.4. Diálisis

Las membranas (Spectrum™) (MW Cutoff 10 kDa) se activaron con Na₂CO₃ 0,1M en ebullición durante 20 minutos, luego se incubaron en etanol al 50 % durante 30 minutos. Finalmente, fueron lavadas con abundante agua destilada. Durante el proceso de diálisis se realizaron dos cambios con agua destilada cada 8-12 horas y uno con PBS-tiourea 5 mM o bicarbonato de amonio 20 mM para las proteínas puras.

5.1.1.5. Ensayos de Aglutinación

Los eritrocitos Tn se prepararon por degradación enzimática a partir de una suspensión de eritrocitos A+ al 2,5 % con 0,05 unidades de Neuraminidasa (N 2133 Sigma-Aldrich), durante 1 hora a 37 °C para exponer el antígeno T y eliminar el ácido siálico (SiA). Las células empaquetadas se lavaron tres veces con PBS 1X, posteriormente se hizo la degradación con 30 unidades de β -galactosidasa para exponer el antígeno Tn, residuos de N-acetilgalactosamina (GalNAc) unidos a Ser/Thr (118,119). La lectura de resultados se estableció a partir de una escala cualitativa de +1 a +4; siendo +1 una aglutinación débil y +4 la máxima aglutinación, comparada con el control (4).

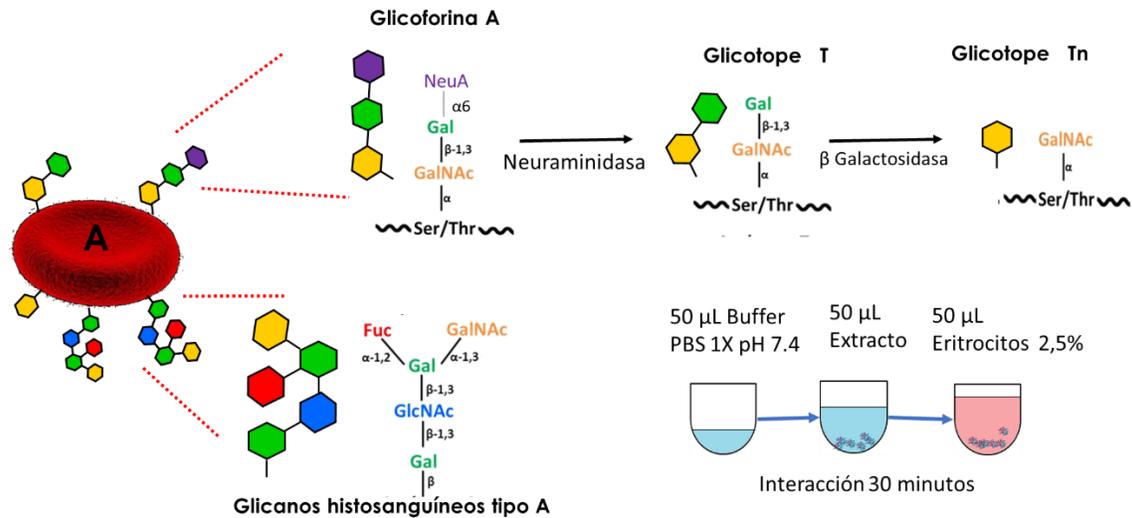


Figura 21. Degradación enzimática con neuraminidasa y β galactosidasa en eritrocitos tipo A para exponer los glicotopes T y Tn para ensayos de eritroaglutinación.

5.1.2. Purificación de inmunoglobulinas g (IgG) anti-LSBo

Previamente se llevó a cabo la inmunización de conejos New Zeland para producir anticuerpos policlonales contra la lectina LSBo-I (85). La purificación de los antiseros se realizó siguiendo la metodología descrita por Baines y Thorpe a partir de un pool de 10 mL de los antiseros, donde se mezclaron 4 mL del conejo 1, 3 mL conejo 3 y 3 mL conejo 4. El título de anticuerpos fue similar para cada uno de los animales inmunizados (Anexo A).

5.1.2.1. Precipitación de proteínas con PEG 8000

10 mL del pool de antiseros fue precipitado con polietilenglicol 8000 20 % (p/v) de (83271Sigma-Aldrich). 2 g de PEG 8000 fue disuelto en 10 mL de agua destilada, la disolución se enfrió a 4 °C y se adicionó gota a gota con agitación suave a 4 °C. El precipitado se obtuvo por centrifugación a 7000 g durante 10 minutos a 4 °C y resuspendió en buffer de fosfatos 70 mM (Na_2HPO_4 y NaH_2PO_4) pH=6,5.

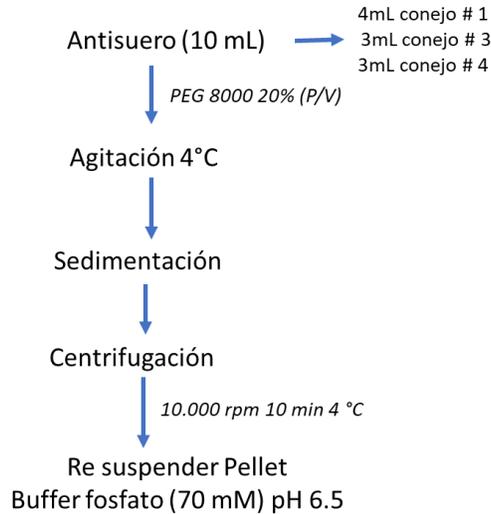


Figura 22. Esquema de purificación de IgGs anti LSBo-I con precipitación empleando polietilenglicol 8000 (PEG 8000).

5.1.2.2. Cromatografía de intercambio iónico

Se usó soporte de DEAE Sephacel™ (GE Healthcare); 5 mL del soporte se equilibró en buffer fosfatos (50 mM) pH 6,5. La fracción no retenida (FNR) se eluyó con el mismo buffer, mientras que la fracción retenida (FR) con buffer fosfatos conteniendo NaCl 1 M. La fracción no retenida se concentró por ultrafiltración (UF) en Amicon con nitrógeno gaseoso (MW cut off 10). Las fracciones de inmunoglobulinas (Iggs) concentradas fueron evaluadas por SDS PAGE y ensayos de ELISA (85). Para cuantificar las fracciones inmunoglobulina G pura (IgG) se empleó el coeficiente de extinción molar, donde la absorbancia (A) a 280 nm, para una solución de 1 mg/mL es igual a 1,38 (120,121). Para algunas fracciones se empleó como método alternativo BCA (ácido bicinconínico) (114).

5.1.2.3. Preparación del soporte de Afinidad Sepharosa 4B-IgG

El soporte para la cromatografía de afinidad (CA) se preparó a partir de Sepharose 4B activada con bromuro de cianógeno (CNBr) (C9142 Sigma-Aldrich). Se pesaron 0,6 g de resina liofilizada y fueron hidratados en HCl 1 mM durante 30 min, posteriormente se realizó un lavado en batch con 200 mL del ácido (HCl), el cual se adicionó con agitación suave. Luego se realizaron tres lavados más con una solución de bicarbonato de sodio 0.1 M (NaHCO₃, NaCl 0,5 M) pH = 8,3. Una vez equilibrado el soporte se adicionó 2,5 mL de IgG (6,67mg) previamente equilibrada en la misma solución de pH 8.3 y se dejó en agitación constante toda la noche a temperatura ambiente (RT). La IgG que no se acopló se recuperó por filtración al

vacío, se dializó y liofilizó para determinar su concentración final y calcular el porcentaje de acoplamiento.

Al gel acoplado se le hicieron dos lavados usando la misma solución de acople; los sitios activos no acoplados fueron bloqueados con una disolución de etanolamina 1 M pH = 8,8 durante 90 minutos. Por último, se realizaron dos lavados con buffer acetato pH = 4,0, conteniendo NaCl 0.1 M y NaCl 0.5M y buffer Tris-HCl (Tris 0,1 M, NaCl 0,5 M) pH = 8,8.

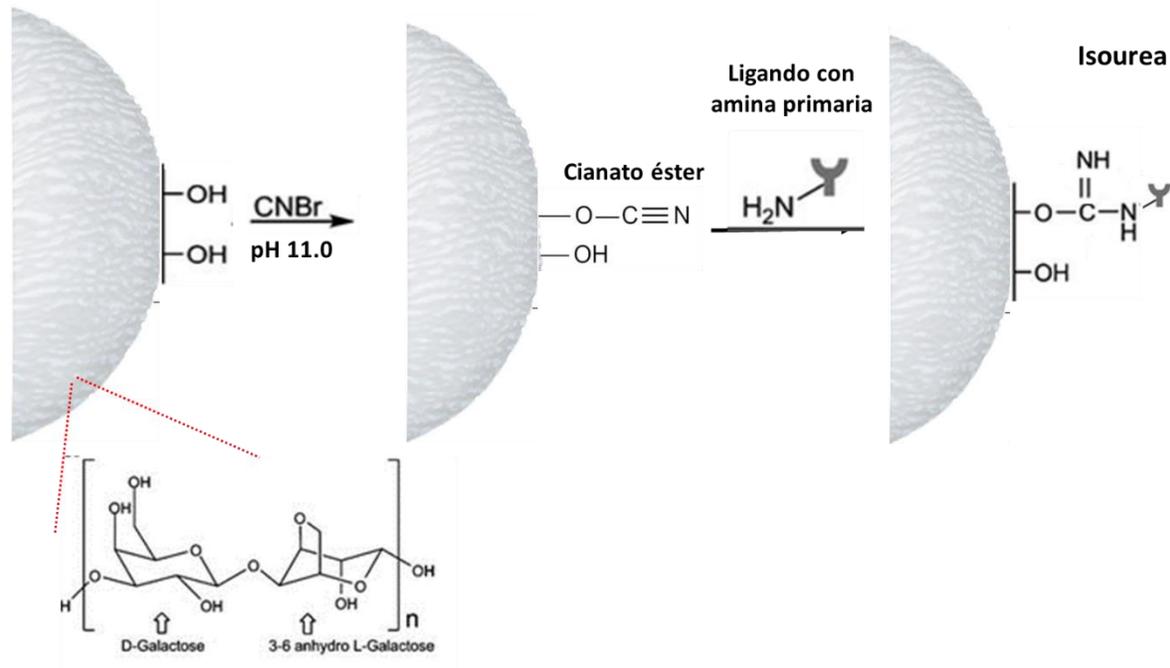


Figura 23. Reacción de acople molecular de un ligando (IgG anti LSB0-I) a una resina de sepharosa 4B pre activada con bromuro de cianógeno (CNBr)

5.1.2.4. Ensayo de ELISA

Para la determinación de la interacción LSB0-I e IgG (anti LSB0-I) se realizó un ensayo de ELISA haciendo dos replicas técnicas y dos replicas biológicas. Las placas se sensibilizaron con la FNR (fracción no retenida obtenida de cromatografía de intercambio iónico sobre DEAE donde está presente la lectina LSB0-I). Se preparó la proteína en una concentración de 0,1 µg/µL, a razón de 10 µg de proteína por pozo en buffer carbonatos 0.1 M pH 9.6 y se sembraron 100 µl a cada pozo, se dejó incubar a 37 °C por 3 horas y toda la noche a 4 °C. Al día siguiente se realizó el bloqueo de la placa incubándola con 200 µl por pozo de PBS-BSA (1.3 %) durante 1 hora a 37 °C.

A los antisueros (antisuero, FNR de intercambio iónico) y suero pre inmune se les realizó diluciones de 6, 12, 25, 50, 100 y 200 µg/mL en PBS-BSA (1.3 %) teniendo un volumen final de 100 µl por pozo, se dejó incubar 1 hora a 37 °C y luego toda la noche a 4 °C. El segundo anticuerpo (IgG generado en cabra contra IgGs de conejo y acoplado a peroxidasa (Sigma immunochemicals) fue preparado en una dilución 1/500 en PBS-BSA (1.3 %), se sembró 100 µl/pozo y se incubó 1 hora a 37 °C. El revelado se hizo con una solución de ABTS (ácido 2,2'-azino-bis(3-etilbenzotiazolina-6-sulfónico)) de 0.3 mg/ml en buffer citrato -fosfato (Ácido Cítrico 0.1 M, fosfato de sodio 0.2 M) pH 5.0 y en el momento de sembrar se agregaron 5 µl de H₂O₂ al 30 % por cada 10 ml de solución. La placa se mantuvo en la oscuridad por 15, 30 y 60 min y se leyó la absorbancia con un filtro de 405 nm. Entre cada paso se hicieron tres lavados agregando 100 µl/pozo de PBS-TWEEN (0.1 %).

5.1.3. Extracción y purificación de la lectina de *S. bogotensis* (LSBo)

5.1.3.1. Obtención del extracto de proteínas

50 gramos de harina de *S. bogotensis* se resuspendieron en PBS 1X tiourea 5 mM en una relación 1:10 (p/v), y se dejaron en agitación a 4 °C durante toda la noche. Al día siguiente se obtuvo el sobrenadante con las proteínas solubles por centrifugación a 7000 g durante 30 minutos a 4 °C. Al pellet se le hicieron dos extracciones más en las mismas condiciones para obtener en total tres (3) extractos proteicos, los cuales se unieron para formar un pool de extractos final.

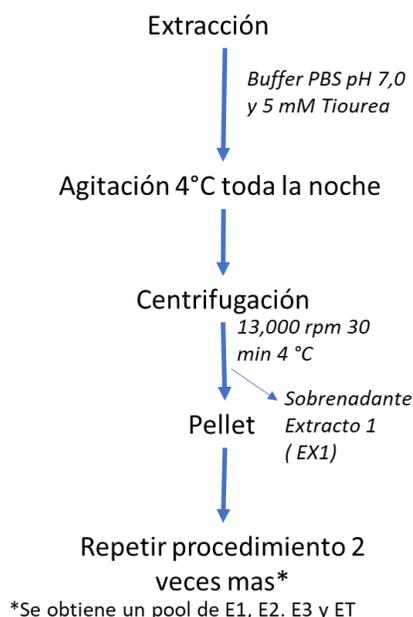


Figura 24. Esquema de extracción de proteínas totales con buffer PBS 1x – tiourea 5mM a partir de harina de semillas de *Salvia bogotensis*

5.1.3.2. Precipitación de proteínas

Se siguió la metodología descrita por Vega y Pérez, (2006) (4). 890 mL del pool de proteínas (5.1.31) fue sometido a precipitación fraccionada con etanol 96% frío entre 0-60 %. El proceso se llevó a cabo a 4 °C y se mantuvo con un goteo lento y constante (1 gota cada 10 segundos). Una vez alcanzado el 60 % de etanol se dejó el extracto durante una hora sin agitación y posteriormente se centrifugó a 10000 rpm a 4 °C durante 30 minutos. El precipitado se solubilizó en PBS 1x tiourea 5 mM y fue dializado contra agua destilada durante 8 horas realizando dos cambios, finalmente fue almacenado en alícuotas para realizar posteriores ensayos de purificación de la lectina (LSBo-I).



Figura 25. Precipitación fraccionada de proteínas con etanol 96% en una concentración de 0-60%

5.1.3.3. Cromatografía de intercambio iónico

Se empleó un soporte de DEAE Sephadex A-50® (Pharmacia) para llevar a cabo la purificación inicial. Este soporte es un intercambiador aniónico débil con un tamaño de partícula entre (40-120) µm y tiene acoplada una amina cuaternaria (116). La columna se empacó con 45 mL de soporte, el cual fue hinchado previamente en PBS 1X procurando no dejar burbujas ni fisuras a lo largo de la extensión de la columna de 50 mL. A continuación, se equilibró con 10 volúmenes de buffer PBS 1X, se midió la absorbancia a 220 y 280 nm para corroborar una la línea base cercana a Abs =0.00. Después se colocaron 20 mL del precipitado de proteínas

soluble, la fracción no retenida (FNR) se eluyó en PBS 1X. Para eluir la fracción retenida (FR) se adicionó al PBS 1x conteniendo NaCl 500 mM. La elución se siguió mediante absorbancia a 220 y 280 nm. El pool de las fracciones de máxima absorbancia se concentró por ultrafiltración (UF) (Amicon, PM 10), entonces se llevaron a cabo ensayos de actividad aglutinante, SDS-PAGE y cuantificación por BCA.

5.1.3.4. Tratamiento con Pectinex®

Debido a su elevada viscosidad, algunos extractos fueron tratado con Pectinex® para degradar carbohidratos (122). Se adicionaron 56 µL de Pectinex® Ultra SP-L (Novozymes corp.) por cada 3 mL de extracto, este proceso se llevó a cabo a un pH 4.5 con buffer acetato/ácido acético (200 mM), durante 12 horas a 37 °C. Cumplido este tiempo, se centrifugó a 15000 rpm durante 30 minutos a 4 °C. Al sobrenadante se le ajustó el pH a 7.0 con NaOH 0.1 M y se almacenó a 4 °C para posteriores ensayos.

5.1.3.5. Cromatografía de inmunoafinidad sobre Sepharosa 4B IgG (antiLSBo-I)

Se llevó a cabo la purificación de la fracción no retenida (FNR) sobre 2 mL de Sepharose 4B-IgG (antiLSBo-I). La elución se llevó con PBS 1X y un flujo de 500 µL/ 5-7 minutos. La fracción retenida (FR) se obtuvo a partir de buffers de diferentes pHs para romper la interacción anticuerpo lectina. Inicialmente con Glicina-HCl (50 mM, NaCl 150 mM) pH = 2,5 y luego con Tris-HCl (50 mM) pH = 8,13. Las tres fracciones se dializaron, liofilizaron y se hicieron ensayos de aglutinación, SDS PAGE y cuantificación.

5.1.3.6. Cromatografía de afinidad sobre Sepharose 4B - Concanavalina A (ConA)

Las fracciones retenidas obtenidas en 5.1.3.5 fueron equilibradas en una solución de NaCl 1 %, 5 mM Ca⁺⁺, 5 mM Mn⁺⁺. Luego fueron sembradas sobre una columna de 1 mL de Con A Sepharose™ 4B (GE Healthcare). La fracción no retenida se eluyó con la solución de equilibrio mientras que las fracciones retenidas se obtuvieron por elución con Glucosa 0.2 M en NaCl 1 % y buffer Gly-HCl (50 mM) pH 2,8. El seguimiento del perfil cromatográfico se realizó colectando volúmenes de aproximadamente 500 µL y midiendo absorbancia a 280 nm. Las fracciones que presentaron mayores absorbancias fueron reunidas en un pool para diálisis y posteriores ensayos de actividad y SDS PAGE.

5.1.4. Determinación de la estructura primaria con la proteína pura

5.1.4.1. Preparación del digerido trípico de la LSB_o-I

La proteína pura (fracción pH 8.3) se resuspendió en buffer MES 20 mM (Ácido 4-morfolinetanosulfónico sal hemisódica), pH 7 4/100 mM NaCl /glicerol 4% y se calentó a 100°C durante 10 minutos. Se corrió un gel en condiciones denaturantes con una concentración de entrecruzamiento de 15%. Se cortó la banda de la proteína con un bisturí y se dividió en trozos de aproximadamente 1 o 2 mm, luego se pasaron a tubos de 1,5 mL. Entonces se adicionaron 200µL de solución de lavado (50% metanol y 5% ácido acético) y se dejó toda la noche a temperatura ambiente.

Al siguiente día, se centrifugó y se removió la solución de lavado con micropipeta, luego se realizó nuevamente un lavado con 200 µl de la misma solución y se dejó por 3 horas a temperatura ambiente, se centrifugo nuevamente y descarto la solución. En seguida se adicionaron 200 µL de acetonitrilo para deshidratar el gel por 5 min, y se centrifugó al vacío durante 3 min. A continuación, se adicionan 30 µL de DTT 10mM por 30 min a temperatura ambiente, pasado este tiempo se descartó el DTT y se adicionaron 30 µL de Iodoacetamida por 30 min a temperatura ambiente. Se hizo de nuevo la deshidratación del gel por 5 min con 200 µL de acetonitrilo, se descartó y se rehidrato dejando incubar la muestra con 200 µL de bicarbonato de amonio 100mM. El proceso de hidratación deshidratación se repitió una vez, así como el secado de los trozos de gel al vacío por 3 min. Ulteriormente, se adicionaron 30 µL de solución de tripsina grado proteómica de 20 ng/uL sobre hielo por 10 min, se removió el exceso y se adicionaron 50 µL de bicarbonato de amonio 50 mM dando vortex y se dejó toda la noche a 37°C.

Los péptidos se recuperaron por adición de 30 µL de bicarbonato de amonio 50 mM e incubación de la mezcla por 10 min. La muestra se centrifugo a 1000 rpm por 30 segundos (s) y se recuperó el sobrenadante transfiriéndolo a un tubo de 0,5 mL. Seguidamente, se adicionaron 30 µL de buffer de extracción (50% acetonitrilo, 5% ácido fórmico) y se incubo nuevamente por 10 min dando vortex ocasionalmente, luego se centrifugó por 30 seg y se recuperó el sobrenadante. Para terminar, se repitió nuevamente el paso anterior incubando por 10 min con buffer de extracción. Finalmente se redujo el volumen por evaporación a 20µL con centrifugación al vacío a temperatura ambiente, sin dejar que la muestra se seque completamente (123).

5.1.4.2. Análisis de los péptidos por nano LC-MS/MS

Las muestras fueron resuspendidas en 5 µl de ácido fórmico al 0,1 %, luego se inyectaron en una columna nanoviper C18 (3 µm, 75 µm X 2 cm, Dionex) a un flujo

de 3 μl / min. Luego se pasaron por una columna EASY spray C18 RSLC (2 μm , 75 $\mu\text{m} \times 25 \text{ cm}$) con un flujo de 300 nl/min conectado a un sistema UltiMate 3000 RSLC (Dionex, Sunnyvale, CA) e interconectado con un Orbitrap Fusion™ Tribrid™ (Thermo-Fisher Scientific, San Jose, CA) espectrómetro de masas equipado con una fuente de nanoiones “EASY Spray” (Thermo-Fisher Scientific, San Jose, CA).

Para la separación de los péptidos, se estableció un gradiente cromatográfico con agua de grado MS (disolvente A) y ácido fórmico al 0,1 % en acetonitrilo al 90 % (disolvente B) durante 30 min de la siguiente manera: 10 min disolvente A, 7–20 % disolvente B en 25 min, 20 % de disolvente B durante 15 min, 20–25 % de disolvente B durante 15 min, 25–95 % de disolvente B durante 20 min y 8 min de disolvente A.

El espectrómetro de masas se operó en modo de iones positivos con un voltaje de nanopulverización establecido en 3,5 kV y una temperatura de la fuente de 280 °C. El calibrador externo incluía cafeína, Met-Arg-Phe-Ala (MRFA) y Ultramark 1621. El espectrómetro de masas se hizo funcionar en un modo dependiente de los datos para cambiar automáticamente entre MS y MS/MS. Los espectros de MS de barrido completo se adquirieron en el analizador Orbitrap, con un barrido de masas entre 350–1500 m/z a una resolución de 120 000 (FWHM), inyección máxima de tiempo a 50 ms, exclusión dinámica 1 a 90S y tolerancia de masa de 10 ppm. Se seleccionó un escaneo de inspección de velocidad máxima durante 3 s para la disociación inducida por colisión (CID) de Orbitrap basada en el árbol de decisión posterior o la fragmentación por disociación de colisión (HCD) de mayor energía (124); para la fragmentación CID, los estados de carga 3 se escanearon en un rango de 650 a 1200 m/z , el estado de carga 4 se escanearon en un rango de 300 a 900 m/z y el estado de carga 5 en un rango de escaneo de 300 a 950 m/z . Todos los datos se adquirieron con el software Xcalibur 4.0.27.10 (Thermo Fisher Scientific).

5.1.4.3. Búsqueda en bases de datos e identificación de péptidos

Los datos sin procesar se analizaron con Proteome Discoverer 2.1 (PD, Thermo Fisher Scientific Inc.) y las búsquedas posteriores se realizaron con el servidor Mascot (versión 2.4.1, Matrix Science, Boston, MA) y SEQUEST HT (125). Se incluyó como parámetros en la búsqueda, especificidad de proteasa trípica completa, se permitieron dos escisiones perdidas, las modificaciones estáticas cubrieron la carbamidometilación de cisteína (+57.021 Da). Además, las modificaciones dinámicas incluyeron oxidación de metionina (+15,995 Da) y desaminación en asparagina/glutamina (+0,984 Da). Para el método MS2, en el que la identificación se realizó a alta resolución en el Orbitrap, se aplicaron tolerancias de iones precursores y fragmentos de $\pm 10 \text{ ppm}$ y $\pm 0,2 \text{ Da}$. Los aciertos de péptidos

resultantes se filtraron para un máximo de 1 % de FDR usando el validador Target Decoy PSM. Se considero una puntuación MASCOT > 20 para proteínas identificadas con dos o más péptidos y una puntuación MASCOT > 34 para proteínas identificadas con un solo péptido.

5.1.4.4. Predicción de la estructura terciaria

Se realizó predicción de la estructura terciaria de la proteína empleando el método de homología con las herramientas bioinformáticas I-TASSER (<https://zhanggroup.org/I-TASSER/>) y Phyre² (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>).

I-TASSER permite el modelado comparativo utilizando una combinación de múltiples estructuras de referencia y una optimización iterativa de alineaciones alternativas. Realiza la comparación y selección de modelos de acuerdo con el mejor resultado (score). Para la evaluación del modelo se utilizarán los parámetros de I-TASSER, el cual arroja un algoritmo de puntuación (C-score) que va de -5 a 2; para la estimación de la calidad de los modelos de predicción, una mayor puntuación indica una mejor predicción en la estructura. Igualmente, el TM-score y el (RMSD) son parámetros estándar que se utilizaran para indicar la cercanía del modelo que se predice con la estructura nativa del molde. TM-score > 0,15 significará un modelo que presenta una topología correcta (126–128).

Phyre² arroja una tabla de resultados en la que se indica estimaciones de confianza, imágenes y enlaces a los modelos pronosticados tridimensionales e información derivada de la base de datos de Clasificación Estructural de Proteínas (SCOP) o del Banco de Datos de Proteínas (PDB), dependiendo de la fuente de la plantilla detectada (129).

5.1.4.5. Refinamiento modelos 3D

La validación de los modelos obtenidos por I-Tasser y Phyre², se realizó empleando las herramientas bioinformáticas: Ramachandran con Procheck vía PDBsum <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html> y MolProbity (servicio web de validación de estructuras para diagnosticar problemas en modelos 3D de proteínas), calculando la validación global y local para contactos de todos los átomos, geometría covalente y conformación tipo Ramachandran y rotámeros.

5.1.4.6. Acople molecular

Se hizo el acople molecular de la LSBo-I con 3 ligandos; se descargó el archivo PDB de la estructura 3D predicha de la lectina, y los 3 ligandos de interés fueron descargados de Pubchem en formato PDB. En Autodock-tools 1.5.6 fueron editados los archivos, se eliminaron aguas y se adicionaron hidrógenos polares. Sobre la lectina modelada se determinaron las dimensiones del gridbox para realizar el acople (x: 40, y: 22, z:46, centrado en x: 19,277, y: 23,761, z: 19,127). Para cada uno de los ligandos se definieron los ángulos de torsión y se guardaron los respectivos archivos en formato PDBQT; en el caso del receptor, este se definió como una estructura rígida. La corrida del acople se realizó usando el programa AutodockVina 1.1.2 (130), y para cada uno de los ligandos fue seleccionado el modelo con la mejor energía de interacción (valor más negativo). Para representar las interacciones en 2D y 3D entre los carbohidratos y la LSBo-I se usó Discovery studio.

5.2. Extracción de ARN total de las semillas de *S. bogotensis*

5.2.1. Extracción con Bromuro de hexadeciltrimetilamonio (CTAB).

Se agregaron en tubos falcón de 15 mL, ~700 mg aproximados de semillas maceradas en nitrógeno líquido y 3 mL de buffer de extracción (CTAB 2%, Tris HCl pH 8.0 100mM, NaCl 1,4M, EDTA 1,5M y PVP40 al 2%), se mezclaron por medio del vórtex durante 10 segundos, posteriormente se llevó a baño de maría por 15 minutos a 65°C. Luego se agregaron 3 mL de cloroformo-IAA (alcohol isoamílico) (24:1), se mezcló con vórtex por 10 segundos y se centrifugó a 8000 rpm por 15 minutos a 15 °C. Se transfirió la fase acuosa (~2 mL, evitando tomar elementos de la fase orgánica) y se agregaron 2 mL de cloroformo-IAA (24:1) se mezcló en vortex por 10 segundos y se centrifugó de nuevo a 8000 rpm por 15 minutos a 15°C. Se transfirió nuevamente la fase acuosa (~ 1 mL) y se agregó 1 mL de LiCl 4M, se dejó a incubación toda la noche a 4°C. Luego se centrifugo a 12000 rpm por 30 minutos a 4°C, se descartó el sobrenadante y resuspendió el pellet en 500µL de buffer TE-SDS (Tris HCl pH 8.0 -1M, EDTA pH 8.0- 0,5 M y SDS 1%). Luego se adicionaron 700µL de isopropanol y 200 µL de NaCl 5M, se mezcló invirtiendo el tubo varias veces y se llevó a -20°C por 30 min. Posterior al tiempo de incubación se centrifugo a 12000 rpm por 10 minutos a 4 °C. Se descarto el sobrenadante y se lavó el pellet con 500µL de etanol al 70% (en agua DEPC). Se centrifugó a 12000 rpm por 2 minutos y 4°C, se descartó el sobrenadante y se secó el pellet a 40°C por 30 min.

Finalmente, se suspendió el pellet en 50 μ L de agua libre de RNAsas. La metodología que se siguió esta descrita en (43).

5.2.2. Extracción de ARN total con Trizol

Se adicionó 1 ml de Trizol, solución monofásica de fenol con isotiocianato de guanidinio, del cual se emplearon dos marcas diferentes, (Zymo RESARCH y Ambion by life) para 0,2 g de material vegetal macerado con nitrógeno líquido, se homogenizó e incubó por 5 minutos a temperatura ambiente, a continuación, se adiciono 200 μ l de cloroformo por 1 ml de Trizol, se homogenizó con vortex por 15 segundos y se incubo de 2 a 3 minutos a temperatura ambiente. Después de la incubación se centrifugo a 12.000 g por 15 minutos a 4°C, se removió la fase acuosa a otro tubo y se adicionó 500 μ l de isopropanol 100% por 1 ml de Trizol, luego se incubo a temperatura ambiente por 10 minutos, se centrifugó a 12.000g por 10 minutos a 4°C y se removió el sobrenadante. El pellet se lavó con 1 ml de etanol 75% por 1 ml de Trizol, se centrifugó a 7.500g por 5 minutos a 4°C y se descartó el sobrenadante, este paso se repitió dos veces más. Finalmente se dejó secar el exceso de etanol sin dejar que el pellet seicara completamente y se resuspendió en agua estéril de 20 μ l a 50 μ l.

5.2.3. Síntesis de cDNA y amplificación de fragmentos de ADN

Como paso previo a la síntesis de primera hebra, el ARN obtenido fue tratado con ADNasa (RNase-Free DNase I®, de Epicentre) utilizando las siguientes cantidades: DNase I 1 μ L, buffer 10X para DNase I 1 μ L, y ARN 10 μ L. La mezcla de trabajo se llevó a 37°C durante 10 minutos y posteriormente se mantuvo a 10°C durante 5 minutos.

A partir del ARN obtenido en el paso anterior y utilizando oligo(dT) como iniciador, se llevó a cabo la síntesis de la primera cadena de ADN (cDNA) en el termociclador (Biorad), utilizando la técnica de transcripción reversa (RT-PCR) (131). Para este proceso se utilizó el sistema comercial RevertAid™ first strand cDNA Synthesis Kit de Thermo Scientific, el cual permite la síntesis de primera hebra a partir de ARN total. La solución de trabajo tuvo un volumen final de 20 μ L con ADNasa, 2 μ L Oligo dT, 4 μ L Buffer de reacción 5X, 1 μ L de inhibidor de RNAsas, 2 μ L de mezcla dNTPs 10 mM y 1 μ L de transcriptasa reversa. La mezcla se mantuvo durante 1 hora a 40°C, posteriormente se llevó a 70°C durante 5 minutos, y para finalizar se conservó -20°C hasta su posterior uso.

5.2.4. Amplificación de fragmentos de ADN

Este proceso se llevó a cabo mediante la técnica de Reacción en Cadena de la Polimerasa (PCR) (132), utilizando la enzima Taq DNA Polimerasa (Safe-Green™ 2x PCR Taq Mastermix Cat #472). La amplificación se realizó utilizando primers que fueron diseñados con secuencias de lectinas de leguminosas, las lectinas de *Glechoma hederácea*, *S. miltiorrhiza* y la secuencia parcial de la lectina *S. bogotensis* LSBo-I (80). Los primers fueron diseñados con BLAST (133), usando las secuencias disponibles de lamiáceas, contra otras secuencias reportadas para lectinas, con el fin de obtener una secuencia consenso que abarcará el gen. Se diseñaron los primers desde los dos extremos de cada gen reportado (desde el amino hasta el carboxilo terminal) y se utilizó el programa Primer3 disponible en NCBI, (<http://www.ncbi.nlm.nih.gov/>), para el diseño de primers forward y reverse con un tamaño de 20 a 23 nucleótidos (nt).

Con el fin de determinar el tamaño aproximado de los fragmentos obtenidos y conocer la eficiencia de la amplificación, el producto fue analizado en gel de agarosa (Alpha Aesar®) al 1.5%, teñido con SYBRr® Safe DNA Gel Stain (Thermo Scientific), sembrando 9 µL de la muestra con 1 µL de buffer de carga (BlueJuice® loading Invitrogen). El marcador de pares de bases empleado fue, O'RangeRuler 100bp + 500bp (Thermo Fisher).

6. Resultados y Análisis

Para entender la función de una proteína es necesario conocer su estructura tridimensional y para llegar a determinarla se debe conocer su secuencia de aminoácidos. En el caso de la LSBo-I, hasta hoy no se ha establecido la totalidad de su estructura primaria. Así mismo, no se conoce la estructura terciaria de ninguna de las lectinas de la familia de *Lamiaceas*, solamente se han propuesto dos por predicción bioinformática a partir de la secuencia del gen. Estas lectinas son interesantes porque reconocen específicamente los glicotopes T/Tn que se expresan como versiones cortas de los oligosacáridos en células tumorales, además se presentan diferencias muy sutiles en el reconocimiento entre las lectinas de la familias Fabácea, Lamiácea y otras que se han reportado son específicas por el determinante Tn (1–3,5,12,14,61,105).

Para establecer la estructura primaria de una proteína, es importante su calidad y cantidad, si se van a obtener resultados a partir de la proteína nativa. En el caso de la LSBo-I, se han realizado estudios enfocados en su purificación, pero los resultados no han sido satisfactorios; adicionalmente, los reportes de lectinas de lamiáceas son escasos comparados con los encontrados para las lectinas que reconocen los determinantes T/Tn de otras familias botánicas. Lo que corresponde a LSBo-I, no se ha determinado un esquema de purificación para obtener una proteína homogénea que permita realizar estudios de caracterización estructural, funcional y biológica. Adicionalmente, la secuencia N-terminal reportada no concuerda con la estructura primaria (5); sin embargo, se tiene una caracterización bioquímica completa. De conformidad con lo anterior, en este trabajo se estableció un nuevo método de purificación y se alinearon nuevos péptidos para aumentar el porcentaje de cobertura de la secuencia, así mismo se hicieron estudios enfocados en la obtención del cDNA como una nueva alternativa para obtener la lectina en forma recombinante y tener la secuencia completa.

6.1. Purificación de lectina de *S. bogotensis* (LSBo-I)

6.1.1. Purificación del anticuerpo (IgG) contra LSBo-I

En trabajos previos se obtuvo el antisuero para aislar los anticuerpos antiLSBo-I (IgGs) y emplearlo en ensayos de caracterización biológica y funcional de la lectina; para ello se inocularon entre 100-270 μg de LSBo-I que permitieron tener respuesta secundaria con incremento de las IgGs específicas (85). Los anticuerpos que se

producen en la primera exposición a un agente extraño son IgMs, sin embargo, una nueva exposición al antígeno produce anticuerpos de clase IgG, los cuales se caracterizan por tener mayor afinidad y ser secretados por células de memoria que se diferencian a células plasmáticas secretoras de anticuerpos (122). Los anticuerpos policlonales (IgGs) se utilizan en cromatografía de afinidad, y corresponden con ~ 85 % de las inmunoglobulinas (IgGs) totales.

La purificación se realizó por intercambio iónico para separar las inmunoglobulinas G (IgG) de las otras clases de IgGs, dado que están en mayor proporción en el antisuero (75 %). Se realizaron ensayos de precipitación con Polietilenglicol 8000 (PEG), para concentrar las proteínas que en su mayoría permanecieron en forma soluble con IgGs activas con título más alto (1:64 vs 1:256) (Tabla 4), respecto al antisuero. El precipitado soluble se aplicó sobre la columna de DEAE-Sephacel (Figura 26-A); en este paso de purificación no se observó incremento importante del título, toda vez que la precipitación pudo denaturar parte de las IgGs (Tabla 4).

Por SDS-PAGE se percibe cómo se fueron eliminando los contaminantes presentes en el antisuero para finalmente obtener las IgGs, observándose las cadenas livianas y pesadas del anticuerpo (Figura 26-B). Una molécula de anticuerpo (IgG), se caracteriza por ser una estructura simétrica compuesta de dos cadenas ligeras idénticas y dos cadenas pesadas idénticas que están unidas covalentemente por medio de enlaces disulfuro (Figura 27). En presencia de reductor se muestra una banda de ~52 kDa que corresponde con la cadena pesada (γ) y una banda difusa ~24-30 kDa característica de la cadena ligera (κ o λ) glicosilada (121,122), con punto isoeléctrico (pI) entre 5.1 y 7,1 (118,119). La albumina que se encuentra en mayor concentración en el antisuero de conejo, con un peso molecular de 67 kDa (120) (figura 26, carril 2); fue eliminada por cromatografía de intercambio iónico. Con este esquema de purificación se obtienen 8.44 mg de IgG totales a partir de 10 mL de antisuero, lo que representa el 3.3 % de la proteína total de partida (Tabla 4).

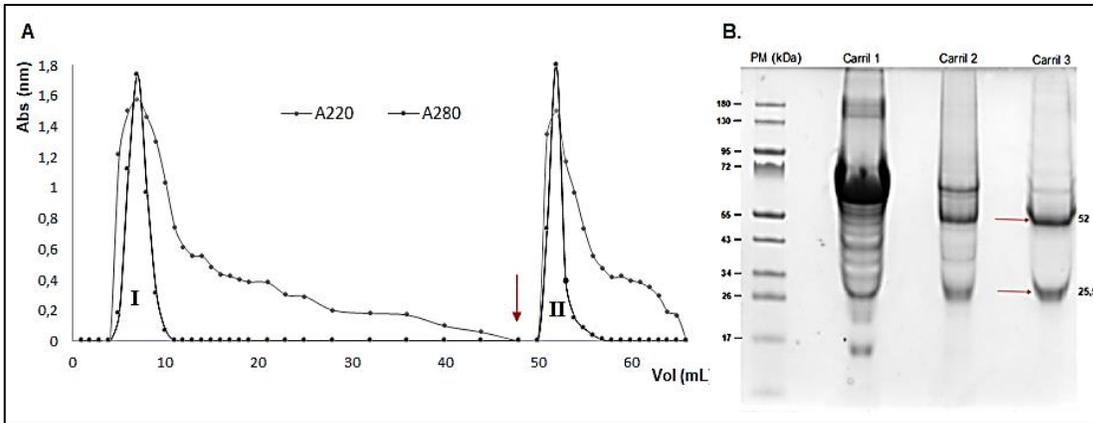


Figura 26. (A). Cromatografía de intercambio iónico sobre DEAE Sephacel empleando buffer de fosfatos 70mM pH 6,5. Elución: buffer fosfatos - NaCl 1,0 M (flecha roja). Se aplicaron 118 mg de proteína obtenida de la precipitación con PEG 8000. FNR (I), FR (II). (B) SDS PAGE Tris- tricina (12,5 %), fracciones preparadas en presencia de reductor (DTT) + calentamiento a ebullición (Δ) 98 °C. Carril 1: Antisuero, Carril 2: Precipitado PEG 10%, Carril 3: Fracción no retenida I (FNR) de la cromatografía DEAE Sephacel (II). Se colocaron 10 μ g de proteína total en cada carril. Tinción usando Comassie G 250 0,1 %.

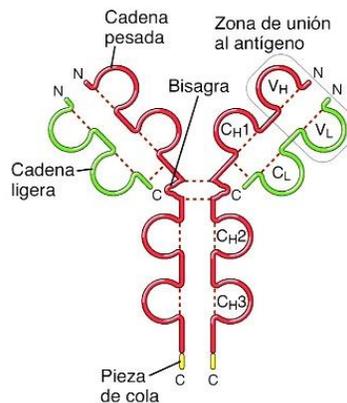


Figura 27. Representación de la inmunoglobulina G (IgG), en rojo, las cadenas pesadas y en verde las cadenas ligeras, también se pueden ver en líneas punteadas los puentes disulfuro intracatenarios e intercatenarios (134)

Tabla 4. Purificación de IgGs totales (antiLSBo-I)

Paso de purificación	Proteína (mg/mL)	Volumen (mL)	Proteína Total (mg)	Título ELISA ^a
Antisuero	25,53	10	255,3	1:64
Precipitado PEG 8000	9,207	3,5	32,22	1:256
Pool FNR	4,22	2	8,44	1:128

a. Ensayo de ELISA: Sé obtuvo a partir de una concentración inicial de proteínas de 200 μ g/mL, con un punto de corte de 0,120 de absorbancia a 405 nm teniendo en cuenta el blanco solo en presencia del anticuerpo secundario.

Las fracciones obtenidas en cada etapa de purificación se evaluaron por ensayo de ELISA en concentraciones de 6, 12, 25, 50, 100 y 200 $\mu\text{g/mL}$ (Figura 28) por duplicado, para corroborar el reconocimiento de la IgG (anti-LSBo-I) por la LSBo-I, a una concentración de 0,1 $\mu\text{g}/\mu\text{L}$, a razón de 10 μg de proteína por pozo. Se observó el reconocimiento por las IgGs presentes en cada una de las muestras evaluadas, donde la mayor interacción se obtuvo con 200 $\mu\text{g/mL}$; además se presentó tendencia lineal a medida que aumenta la cantidad de anticuerpo (IgG). La reacción antígeno – anticuerpo (Ag-Ab) se caracteriza por un equilibrio, donde a menor y mayor dilución del anticuerpo (Ab), la reacción se desplaza hacia la formación de una menor cantidad del complejo (Ag-Ab), observándose que todas las fracciones presentaron reconocimiento dependiente de la concentración, al ser comparadas con el suero pre inmune cuya mayor absorbancia fue de 0,240 nm con una concentración 200 $\mu\text{g/mL}$. El título fue mayor para la fracción precipitada con PEG 8000, donde probablemente la concentración de IgG fue mayor debido a la retención de la actividad de los anticuerpos funcionales (135). Otros métodos de purificación de IgGs (ratón) se realizan por cromatografía de afinidad con Proteína A de *Staphylococcus aureus*, (136); sin embargo, los resultados son similares a los obtenidos en intercambio iónico.

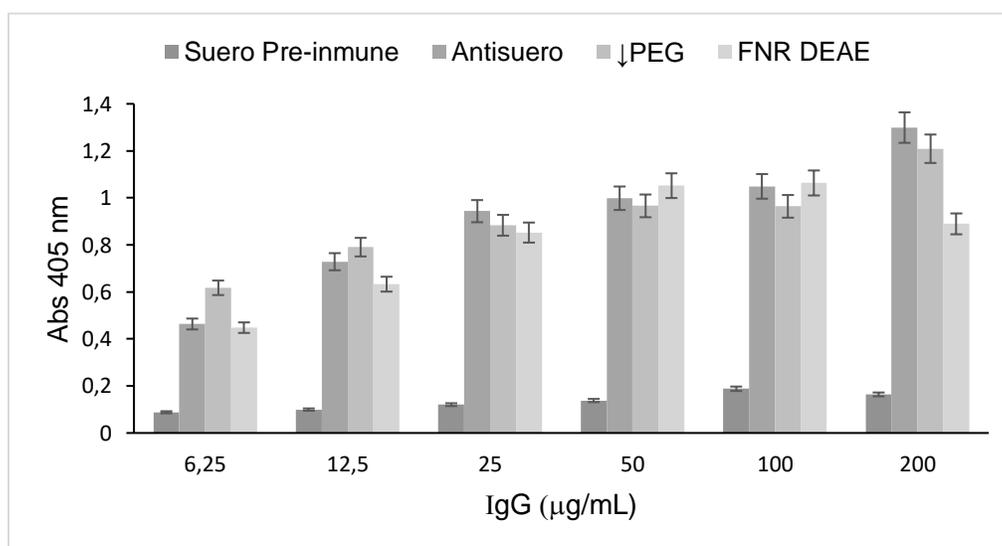


Figura 28. Ensayo de ELISA para la detección de la interacción LSBo-I vs IgGs. Se llevó a cabo con el Suero pre inmune, antisuero, ↓PEG 8000 10 %, lectina de *S. bogotensis* (LSBo-I), fracción no retenida (FNR) (IgG) obtenida por intercambio iónico.

El porcentaje de acople de la IgG a la resina de Sepharose 4B estuvo entre el 90-98 % (Tabla 5), pero la cromatografía de afinidad realizada con la columna del acople 1 no fue funcional, teniendo en cuenta que no hubo retención de la LSBo-I;

posiblemente la concentración de IgG (3,8 mg/mL) no permitió por impedimento estérico la interacción; o el número de moléculas efectivas de IgG luego del acople fue muy bajo (22) debido a la inmovilización en una orientación aleatoria, lo que genera una disminución en la eficiencia de unión (137). Estudios similares, permiten la purificación de transferasas en soportes con IgG de 5,3 mg/mL, con un porcentaje de acople del 99 % (127). Al disminuir la concentración de IgG (2,6 mg/mL), acople 2, se observó retención de la LSBo-I, por lo tanto, los ensayos conducentes a la obtención de la proteína pura se hicieron con este soporte de afinidad.

Tabla 5. Acople de IgGs a Sepharose 4B

Acople	Resina (g)	Resina (mL)	IgGs (mg)		Resina (mg/mL)	Porcentaje de acople (%)
			Iniciales	no acopladas		
1	0,54 g	2,0	8,44	0,84	3,8	90
2	0,60 g	2,5	6,67	0,07	2,6	98

Los ligandos utilizados normalmente para la separación por cromatografía de afinidad pueden incluir compuestos orgánicos, metales inorgánicos que forman complejos de coordinación con los aminoácidos de las proteínas, proteínas que pueden interactuar con otras proteínas por reconocimiento biológico y como en este caso; anticuerpos, que pueden diseñarse para unir un epítoto específico de una proteína (138). No obstante, se deben establecer las condiciones que permita que la cromatografía de afinidad sea satisfactoria; desde la selección de soporte, inmovilización de ligandos, hasta las condiciones de elución. Algunos ejemplos, incluyen la separación de enzimas como la neprilisina con anticuerpos monoclonales (139).

6.1.2. Extracción y purificación de la lectina (LSBo-I).

La extracción de proteínas se hizo con 50,0 g de harina de *S. bogotensis* (SBo) en PBS 1X; teniendo en cuenta que los extractos se pigmentan durante el proceso se agregó tiourea 5 mM. De acuerdo con estudios previos la tiourea inhibe la acción de las polifenoloxidasas (122,140,141), de modo que se puede evitar la oxidación de compuestos fenólicos que pueden interferir en posteriores procesos de purificación o que disminuyan la actividad de la lectina. Los extractos obtenidos a partir de las tres extracciones se unieron en un pool y se precipitaron con etanol entre el 0-60 % como había sido reportado previamente (4). Las proteínas se solubilizaron en PBS 1X (Figura 29), y por SDS PAGE (Figura 30), se observó un gran número de bandas de alto, medio y bajo peso molecular para las tres extracciones y el pool de extractos. En cambio, para el precipitado de proteínas con etanol del 96 %, las

bandas fueron difusas y en menor cantidad, debido a la coprecipitación de polisacáridos que pueden interferir en la resolución de las bandas de proteína.



Figura 29. Extracto de proteínas precipitado con etanol 0-60 % y solubilizado en PBS 1X, en membrana de diálisis (MW cut off 10 kDa). Se observa que el extracto es denso y pigmentado debido a la presencia de polisacáridos.

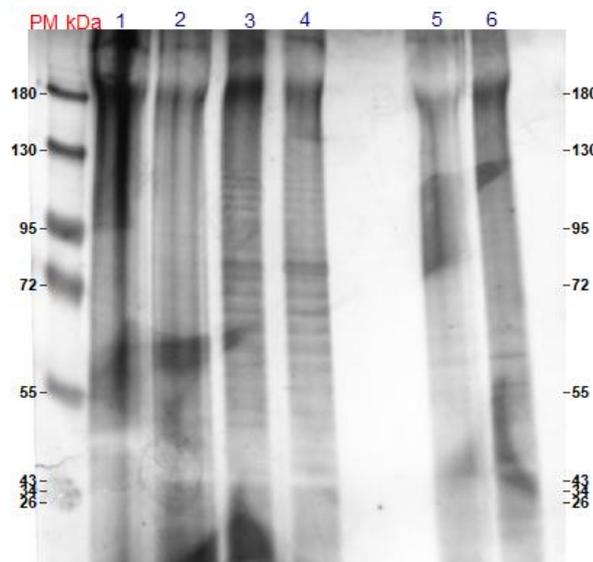


Figura 30. SDS PAGE Tris Glicina 12,5 %. Extracto proteico de semillas de *S. bogotensis* (Sbo), en condiciones reductoras (DTT) + calentamiento a 98 °C (Δ). PM (patrones de peso molecular). Carril 1, Extracto I. Carril 2 Extracto II, Carril 3. Extracto III. Carril 4. Pool de extractos, Carril 5. Precipitado de proteínas con etanol 0-60 %. Carril 6. Extracto precipitado con etanol 0-60 % dializado. Se sembraron 10 μ g de proteína total en cada carril, tinción de plata.

El precipitado de proteínas se sometió a cromatografía de intercambio iónico sobre Sephadex DEAE A-50; se realizaron en total seis (6) corridas cromatográficas y en cada una se agregaron 40 mL del precipitado soluble que contenían un total de ciento dieciocho (118) mg de proteína (Figura 31-A). Las fracciones no retenidas (FNR) se agruparon en un pool (269 mL), se concentraron por ultrafiltración con membrana de 10 kDa, para llevar a cabo los ensayos de afinidad; sin embargo, la fracción conservó parte de los pigmentos, la mayoría quedó retenida de forma irreversible en el soporte cromatográfico.

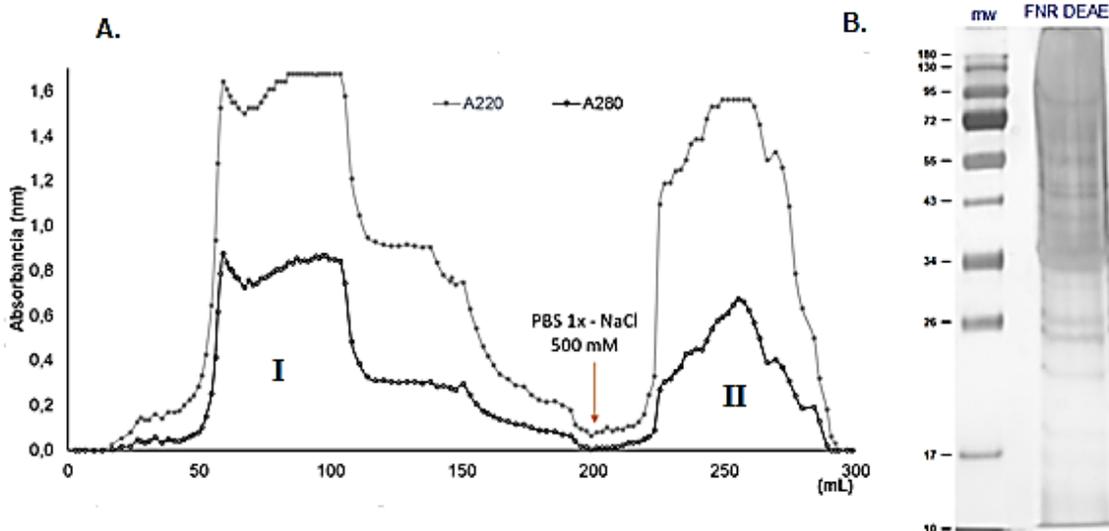


Figura 31. (A) Cromatografía DEAE Sephadex A-50®, 118 mg de proteína total. (I) Fracción no retenida FNR, (II) Fracción retenida FR, elución con alta fuerza iónica (PBS 1X - NaCl 500 mM). (B) Fracción no retenida DEAE - SDS PAGE Tris – Glicina 12,5 %, +DTT, +Δ 98 °C. (mw): Patrones de peso molecular (kDa). FNR DEAE: 25 µg de proteína total. Tinción de plata.

Los eritrocitos de diferente origen se emplean rutinariamente para determinar la actividad de las lectinas en un ambiente de membrana celular, y pueden ser tratados con enzimas para hidrolizar oligosacáridos, por ejemplo, la glicoforina A (GPYA), es una proteína transmembranal altamente glicosilada (60 %), que contiene diez y seis (16) cadenas de O-glicanos (NeuAc α 2,3Gal β 1,3(NeuAc α 2,6)GalNAc), una cadena de N-glicano triantenario (NeuAc α 2,6 y núcleo fucosa α -Fuc1,6), además de los determinantes de los de grupos-histosanguíneos, lo que hace posible realizar ensayos de actividad aglutinante (142). Igualmente, se exponen estructuras de carbohidratos por degradación enzimática que ayudan a la interacción de estas proteínas con el glicano (118). Conforme con lo anterior, la fracción no retenida (FNR) aglutinó eritrocitos T (+3) que se obtienen por degradación de eritrocitos A con neuraminidasa (Sialidasa) para exponer residuos de galactosa (Gal). Así mismo, aglutinó eritrocitos Tn (+3), como se había reportado en trabajos previos (4,5). En este caso, los eritrocitos T, son tratados con β -galactosidasa para exponer residuos de α GalNAc; los eritrocitos tratados fueron aglutinados (+1) con 1,5 µg de proteína total.

Las lectinas que reconocen GalNAc a menudo muestran un amplio rango de especificidad de acuerdo con la configuración de los enlaces (α o β), ubicación en el oligosacárido (interna o externa) y la participación de los azúcares vecinos (11,143). Además, el modo de presentación de GalNAc, monovalente, polivalente multiantenaria o en clúster, puede influenciar la interacción de la lectina (144); en

este caso, los eritrocitos exponen varias unidades estructurales de los glicotopes T (Gal β 1,3GalNAc α 1-Ser/Thr) y Tn (GalNAc α 1-Ser/Thr), presentes en O-glicanos de la GPYA en la superficie celular. Por otro lado, la LSBo-I, mostró especificidad por el glicotipe Tn en forma polivalente con la mucina submaxilar ovina y bovina (MSO, MSB) (1,145). Teniendo en cuenta que los resultados de aglutinación permitieron detectar con pequeñas cantidades (1.5 μ g) la actividad de lectina, se seleccionaron estos ensayos para continuar el seguimiento de la purificación.

Por SDS PAGE en la FNR se observó bandas intensas definidas para proteínas de peso molecular medio (>26 kDa), y con menor intensidad de bajo peso molecular. Al comparar estos resultados con los obtenidos previamente (4,5), hay una mejor resolución de bandas de proteínas, lo que muestra una mejor calidad del extracto de proteínas (Figura 31-B).

Luego del proceso de concentración de la fracción de proteínas (FNR) por ultrafiltración, aumentó la viscosidad dificultando los ensayos de purificación; por tal razón, se realizó digestión enzimática con una mezcla comercial de poligalacturonasas (Pectinex®). Estas enzimas son específicas para hidrolizar pectinas y mucílagos solubles presentes en extractos de semillas; en este caso, degradó los polisacáridos que pudieron precipitar con etanol al 96 % y que estaban presentes luego de la cromatografía sobre DEAE-Sphadex A-50. El proceso de digestión mejoró la viscosidad del extracto y se conservó la actividad de la lectina dado que aglutinó eritrocitos T al 2.5 % con 1.3 μ g de proteína total. La adición de estas enzimas a los extractos crudos obtenidos de semillas de diferentes especies de Lamiáceas mejoró la interacción de las lectinas por glicoproteínas unidas a superficies de poliestireno cubiertas con aMSO (4,5,87,106,146).

En la FNR se obtuvo 64,5 mg de proteína total con 15 μ g/mL de título de aglutinación; por consiguiente, se llevó a inmunoafinidad sobre el soporte Sepharosa 4B-IgG (AntiLSBo). Se obtuvo dos fracciones retenidas y eluidas usando buffers de Tris HCL (50 mM) pH 8,3, y Gly-HCL (50 mM) pH 2,8 (FR 2,8) (Figura 32). La mayor cantidad de proteínas, ~93 %, fue obtenida en la FNR con aglutinación débil (\pm) causada por 20.4 μ g de proteína total (Tabla 6). Paralelamente, las fracciones purificadas presentaron actividad aglutinante (+1) con 0,16 y 0,06 μ g de proteína total para las fracciones obtenidas a pH 8,3 y 2,8, conjuntamente con grados de purificación más altos (4716 veces), respecto a lo reportado previamente (2000 veces) (4); en este último caso, se aisló la LSBo-I por cromatografía de afinidad sobre Sepharosa 4B-aMSB (asialo mucina submaxilar bovina) y una elución a pH 11.4. La elución a pHs extremos básicos causa denaturación y reducción de puentes disulfuro intracatenarios, que pueden reoxidarse de forma

diferente en la proteína nativa, y formar agregados de alto peso molecular que no se pueden disociar fácilmente, así por SDS PAGE, las bandas de proteína son heterogéneas, delgadas y no resueltas (4,5). Lo anterior se vio reflejado al comparar los perfiles electroforéticos, donde se resalta la calidad de la lectina purificada (15,75) (Figura 33) mostrando bandas gruesas y definidas.

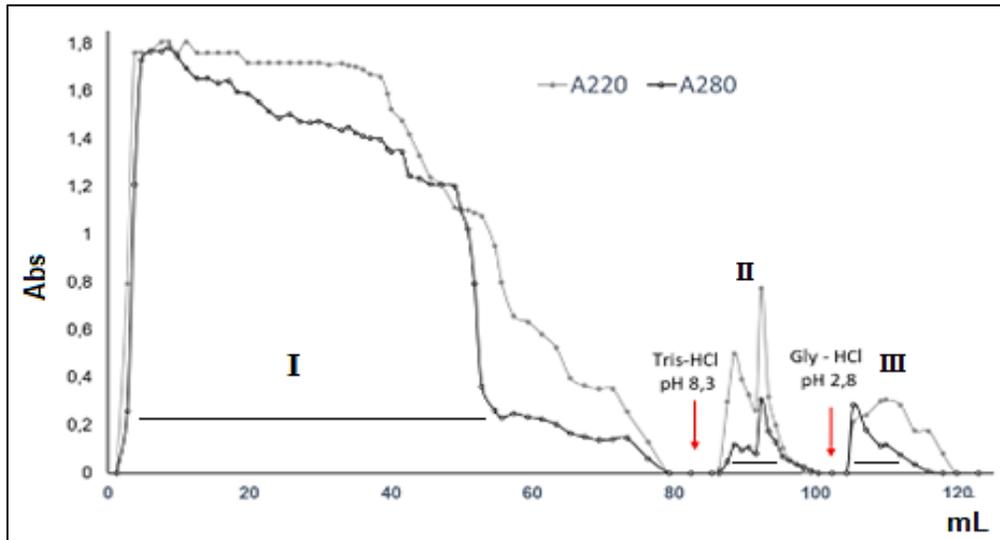


Figura 32. Cromatografía de Inmunoafinidad usando un soporte de Sepharosa-4B inmunoglobulina G (IgG:Anti- LSBo-I); Se sembraron 17 mg de la FNR (DEAE) tratada con Pectinex. Buffer de equilibrio: PBS 1x, Buffer de elución: Tris -HCl (50 mM) pH 8,3 y Gly (Glicina) -HCl (50 mM) pH 2,8. I (FNR), II (FR 8,3) y III (FR 2,8)

Con este método de purificación la elución de la lectina se llevó a cabo en condiciones de elución más suaves, por SDS PAGE. Para la FR 8,3 se observó una banda intensa alrededor de los 36 kDa (80 %) (Figura 33-A, carril 1), que correspondería con el monómero de la LSBo-I, reportada previamente con un peso molecular de 35.7 kDa (15), también hay bandas de menor intensidad (10 %) en 29 kDa y 45 kDa, posiblemente se presentan interacciones inespecíficas o reactividad cruzada mediada por el anticuerpo policlonal (119), aunque para la FR 2,8 se encontró una única banda alrededor de 36 kDa (100 %) (Figura 33-A, carril 2) que corresponde con el monómero de la lectina (LSBo-I). Acorde con la técnica de afinidad se logró aislar la LSBo-I a partir de un extracto crudo, con las condiciones establecidas que incluyeron la selección del ligando (IgG), interacción (LSBo-I:IgG) y condiciones de elución de la fracción retenida (LSBo-I).

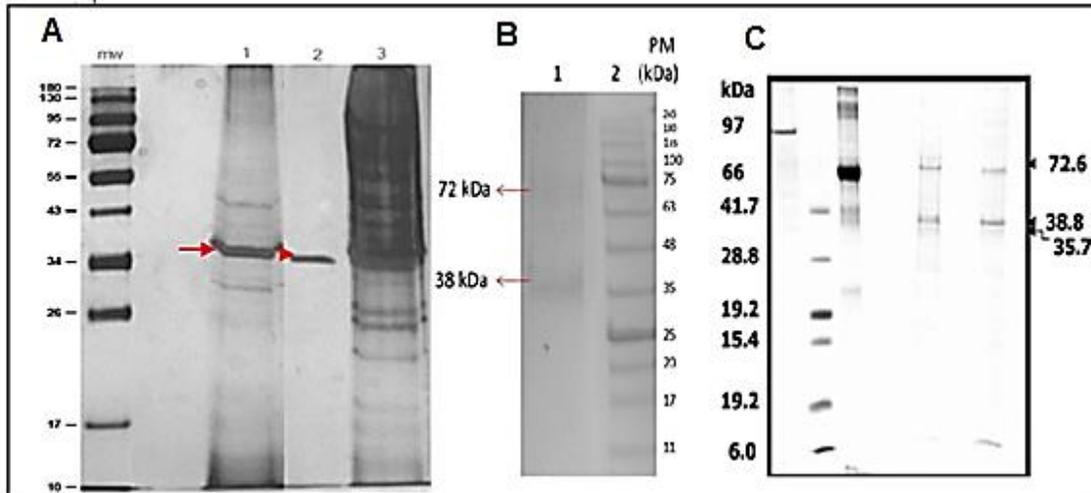


Figura 33. (A). SDS PAGE 12,5 %, +98 °C para la LsBo-I obtenida por inmunoafinidad sobre Sepharosa 4B IgG. mw: patrones de peso molecular, Carril 1. Fracción retenida (FR), pH 8,3. Carril 2. Fracción retenida (FR), pH 2,8. Carril 3. Fracción no retenida (FNR). Se colocaron 30 µg de proteína total en cada carril. (B)(C). SDS PAGE 12,5 %, +98 °C para la LsBo-I obtenida por Cromatografía de afinidad sobre Sepharosa 4B-aMSB. Imágenes tomadas de (4,80)

Es importante mencionar que en las semillas de SBo (*S. bogotensis*) hay una lectina en bajo porcentaje con especificidad por residuos de manosa (LsBo-II) que aglutinó eritrocitos A, con un monómero de ~29 kDa (75). Con el objetivo de separar las dos lectinas, se incluyó la purificación sobre DEAE sephadex A-50, aprovechando la diferencia en los puntos isoeléctricos (pI) (8.0 vs 6.5) (80). Sin embargo, no toda la lectina específica de manosa (LsBo-II) (pI=6.5) se retuvo, de modo que se incluyó un proceso adicional de purificación para la FR 8,3, empleando soporte de Sepharosa 4B ConA (Concanavalina A) (Figura 34) (Anexo B).

La elución de la fracción retenida se realizó con una solución de NaCl 1 % conteniendo glucosa (Glc) 0.2 M, teniendo en cuenta que este carbohidrato es reconocido por Con A, y posteriormente con un buffer Gly-HCl (50 mM) pH 2,8, para eluir la proteína retenida. La obtención de una fracción a pH 2,8 se debe a un comportamiento característico de las lectinas de la tribu Diocleae, que presentan un equilibrio dímero – tetramero dependiente del pH, donde cada uno de los bucles que conforman la cavidad central de la estructura cuaternaria tienen una orientación que permite que algunos aminoácidos cargados interactúen para estabilizar el tetramero, por lo que al disminuir drásticamente el pH, se presenta la disociación de subunidades de Con A, afectando su afinidad por las moléculas que reconoce (147).

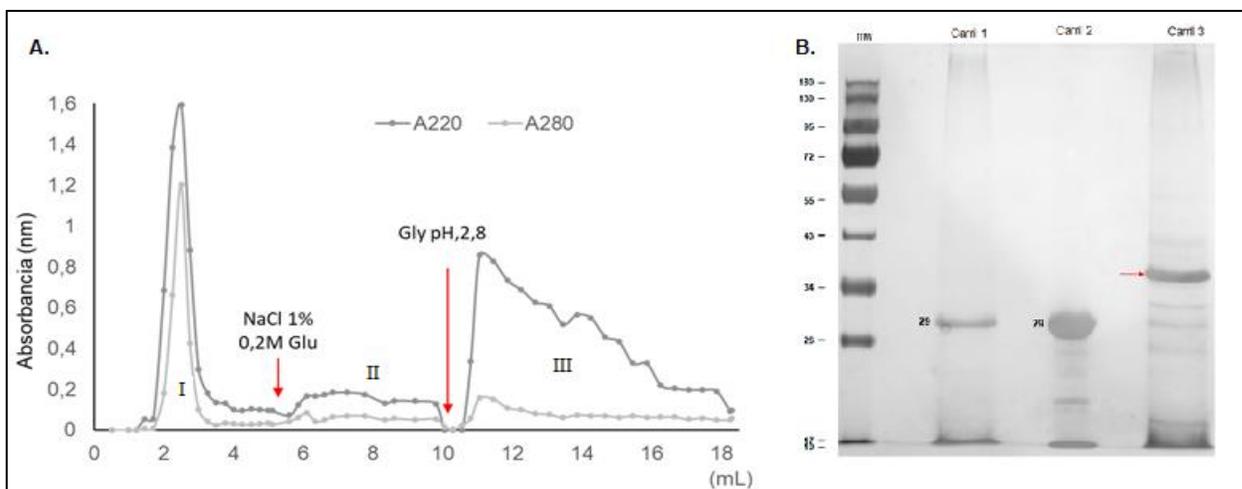


Figura 34. (A) Cromatografía de afinidad sobre Sepharose 4B- ConA, Muestra: FR pH 8,3 (203 μ g de proteína total). Buffer de equilibrio: NaCl 1 %, CaCl₂ 5mM, MnCl₂ 5mM. Elución: NaCl 1 % - Glucosa 0,2 M, buffer Gly- HCl (50 mM) pH 2,8. (I) FNR, (II) FR Glc, (III), FR 2,8. (B) Fracciones de cromatografía de afinidad - Sepharosa 4B ConA, SDS PAGE 12,5 %: Δ (98 °C) +DTT. mw: Patrones de peso molecular, carril 1: FR eluida con NaCl 1 % - Glucosa 0,2 M (5 μ g de proteína), Carril 2: FR eluida con Gly HCl pH 2,8 (10 μ g de proteína), Carril 3: FNR (5 μ g de proteína). Tinción de plata

En el perfil de elución (Figura 34-A) se observó una fracción mayoritaria simétrica no retenida (FNR) respecto a las dos fracciones retenidas (FR), lo que demuestra la separación de los componentes minoritarios. Sin embargo, por SDS PAGE (Figura 34-B, carril 3), se observó la banda de 36 kDa (97 %) correspondiente al monómero de LSB_o-I, acompañada de bandas contaminantes que están en baja proporción y que probablemente se derivan del manejo de la muestra o interacciones proteína – proteína que se pudieron dar y coeluyeron juntas.

Considerando que la banda de lectina es de buena calidad, se continuaron los ensayos de secuenciación *in situ*, por extracción a partir de un gel de electroforesis como se indica más adelante. La misma fracción aglutinó eritrocitos Tn (+3) (Figura 35-D); en efecto para las fracciones retenidas (FR), se observó la banda de ~29 kDa que corresponde al monómero de la LSB_o-II que reconoce manosa (5).

Algunas lectinas de Fabáceas, y específicamente de la tribu Diocleinae con afinidad por Man/Glc tienen un peso molecular aparente de 30 kDa en condiciones denaturantes (148,149). Adicionalmente, algunas tienen un bajo porcentaje de glicosilación y un oligosacárido que es reconocido por la lectina de Concanavalina A (ConA) (150–152), de modo que por afinidad se pueden aislar N-glicoproteínas típicas de proteínas vegetales ricas en manosa [(Man α 1-6[(Man α 1-3)(Xyl β 1-2)]Man β 1-4-GlcNAc β 1-4(Fuc α 1-3)GlcNAc] y [(Man α 1-3/6)(Xyl β 1-2)]Man β 1-4-GlcNAc β 1-4(Fuc α 1-3)GlcNAc (14,80). Estos resultados indican que ambas lectinas

se diferencian en sus propiedades moleculares y bioquímicas, probablemente LSBo-I no es glicosilada o el glicano tiene una estructura diferente que no es reconocida por ConA. Por su parte las glicoproteínas de tipo O han sido poco estudiadas (153) en plantas; en consecuencia, no se puede asumir que no se presenten en glicoproteínas aisladas de semillas.

En la figura 35 se observan aglutinados de eritrocitos Tn de diferente tamaño, promovidos por LSBo-I en las diferentes etapas de purificación; de igual modo la cantidad de proteína necesaria para observar la actividad disminuyó en cada etapa, por ejemplo, con 4,5 µg de proteína total de la FNR (DEAE) se encontró actividad (+2), mientras que para las fracciones retenidas de cromatografía de afinidad, FR pH 2,8 y FR 8,3, con 0,18 y 0,64 µg de proteína total la actividad fue caracterizada por grupos grandes de eritrocitos (+4). En estudios previos se reportaron aglutinaciones (+3) con 2,1 µg de proteína total (80), y una concentración mínima de 0.17 µg/mL de LSBo-I para generar aglutinación (+1) (4).

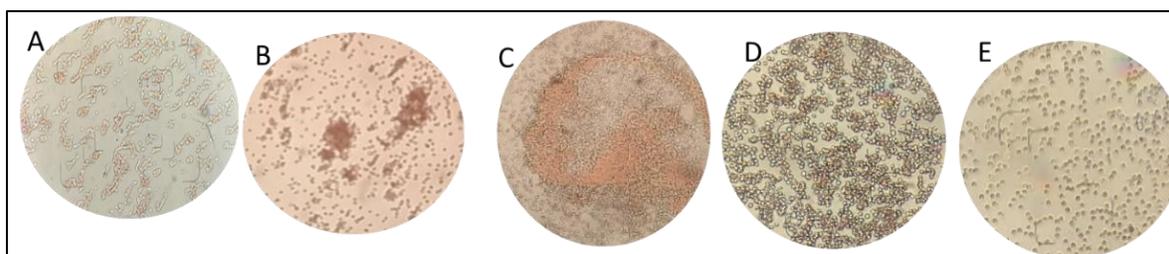


Figura 35. Aglutinación de eritrocitos Tn al 2.5 % con la LSBo-I^b (A)^a FNR DEAE (4,5 µg de proteína) (+2), (B) FR pH 2,8 (0,18 µg de proteína) (+4), (C) FR pH 8,3 (0,64 µg de proteína) (+4), (D) FNR -Con A, (0,56 µg de proteína) (+3) (E) Blanco (eritrocitos sin lectina).

- Las fracciones se obtuvieron a partir de la FNR DEAE tratada con Pectinex® Ultra SP-L.
- Escala cualitativa de aglutinación de +1 a +4, siendo (+1) una aglutinación débil, caracterizada por racimos de 3 a 4 células. (+4) es la máxima aglutinación caracterizada por racimos grandes de células.

En la tabla 6 se presentan los diferentes pasos de purificación realizados para obtener LSBo-I. A partir de 50,0 g de harina, se extrae 75 % de proteínas que precipitan con etanol al 60 %, de las cuales el 0.75 % (14.9 mg) corresponde a LSBo-I. El porcentaje de proteínas solubles que se extrajo de la harina fue del 4 %, donde, el 0.03 % corresponde a LSBo-I, la cual se obtuvo con alto grado de purificación (4716 veces) y actividad específica (0,06 µg). Los resultados son similares a los obtenidos con el esquema de purificación reportado previamente (4), puesto que partiendo de 100 g de semillas maduras se obtiene 18.9 mg de LSBo-I; cuatro veces mayor que lo obtenido para la lectina de *S. sclarea* (5 mg/100 g semillas) (13); con la lectina de *M. laevis* (MLL), a partir de 10 g de semillas, el grado de purificación (5-10 veces) fue bajo, al igual que la actividad aglutinante (139). Es

muy importante resaltar que los métodos de purificación previos generaron la desnaturalización de la proteína y por lo tanto pérdida de su actividad (4,80), por lo que el grado de purificación alcanzado en este trabajo es considerable y dada la remoción de polisacáridos, la eliminación de los pigmentos, y el método de inmunoafinidad empleado, permite llevar a cabo una elución menos drástica durante el proceso de purificación.

Tabla 6. Etapas de purificación de la lectina de *S. bogotensis* (LSBo-I)

Etapa	Fracción	Proteína (mg/mL) ^g	Vol (mL)	(mg) Totales	^a Actividad Específica	Grado de Purificación (Fold)
Extracción de proteínas ^f	Extracto I	3,60	250	900	---	----
	Extracto II	2,70	240	648	----	----
	Extracto III	1,80	405	729	---	----
	Pool de extractos	2,35	890	2.091,5	283	1
Precipitación con etanol al 96 %	0-60 %	3,95	400	1.580	123	2,3
Cromatografía sobre DEAE	Pool FNR DEAE ^b	0,24	269	64,50	1.5	188
	Pool FNR DEAE PEC ^{c,d}	0,88	60	53,04	1.3	217
Cromatografía de Afinidad (Anti- LSBo-I)	FR pH 8,3	0,83	0,50	0,415	0,16	1768
	FR pH 2,5	0,06	0,50	0,030	0,06	4716
	FNR Afinidad	0,83	60,0	49,62	20.75 ± ^e	----

- La actividad específica de aglutinación se define como la cantidad mínima de proteína (μg) total que se requiere para tener aglutinación (+1). Se utilizó una suspensión de eritrocitos Tn al 2.5 %.
- A partir de 118 mg de proteína aplicados sobre DEAE-Sephadex
- A partir de 64,5 mg de proteína tratados con pectinex y aplicados sobre Sepharosa 4B-IgG
- Método de concentración por ultrafiltración con gas Nitrógeno
- ±: Aglutinación débil.
- Extracción a partir de 50,0 g de Harina de *S. bogotensis* (SBo)
- Cuantificación por el método del ácido bicinconinico (BCA)

6.1.3. Análisis de péptidos trípticos obtenidos de la LSBo-I por nano LC-MS/MS para determinar la secuencia de aminoácidos

La banda del monómero de la LSBo-I, fue extraída del gel (Figura 36) y tratada con tripsina. Para evaluar la eficiencia de la hidrólisis, los digeridos fueron sometidos inmediatamente a cromatografía líquida y posteriormente a espectrometría de masas. Los péptidos trípticos fueron analizados y después de eliminar redundancias, se seleccionaron los más abundantes o principales. Se obtuvo (12) doce secuencias de péptidos, que fueron alineados con proteínas reportadas en Uniprot (<https://www.uniprot.org/blast/>), tipo leguminosa (LegB). Igualmente se realizó BLASTp para contrastar el alineamiento (Anexo C) y verificar la identidad de

los péptidos contra lectinas, encontrándose entre el 70 – 100 % identidad con proteínas que tienen en su estructura dominios de lectina tipo Leguminosa (LegB) y porcentaje de cobertura entre 2 y 9 % (Tabla 7). Solamente los péptidos 1 y 2 encajan en el dominio tipo LegB, los demás en el dominio kinasa u otra parte de la secuencia que no pertenece al dominio.

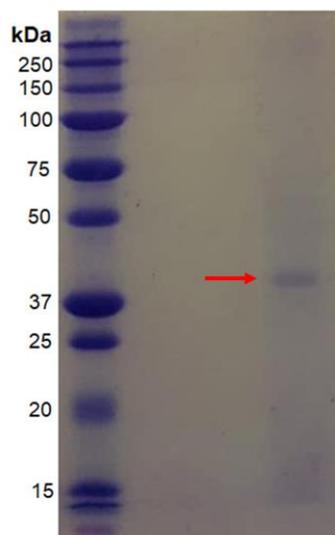


Figura 36. SDS-PAGE 15 % LSB0-I. Se observa la banda de proteína que fue cortada y sometida a digestión con tripsina.

Tabla 7. Péptidos obtenidos por digestión trípica y secuenciación (MS/MS)

N°	Péptido	% de cobertura	Proteína	Pfam	Características moleculares	Identificación UNIPROT
1	TQGWDIYEA GSRNLEVR 184-200	5	Proteína de <i>Malus baccata</i> ,	Lectina LegB 28-267	351 aminoácidos 37,803 Da. Proteína inferida por homología ^a	A0A540KTT4
2	IGFSASTGD WVEVHNILS WSFNSSLEIN NGKNNMGLII GLVVEPSK 240- 285	9	Proteína Quinasa ser/Thr no específica de <i>Acer yangbiense</i>	Lectina LegB 28-275	536 aa 59,653 Da Proteína predicha ^b	A0A5C7HSP9
3	MEISRLDQDN QQLMAIGSGI R 1-21	8	Proteína de <i>Camellia sinensis</i> var. <i>sinensis</i>	Lectina LegB 195-279	280 aa 30,649 Da Proteína predicha ^b	A0A4S4D761

4	GNGSSWRLS ATLAEPFTDD R 714-733	2	Dominio de lectina de leguminosa <i>Paenibacillus thiaminolyticus</i>	Lectina LegB 84-325	807 aa 89,034 Da Proteína predicha ^b	A0A378ZGP7
5	QVLADVAEG LNYLHHGWD QVVVHRDIKS SNILLDSDMR 406-503	6	Proteína Quinasa de <i>Populus trichocarpa</i>	LectinaLeg B 19-273 p-kinasa 347-617	655 aa 71,933 Da Proteína inferida por homología ^a	B9IEY2
6	LKLNNGVTV GSFVVLVLIAL FSLFLIKR 192-219	5	Protein kinase domain- containing protein <i>Gossypium barbadense</i>	Lectina LegB 25-205 p-Kinasa 254-522	589 aa 66,100 Da Proteína inferida por homología ^a	A0A5J5R3X5
7	LSLGEIKLAT MGFNQNRVV GEGGSATVY K 358-386	4	Dominio de proteín quinasa de <i>Cucumis sativus (Cucumber)</i>	Lectina LegB 19-268 p-kinasa 371-642	695 aa 76,864 Da Proteína por homología ^a	A0A0A0KD87
8	QFVAEVVMT GNIQHRNLVP LLGYCR 390-414	4	Lectina tipo L con dominio de receptor quinasa de <i>Arabidopsis thaliana</i>	Lectina LegB 24-273 p-Kinasa 350-618	667 aa 74,097 Da Proteína inferida de la homología ^a	Q9M3D7
9	QLAVELDTS MNDFDPDGN HVGIDTRSIH PFAAK 149-182	5	Probable lectina Tipo L-con receptor quinasa de <i>Juglans regia</i>	Lectina LegB 24-277 p-Kinasa 346-609	668 aa 73,254 Da Proteína inferida de la homología ^a	A0A214E0E5
10	VQYDGNKISV FVPLVNLST NMNGPVLTL PLNLSDYLP E K 197-236	6	Serina/threonina proteína quinase de <i>Trema orientale</i>	Lectina LegB 28-284 p-Kinasa 346-622	669 aa 74,392 Da Proteína inferida de la homología ^a	A0A2P5B974
11	IFPNSSKLFV GFSAATGNA VSGHYILWW SFSTGRGSP HGLDISK 293-326	6	Proteína Quinasa de <i>Brassica campestris</i>	Lectina LegB 46-290 p-kinasa 359-626	713 aa 79,011 Da Proteína inferida de la homología ^a	A0A3P6CHY4
12	VSKSSKQGIN EYASEVR 283.-326	3	Proteína quinasa de <i>Helianthus annuus</i>	Lectina LegB 22-263 p-kinasa 347-616	62 aa 74,154 Da Proteína inferida de la homología ^a	A0A251RVA4

- a. Indica que la existencia de una proteína es probable porque existen ortólogos claros en especies estrechamente relacionadas.
- b. Se utiliza para entradas sin evidencia a nivel de proteína, transcripción u homología.

Se realizó un alineamiento múltiple de los 12 péptidos con la secuencia parcial de la lectina LSBo-I (Anexo D), mediante el programa bioinformático Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), de acuerdo con estos resultados se seleccionaron los péptidos 1, 2, 3, 11 y 12 debido a que mostraron ser los que alinearon con una mayor cantidad de aminoácidos conservados y en forma más completa.

Con los péptidos seleccionados, se hicieron alineamientos pareados obteniendo una mayor cobertura. El péptido N° 1, con un total de 17 aminoácidos, alineó doce (12) de ellos en las posiciones 189 a 201, abarcando una sección de la secuencia que no se había determinado, permitiendo postular 5 aminoácidos, **EAGSN**, en las posiciones 192-196. Respecto a la secuencia de LSBo-I, se presentaron cambios, **N/D** (190), **L/V** (200), propios de la técnica de secuenciación por masas, que se basa en la diferencia de peso molecular para identificar los residuos, siendo menos exacta y más subjetiva que la obtención de secuencia por el método Edman, donde se identifica con exactitud cada aminoácido. Otras sustituciones, están dadas por cambios como **Q/R** (201), **Y/R** (197) y **E/L** (98) (Figura 37-A).

El péptido N° 2 con 46 aminoácidos alineó en la secuencia hacia el extremo C-terminal, desde la posición 228 hasta la 274. Se conservan 10 aminoácidos y se presentan sustituciones en 7 posiciones con aminoácidos similares, **V/I**, **L/F**, **A/S**, **K/H**, **N/D**, **A/S**, **M/S**. Se postuló en la secuencia 18 aminoácidos que no se habían asignado (figura 37-B). El péptido N° 3 con 21 aminoácidos, presentó alineamiento de 11 aa de la posición 14 a la 24 de la secuencia, con 4 aminoácidos conservados y una sustitución idéntica (**T/S**), este se alineó hacia el extremo N-terminal (Figura 37-C). El péptido N° 11 con 44 aminoácidos, se alineó en extremo C-terminal de la secuencia, se conservaron 9 aminoácidos y hay 6 sustituciones con aminoácidos de propiedades bioquímicas similares (Figura 37-D). Finalmente, el péptido N°12 con un total de 17 aminoácidos, alineó 13 de ellos en la posición 207 a 219; con 3 aminoácidos conservados, 2 sustituciones por aminoácidos idénticos, que permitió añadir cuatro aminoácidos que no se habían determinado como se muestra en la figura 37-E.

A.

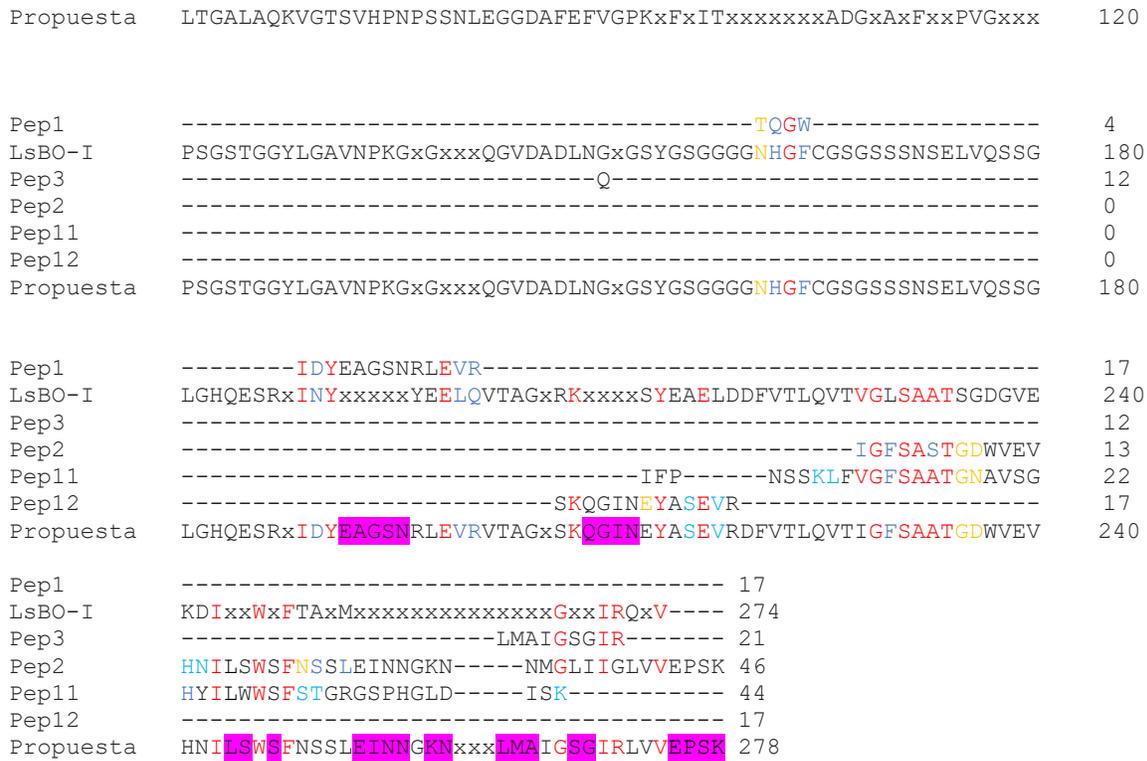


Figura 38. Alineamiento múltiple de los 5 péptidos seleccionados (1,2,3,11 y,12) con la lectina parcial de LSBo-I y la secuencia nueva (propuesta) de 278 aa.

En Lamiáceas únicamente se ha reportado la secuencia de lectinas de las especies *Glechoma hederacea* (Gleheda) y *S. miltiorrhiza* (LSM), específicas para el antígeno Tn. La secuencia de la LSM fue obtenida a partir de clonación molecular del gen codificado por 822 pb, y fue comparada con varias secuencias de lectinas de Leguminosas, mostrando un rango de identidad entre el 29 y el 43 % (154). Por su parte Gleheda (lectina obtenida de las hojas de *Glechoma hederacea*) se obtuvo por clonación de un gen de 819 pb que codifica los 273 aminoácidos reportados. Las dos especies están clasificadas entre las “lectinas tipo Leguminosa”. El alineamiento de estas dos secuencias mostró 42 % de identidad entre ellas (5). En la figura 34 se presenta la secuencia primaria de la LSBo-I propuesta en este trabajo.

```

      10      20      30      40      50      60
MATLLQNLIT ATSMEISRLD QDNQRXSXTL LTHPNFNGDL GHGQHGS GSG QSPSPSRFXA

      70      80      90      100     110     120
LTGALAQKVG TSVHPNPSSN LEGGDAFEFV GPKXFXITXX XXXXXADGXA XFXXPVGXXX

     130     140     150     160     170     180
PSGSTGGYLG AVNPKGXGX XQGVADLNG XGSYSGGGG NHGFCGSGS SENSELVQSSG

```

190 200 210 220 230 240
 LGHQESRXID YEAGSNRLEV RVTAGXSKQG INEYASEVRD FVTLQVTIGF SAATGDWVEV

 250 260 270
 HNILSWSFNS SLEINNGKNX XXLMAIGSGI RLVVEPSK

Figura 39. Estructura primaria parcial LSBo-I propuesta. Los residuos resaltados en rojo corresponden a aminoácidos importantes para el reconocimiento de los antígenos T y Tn, que han sido reportados en otras lectinas con especificidad por el antígeno T o Tn. La (x) indica la posición en la que no se identificó el aminoácido.

Se realizó el alineamiento múltiple de LSBo-I con 10 lectinas específicas por el glicotope Tn y que pertenecen a diferentes familias, de la familia Lamiacea: Gleheda y *S. miltiorrhiza*; de la familia Fabacea: *V. villosa* isolectina B4 (VVB4), *Arachis hypogaea* (PNA), *V. macrocarpa* (VML), *P. tetragonolobus* (WBAI) y *G. simplicifolia* (GcLA4); de la familia Moraceae: *Morus nigra* (Morniga- G) y *Artocarpus integrifolia* y de la familia Amarantacea, la lectina *Amaranthus caudatus*, para esto se empleó el programa Clustal Omega 12.1 (Figura 40). En azul claro se resaltan los aminoácidos altamente conservados en las diferentes especies. Por ejemplo, las glicinas (G) en las posiciones 126 y 130 de LSBo-I, se encuentra conservada en todas las especies de lectinas (*), estas han sido reportadas como aminoácidos importantes en la interacción con el carbohidrato en algunas de ellas (2,154,155).

```

Amarantina      ---MAGLPVIMCLKSNNNQKYLRYSQDNIQQYGLLQFSADKILDPLAQFEVE--PSKTYD 55
Morniga-G       MASSSFL-----S----- 8
Jacalina        ----- 0
WBAI            -----ETQSFNFD-----HFEENS-KELNLQ 20
VVB4            -----SESTSFST-----NFNPQ-ENLILQ 21
VML             -----SEVVSFSFT-----KFNPQ-KDIILQ 21
PNA            MASNSKIPLYLPLLASM--AIFLILLHSVNSVDVTTFLFD-----RFDQNN-ENLI IQ 50
GsLA4          -----XNTVNFTYPDFW-----SYSLKNGTEITFL 25
LSBo-I         ---MATLLQNLLTATSM--EISRLDQDNQRXSXTLLTHPN-----FNGDLGHGQHGS 47
SML            ---MAKLLQNLIPLLSA--IVLLAAANTVRSQTTSTFTYD-----FWGDQPNDLIYQ 47
Gleheda        -MANQKLPQTLVSI-----AALFLMVANMALSKTTHF-----AVPPALTFQ 40
  
```

```

Amarantina      GLVHIKSRYT-----NKYLVRWSPNHYWITASANEPDENKSNWACTLFKELYVEEG 106
Morniga-G       -----LSFLILLFSISSANTRKWSLSN-----GLDQKPIISIIEA 42
Jacalina        ----- 0
WBAI            RQASIKSSGV-LELTKLTKNGVPVWKSTGR-----ALYAEIKIWD 61
VVB4            EDALVNSKGT-LELT---KNGKVPVESLGR-----NC----- 49
VML             GDALVTSK GK-LQLTKV-KDGKPVDSHSLGR-----ALYAAF IHIWDD 61
PNA            GDASVSSNGA-LQLTRVDSSGVPQGGSVGR-----ALYSEPI SLYDK 91
GsLA4          GD-ATRIPGA-LQLTKTDANGNPVRSSAGQ-----ASYSEPVFLWDS 65
LSBo-I         GSGQSPSPSRFXALT-----GALAQKVG T-----SVHPN P S S N L E G 83
SML            GSAHFPSRRTTFLRLT-----DALSSQVGR-----VVHSN P V Q F S Q G 83
Gleheda        GDAFDPNDTSFIRLT-----TSHTWSVGR-----VMYSKELT DWGE 76
  
```

```

Amarantina      NMKKVRLRHVQLGHYTQNYTVGGSFVSYLFAESSQ---IDTGSKDVFHVIDWKSIFQFPK 163
Morniga-G       AIGVSEDL-----LNLNGMEAKNN---QQSGKSQTIVVGTWGAQVTS-- 81
Jacalina        -----MASQTITV P WGGP----- 14
WBAI            TTGNVASF-----ETRFSE NITQPYAYPEPADGLTFFMVPNSP--G-- 101
VVB4            --TTLASF-----TTSFSFVMSAPNS-LDVADGLAFFLAPP T Q--P-- 86
VML            STDRVASF-----ATSFSFVVEAPDE-SKTADGIAFFLAPPD T Q--P-- 100
  
```

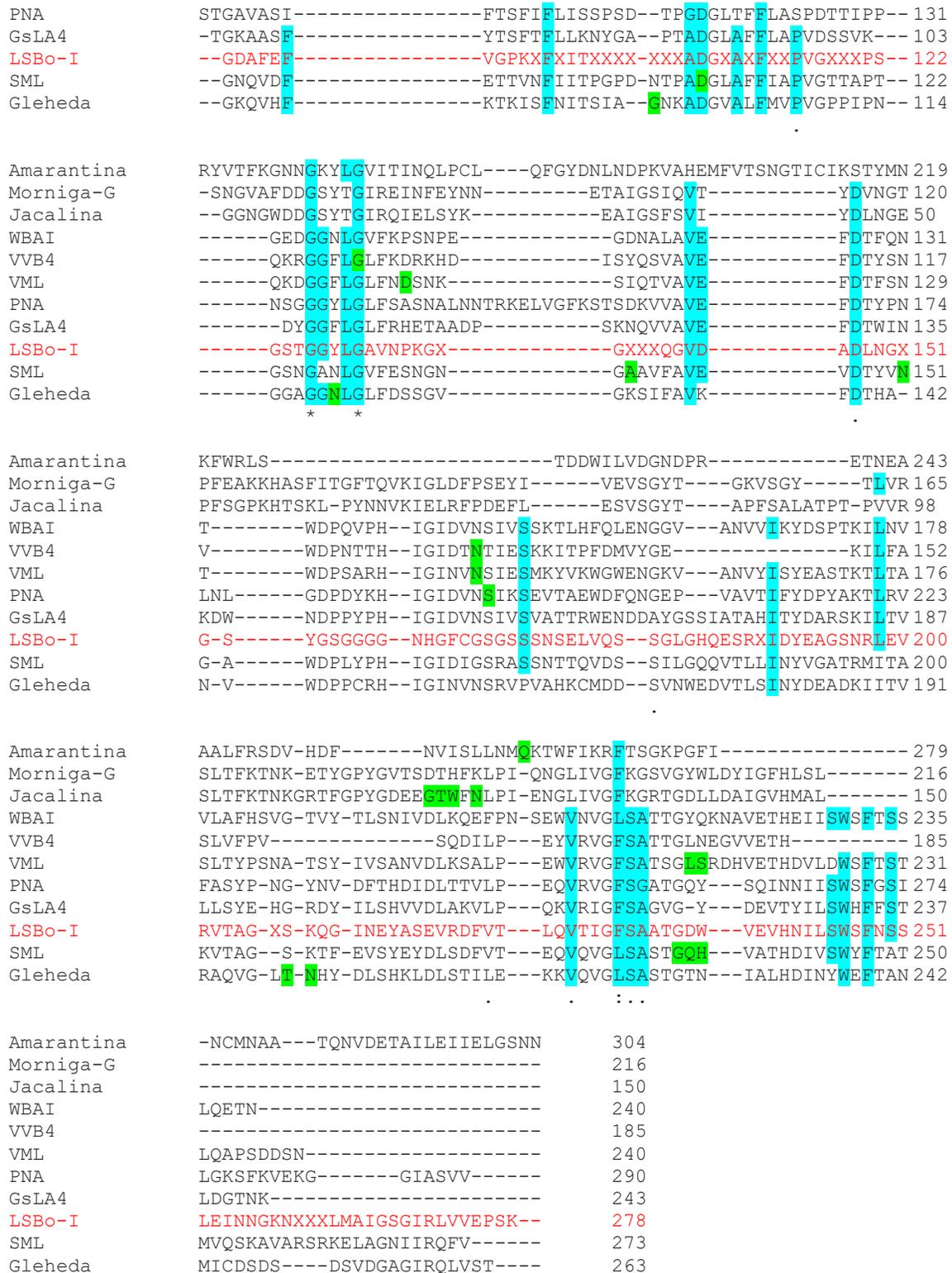


Figura 40. Alineamiento múltiple de LSBo-I (*S. bogotensis* l) (en rojo), con 10 lectinas específicas por el glicótopo Tn pertenecientes a diferentes familias. Amarantina (*Amaranthus caudatus*), Morniga-G (*Morus nigra*), Jacalina (*Artocarpus integrifolia*), WBAI (*P. tetragonolobus*), VVB4 (*Vicia villosa* B4), VML (*V. macrocarpa*), PNA (*Arachis hypogaea*), GsLA4 (*Griffonia simplicifolia*), SML (*S. miltiorrhiza*), Gleheda (*G. hederácea*).

En rojo (LSBo-I), Azul (aminoácidos conservados en las secuencias), verde (aminoácidos que interactúan en el reconocimiento al carbohidrato).

LSBo-I mostró mayor porcentaje de identidad 35,16 y 23, 37 % con las Lectinas de *S. miltiorrhiza* y Gleheda pertenecientes también a la familia Lamiaceae, y menor porcentaje (11,68 %) con la lectina de *Amaranthus caudatus* perteneciente a la familia Amaratacea, con las demás presentó porcentajes de identidad entre 15 y 21 % (Figura 41-A). Estos porcentajes se corroboran con al árbol filogenético, en el que se muestra que LSBo-I está más relacionada con la lectina de *S. miltiorrhiza* y Gleheda (las tres pertenecen a la familia Lamiaceas), pero a su vez también están filogenéticamente relacionadas con las 5 lectinas de la familia Fabaceae (WBAI, VVB4, VML, PNA y GsLA4) todas ellas con plegamiento de lectina tipo Leguminosa. También se observa que LSBo-I esta menos relacionada filogenéticamente con las lectinas de la familia Moraceae (Morniga-G y *A. integrifolia*) cuyas lectinas pertenecen a la familia Jacalina y de la familia Amarantacea, la *A. caudatus*, de la familia de lectinas Amarantina (Figura 41-B).

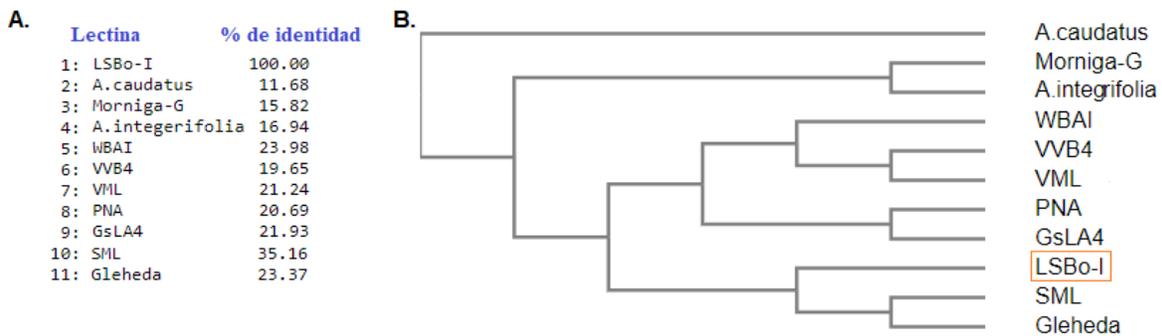


Figura 41. (A) Porcentaje de identidad entre las lectinas alineadas con LSBo-I. (B) árbol Filogenético.

Para las lectinas de Lamiaceae, las interacciones con el glicotipe T se dan a través de puentes de hidrógeno que se establecen con tres aminoácidos Asp, Gly, Asn implicados directamente en el reconocimiento. Con la lectina de Gleheda se establecen puentes de hidrógeno entre los OH, 3, 4, 5, 6 y el grupo N-acetil del GalNAc con los aminoácidos Asp73, Gly93, Asn117, Thr201 y Asn202 y la His115 que interactúa con el anillo piranosa del azúcar (3). En la interacción con la lectina de *S. miltiorrhiza* son importantes los aminoácidos Asp107, Ala139, Asn151, Gly234, Gln235 e His236 en el sitio de interacción con el carbohidrato (2); en la secuencia de LSBo-I los aminoácidos Asn149 y Gly236 están en posiciones cercanas, lo que podría sugerir que son también residuos críticos en el reconocimiento del antígeno Tn (5).

En fabáceas, las lectinas de *Vicia villosa isolectina* B4 (VVLB4), los residuos Asp85, Gly103 y Asn129 hacen parte de una red de ocho puentes de hidrógeno formados con OH, 3, 4, 6 y el grupo N-acetil del GalNAc del antígeno Tn. La Tyr127 se identificó como un aminoácido importante en la interacción de VVLB4 con el anillo piranosido del GalNAc, interacciones que son determinantes para la especificidad de esta lectina (155). Con la lectina de *V. macrocarpa*, el anillo interactúa con el grupo aromático de Phe129 y la cadena alifática de Leu215. Los puentes de hidrógeno se presentan entre los grupos OH 3, 4 y 6-OH, así como con el oxígeno del grupo carbonilo del grupo N-acetil con Gly107 y las cadenas laterales de Leu215, Ser216, Asp109 y Asn131 (105). En la lectina de *A. hypogaea* (PNA) se reporta como aminoácidos importantes en la interacción con el carbohidrato, Asp83, Gly104, Asn127 y Ser211. El anillo piranosido de Gal interactúa con la Tyr125. El residuo GalNAc sobresale del sitio de unión a carbohidrato e interactúa mediante dos enlaces de hidrógeno con Ser211 y Gly212 (156).

Para la lectina de *A. caudatus* (Amarantina), perteneciente a la familia Amarantacea, se reportó, que los residuos que forman el sitio de unión a carbohidrato, reconociendo específicamente el antígeno T son: Asn74, His75, Tyr76, Trp77 y Gln262, y presenta interacciones mediadas por cinco puentes de hidrógeno con Gal y tres con GalNAc (5). La especificidad en el reconocimiento también está generada por algunas moléculas de agua presentes, que forman puentes de hidrógeno con los dos monosacáridos del antígeno T (157).

En la familia Moraceae se ha encontrado que el sitio de unión a carbohidrato de la jacalina, lectina de *Artocarpus integrifolia* (158) que interactúa con el antígeno T, mediante el residuo de GalNAc a través de la formación de nueve puentes de hidrógeno entre O4, O5, O6 y el grupo N-acetil con Gly8, Gly121, Tyr122, Trp123 y Asn125.

Con respecto al análisis de aminoácidos de la LSB0-I (Tabla 8), se observa que la proteína se compone de alto porcentaje de residuos de glicina (Gly) 14,4%, serina (Ser) 12 %, leucina (Leu) 7,6 %, alanina (Ala) 5,8 % y glutamina (Gln) /ácido glutámico (Glu) 4,3 %. En los péptidos analizados no se encontró ningún residuo de cisteína de los cuatro (4) que fueron propuestos, y que forman dos puentes disulfuro intracatenarios (15,16); lo que indica que la hidrólisis y el método de secuenciación realizado no fue adecuado para hacer su detección. Adicionalmente, el análisis de aminoácidos que incluyó un proceso de oxidación peróxica y mostró la presencia de ácido cisteco, y ausencia de metioninsulfona. Contrario a las dos secuencias propuestas que incluyeron dos (2) y tres (3) residuos de metionina en su estructura, lo que es inusual debido a que este aminoácido es escaso en proteínas. Para las lectinas de Gleheda, *S. sclarea* y *S. miltiorrhiza* se reportaron en el análisis de

aminoácidos, uno (1) y tres (3) residuos de metionina (Met) respectivamente. De acuerdo con (159), no hay enlace S-S presente en la subunidad de 26 kDa de la lectina de *Molucella laevis* (MLL), mientras que un enlace S-S une las dos subunidades para *S. sclarea* (SSL) y Gleheda (14,82).

Tabla 8. Análisis de aminoácidos obtenidos de la LSBo-I

N°	Aminoácido		LSBo I, Wilches (5)	LSBo_I, Este trabajo**	Vega, Pérez (4)
1	Ala	A	16	16	36
2	Cys	C	1	1	4
3	Asp	D	11	10	*
4	Glu	E	9	12	*
5	Phe	F	9	10	4
6	Gly	G	37	40	92
7	His	H	6	7	4
8	Ile	I	4	9	6
9	Lys	K	5	6	21
10	Leu	L	21	21	10
11	Met	M	2	3	0
12	Asn	N	11	17	15 (Asx)
13	Pro	P	9	10	13
14	Gln	Q	10	11	42 (Glx)
15	Arg	R	6	8	3
16	Ser	S	25	33	60
17	Thr	T	16	13	13
18	Val	V	13	16	10
19	Trp	W	1	2	9
20	Tyr	Y	5	4	5
No determinados (x)			58	31	0
N° total de residuos de la secuencia			274	278	
No Residuos			216	247	347
% cobertura			62	71	
Masa molecular (Da)			22119	25710	32509

(5) Wilches Torres MA. Aproximación a la estructura primaria de lectinas específicas para el antígeno Tn e identificación de nuevas lectinas específicas para glucosa/manosa en semillas de *S. bogotensis* y *Lepechinia bullata*. Universidad Nacional de Colombia-Sede Bogotá; Análisis de aminoácidos obtenido de la secuencia parcial de LSBo-I.

(4). Vega N, Pérez G. Isolation, and characterisation of a *S. bogotensis* seed lectin specific for the Tn antigen. *Phytochemistry*. 2006;67(4):347–55. Análisis de aminoácidos realizado por hidrólisis total de la proteína en condiciones ácidas y/o por oxidación perórmica. Peso molecular 38702 Da con 16 % de carbohidratos.

** Este trabajo. Análisis de aminoácidos obtenido de la secuencia parcial de LSBo-I.

Si se tiene en cuenta el análisis de aminoácidos faltaría por añadir en la secuencia, residuos de Ala, Cys, Gly, Lys, Ser y Trp, aunque se observa también un exceso de aminoácidos hidrofóbicos. Para LSBo-I se logró secuenciar 25710 Da de masa molecular (https://web.expasy.org/compute_pi/) de los 32509 Da (16) determinados previamente, aún faltan por secuenciar 6799 Da que corresponden con ~62 aminoácidos. En la secuencia reportada para la lectina de *S. miltiorrhiza* no se reportaron cisteínas (2), mientras que en Gleheda se reportaron tres, dos que forman un puente disulfuro y un tiol libre (154).

6.1.3.1. Modelamiento de la estructura terciaria

Se realizó predicción de la estructura terciaria de la proteína empleando el método de homología con los servidores I-TASSER (<https://zhanggroup.org/I-TASSER/>) y Phyre² (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>). En cada programa se cargó la secuencia de aminoácidos en formato FASTA, en las diferentes herramientas conforme al método de predicción, se seleccionó un análisis de construcción intensa y por defecto, con la finalidad de obtener resultados más exactos y precisos.

El programa I-TASSER (Intractive Threading Assembly Refinement) es una plataforma unificada y automatizada, que permitió la predicción de la estructura de la proteína mediante threading (enhebrado) en la que usó como plantilla cinco (5) estructuras cristalográficas de lectinas de fabáceas, la isoforma de la Lectina de *Tetragonolobus lotus* (PDB: 2EIG,) con 234 aminoácidos que reconoce L-fucosa, lectina de *Maackia amurensis* (MAL) (1DBN) con 239 aminoácidos que reconoce α 2-3 ácido siálico, lectina de *Dioclea altissima* (7LJG) con 237 aminoácidos específica por residuos de manosa, fitohemaglutina (PHA) (1G8W) de *Phaseolus vulgaris* con 233 aminoácidos y la lectina de *Pterocarpus angolensis* lectin (1N3O), con 252 aminoácidos y específica por manosa (Man)/ Glucosa (Glc). Como resultado de la predicción, se generaron cinco (5) modelos para LSBo-I; la mejor predicción arrojó un C-score=-0.48, TM-score = 0.65±0.13 RMSD = 6.8±4 (Figura 42). El sistema de puntuación C-score, está relacionado con la calidad del modelo, generalmente está en el rango de [-5,2], donde un puntaje C de valor más alto significa un modelo confiable y viceversa. Por su parte, TM-score y RMSD miden la similitud estructural entre dos estructuras y la calidad del modelo.

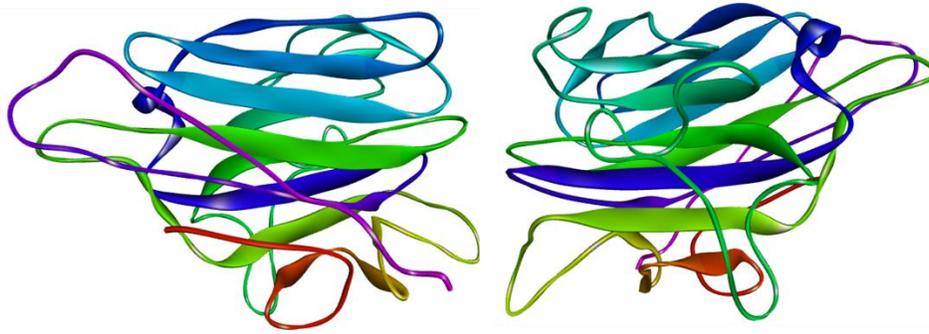


Figura 42. Estructura obtenida con I-Tasser sin minimización de energías. Izquierda, vista posterior de 6 hebras antiparalelas. Derecha, vista frontal bolsillo de unión a ligando característico de leguminosas.

La validación del modelo se realizó por Ramachandran con Procheck vía PDBsum (<http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>) y MolProbity (servicio web de validación de estructuras para diagnosticar problemas en modelos 3D de proteínas), calculando la validación global y local para contactos de todos los átomos, geometría covalente y conformación tipo Ramachandran y rotámeros. El modelo predicho por I-Tasser (Figura 42) presentó una geometría baja, solamente el 57 % de los ángulos diedros de residuos distintos a prolina y glicina dentro de las regiones más favorecidas y 10 % de los residuos se encuentran por fuera de ellas según las validaciones realizadas por MolProbity y por Ramachandran (Anexo E).

El modelo obtenido por Phyre², utilizó como modelo la lectina de *Bowringia milbraedii* con 240 aminoácidos (PDB 2FMD), que reconoce Man (α 1-2) Man. Esta predicción presentó buena estereoquímica ya que el 98,7 % de los ángulos diedros de residuos distintos a prolina y glicina se encuentran en las regiones más favorecidas respecto en el diagrama de Ramachandran (Anexo E). Se realizó minimización de energías de los dos modelos propuestos I-Tasser y Phyre2 con el servidor GalaxyWEB ya que fue el único que mejoró la mayoría de los ítems que evalúa MolProbity. Sin embargo, el refinamiento del modelo obtenido por I-Tasser no optimizó los valores de MolProbity, en comparación con las correcciones estereoquímicas que obtuvo el modelo de Phyre2. A continuación, se presentan los resultados de validación por MolProbity y Prochek del modelo predicho por Phyre2 (Figuras 43 y 44).

Protein Geometry	Poor rotamers	2	1.30%	Goal: <0.3%
	Favored rotamers	150	97.40%	Goal: >98%
	Ramachandran outliers	1	0.50%	Goal: <0.05%
	Ramachandran favored	186	93.00%	Goal: >98%
	Rama distribution Z-score	-1.73 ± 0.50		Goal: abs(Z score) < 2
	C β deviations >0.25Å	8	4.82%	Goal: 0
	Bad bonds:	4 / 1489	0.27%	Goal: 0%
	Bad angles:	45 / 2018	2.23%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 9	0.00%	Expected: ≤ 1 per chain, or $\leq 5\%$
Low-resolution Criteria	CaBLAM outliers	11	5.6%	Goal: <1.0%
	CA Geometry outliers	7	3.54%	Goal: <0.5%
Additional validations	Tetrahedral geometry outliers	1		

In the two column results, the left column gives the raw count, right column gives the percentage.
Key to table colors and cutoffs here: [?](#)

Figura 43. Validación predicha del modelo obtenido por Phyre2 refinado por GalaxyWEB empleando MolProbity.

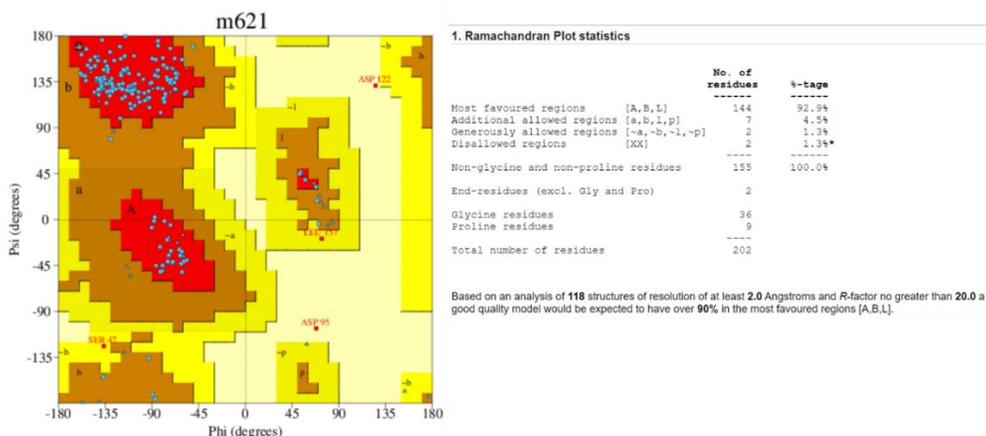


Figura 44. Validación predicha del modelo obtenido por Phyre² refinado por GalaxyWEB empleando Ramachandran.

La estructura terciaria modelada para LSBo-I por Phyre² con minimización de energías por GalaxyWEB, presenta características estructurales comunes en las lectinas de tipo vegetal. En general, este tipo de lectinas se caracterizan por tener una estructura tridimensional formada principalmente por hojas β que están conectadas por giros α y β . La estructura presenta rasgos característicos a las lectinas tipo leguminosa (L) que tienen una hoja β frontal curva con 5 a 7 hebras β . En este caso tiene 6 hebras (amarillas) (Figura 45-A), una hoja β posterior plana formada por seis (6) hebras antiparalelas, como se muestra en la figura 45-B y adicional a esto, aunque en este modelo no se alcanzan a percibir, posiblemente por la falta de los aminoácidos que componen la estructura, las lectinas tipo L también tienen una hoja superior de cinco (5) hebras que las mantiene unidas. Estas hojas forman una estructura aplanada en forma de domo con bucles localizados en la parte superior del monómero formando el dominio de reconocimiento de carbohidratos (CDR), el cual se puede ver en la vista lateral del modelo señalado con una flecha roja (Figura 45-C) (76).

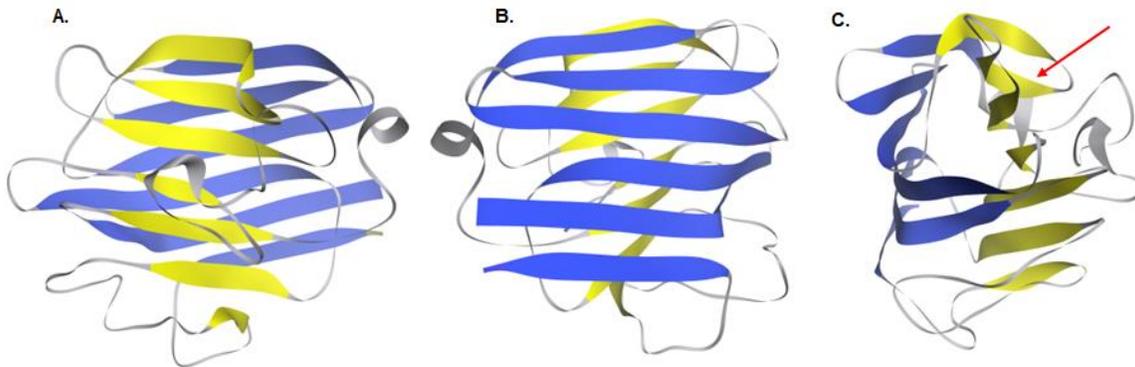


Figura 45. Estructura predicha por Phyre² con minimización de energías por GalaxyWEB para LSBo-I. **(A)** Hoja β frontal con 6 hebras (amarillas). **(B)** hoja β posterior con 6 hebras antiparalelas planas (Azules). **(C)** Vista lateral, espacio entre las dos hojas (frontal y posterior) y formación del bolsillo de unión a ligando característico de leguminosas (flecha roja) determinado por Discovery studio.

El bolsillo de unión a carbohidrato se puede percibir mejor en la Figura 46, en ella se resaltan en rojo los aminoácidos importantes en la formación del bolsillo, entre ellos se encuentran tanto los aminoácidos de interacción directa con el carbohidrato, en este caso con GalNAC, como los aminoácidos que ayudan a estabilizar el bolsillo. En leguminosas, el bolsillo CRD está formado por cuatro bucles adyacentes entre sí, estas regiones presentan una alta variabilidad de residuos involucrados en la determinación de la especificidad, y se encuentra muy cerca de los sitios de unión de metales, lo que puede ayudar en la unión (76,77).

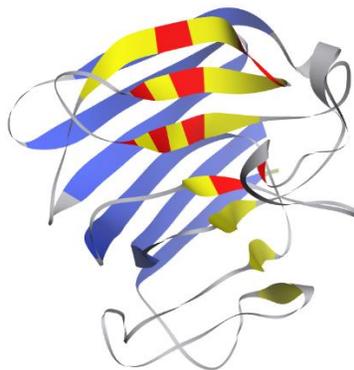


Figura 46. Predicción de la estructura 3D de LSBo por Phyre² en la que se resalta en rojo el bolsillo de unión a ligando a GalNac determinado por Discovery studio.

Las hojas frontal y posterior en las lectinas tipo L son casi paralelas y se pueden encontrar a una distancia cercana a 13 Å, en este modelo se determinaron distancias promedio de 12,8 Å entre las hojas frontal y posterior (Figura 47). Se ha encontrado que esta distancia es importante para la formación de un cluster de

residuos hidrofóbicos cuyas cadenas laterales están posicionadas entre las hojas frontal y posterior dando un plegamiento estable (76).

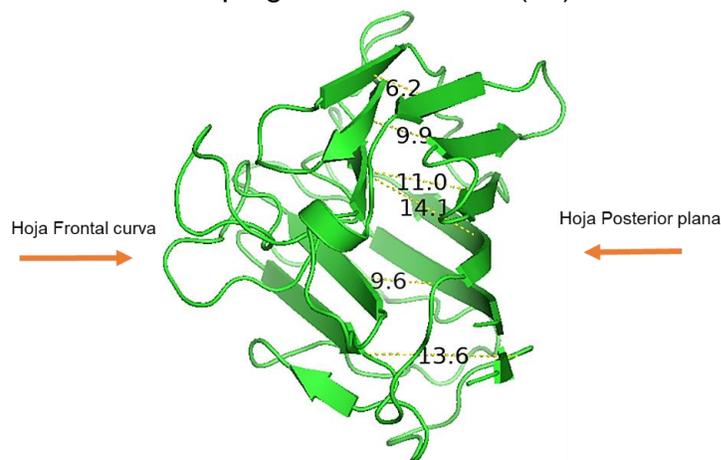


Figura 47. Predicción de la estructura tridimensional de LSB0-I predicha por Phyre² con distancias entre las hojas frontal y posterior determinadas con Pymol.

6.1.3.2. Acople molecular

Teniendo el mejor modelo por Phyre² se realizó la simulación de acople molecular con Autodock Vina, que se basa en la predicción de modos de unión característicos de un ligando a una proteína, cuya estructura tridimensional es conocida. Esta técnica utiliza una función de puntuación basada en la energía de interacción durante el acoplamiento en términos de la energía libre ΔG (160) o energía del “enlace” no covalente entre ligando y receptor, y mediante un gradiente de optimización evalúa la mejor posición dentro de la macromolécula el cual estima la pérdida de entropía al momento de unión con la proteína, donde se da un total de seis evaluaciones por pares, cada una incluye términos energéticos de Van Der Waals, enlaces de hidrogeno, interacciones de Coulomb y solvatación (161).

La simulación de acople molecular con Autodock-vina se realizó empleando tres ligandos: N-acetilgalactosamina (GalNac), glicotopes T (Gal β 1,3GalNAc α 1-Ser/Thr) y glicotope Tn (GalNAc α 1-Ser/Thr). El grid de acoplamiento fue: Coordenadas npts x: 40, y: 22, z:46, centrado en x: 19,277, y: 23,761, z: 19,127. En la Tabla 9 se presenta un resumen de la afinidad y los sitios de unión que se obtuvo con cada uno de los ligandos. El complejo que presentó menor energía fue el de LSB0-I – glicotope T (-6.9 Kcal/mol) (Figura 48), seguido de glicotope Tn (-5,9 Kcal/mol) (Figura 49), y por último GalNac (-5.3 Kcal/mol) (Figura 50). Estos datos son comparables con los acoples realizados en trabajos previos, en los que las energías de acople de la lectina LSB0-I con estos tres ligandos fue de (-6.6) para el ligando T, (-6.1) para Tn y (-5.8) para GalNac (5). Se puede determinar que en

este trabajo aumentó la afinidad a ligando con el glicotope T, ya que su afinidad aumento y también hay un mayor número de aminoácidos involucrados en las interacciones con este ligando.

Tabla 9. Afinidad que presenta el modelo a diferentes ligandos.

	Ligando		
	Glicotope T	Glicotope Tn	GalNac
Afinidad (kcal/mol)	-6,9	-5,9	-5,3
Aminoácidos de unión	N137, Y130, N126, G108, G128, S146, S145, D95, K116	D95, G108, N126, S146, G144	S149, N137, Y130, N126, G144, S146, S145

En las tablas 10, 11 y 12 se describen las diferentes interacciones que se dan entre los aminoácidos de LSBo- I y los tres ligandos seleccionados. El complejo LSBo-I – glicotope T, muestra la posible interacción del antígeno T con los aminoácidos Asn126, Asn137, Tyr130, Gly108, Gly128, Ser146, Ser145, Asp95, Lys116 (Figura 48-C) y cuya energía de unión fue de -6.9 Kcal/mol; la menor encontrada entre los modelos generados con diferentes ligandos, lo que podría sugerir que la estructura tiene preferencia por la unión con este antígeno; su sitio de reconocimiento puede encajar trisacáridos, corroborándose con los resultados obtenidos en trabajos previos donde también presentó mayor interacción con este ligando (5).

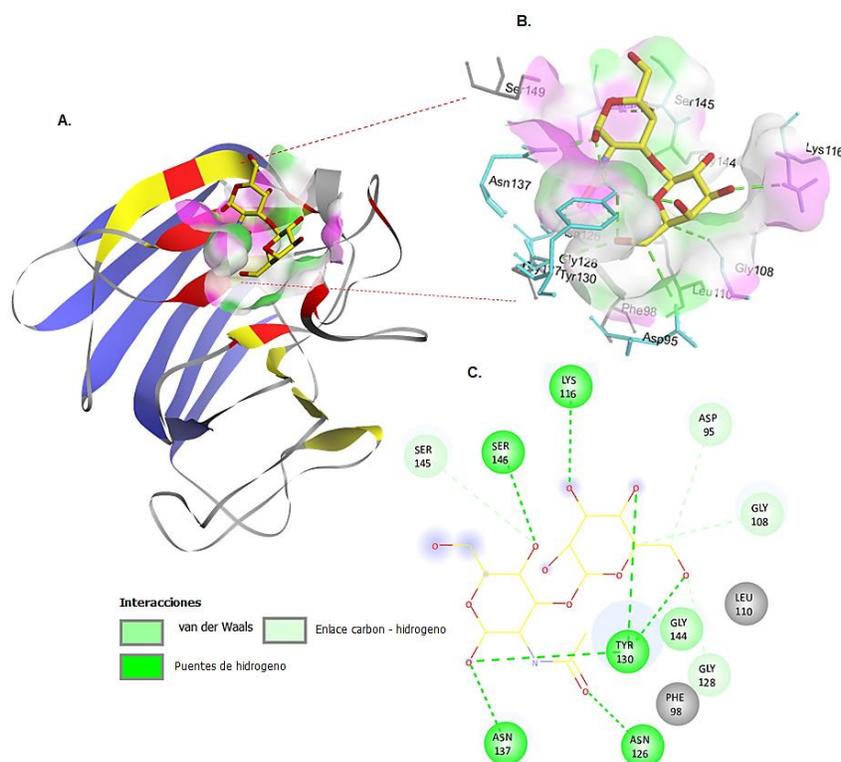


Figura 48. Modelo de acoplamiento entre LSB0-I y el glicotipe T obtenido por Autodock Vina. (A) Localización de bolsillo en LSB0-I (B) Interacciones receptor – ligando, (Azul claro) aminoácidos que interactúan directamente con el ligando, (gris) aminoácidos que estabilizan el bolsillo (C) Visualización 2D de las diferentes interacciones receptor – ligando determinado por Discovery studio.

Tabla 10. Descripción de las interacciones de los aminoácidos en el sitio de reconocimiento con el carbohidrato T.

LSB0-I – glicotipe T		Armstrong (Å)
Lys116	H de la amina de K116 al O3 de Gal	2,7
Asn126	H de la amida de N126 al O7 del acetilo de GalNac	2,4
Tyr130	H de fenol de Y130 al O1 de GalNac	2,3
Asn137	H de la amida de N137 al O1 de GalNac	2,1
Ser146	H del amino peptídico con O4 de GalNac	2,8
Ser145	C α O4 de GalNac	3,2
Gly128	C α al O6 de Gal	3,8
Gal	O4 al H del fenol de Y130	3,1
	O6 al H del fenol de Y130	3,0
	C6 al O del carboxilo de D95	3,6

*En color se muestran las interacciones que son iguales con otro ligando

El acople de LSB0-I con el glicotipe Tn, predijo interacciones del ligando con los aminoácidos, Asp95, Gly108, Asn126, Ser146 y Gly144 (Figura 49-C), en este caso, se presentó una afinidad de -5,9 Kcal/mol, que se ve representada en menor número de interacciones con el ligando, respecto a las presentadas por el glicotipe T.

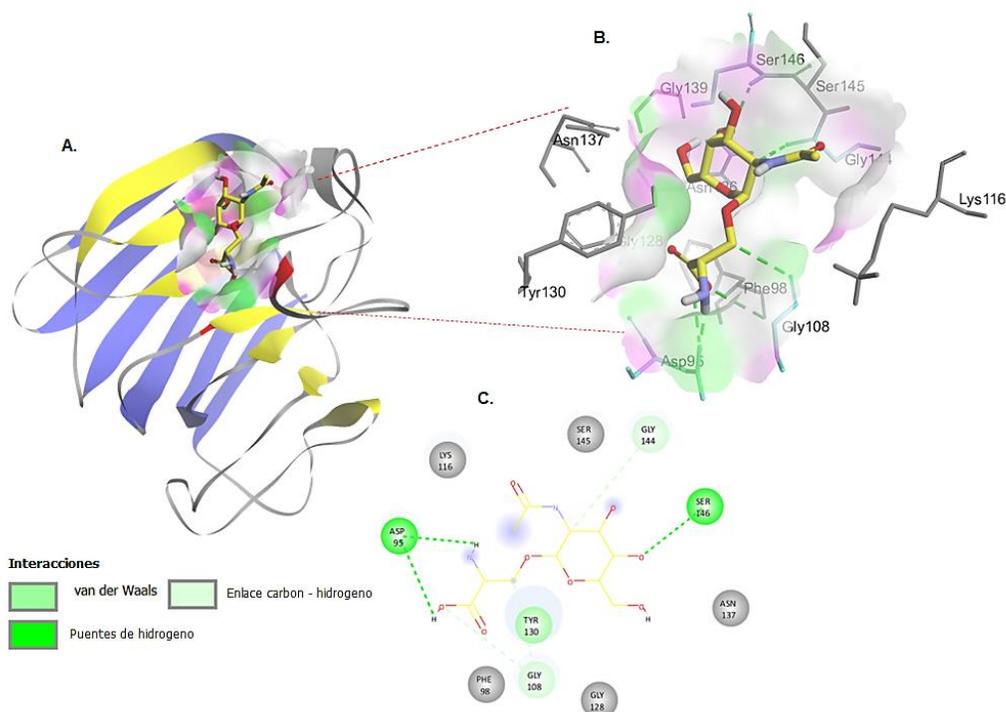


Figura 49. Modelo de acoplamiento entre LSB0-I y glicotopo Tn, obtenido por Autodock Vina. **(A)** Localización de bolsillo en LSB0-I **(B)** Interacciones receptor – ligando, (Azul claro) aminoácidos que interactúan directamente con el ligando, (gris) aminoácidos que estabilizan el bolsillo **(C)** Visualización 2D de las diferentes interacciones receptor – ligando determinado por Discovery studio.

Tabla 11. Descripción de las interacciones de los aa en el sitio de reconocimiento con el Ligando – Antígeno T.

LSBo-I – Tn glicotopo		Armstrong (Å)
Tn	C2 GalNac a carboxilo alfa de G144	3,4
Tn	C _β Serina de Tn a carboxilo alfa de G108	3,4
Ser146	H del amino alfa con O4 de GalNac	2,3
Tn	OH de Serina de Tn con O carboxílico R de D95	2,4
Tn	H amino de la serina de Tn a O carboxílico R de D95	1,9
Gly128	C _α al O7 del OH de la serina de Tn	3,5

*En color se muestran las interacciones que son iguales con otro ligando

Por último, los aminoácidos implicados en la interacción con GalNac, son Ser149, Asn137, Tyr30, Asn126, Gly144, Ser146 y Ser145 (Figura 50-C). Como se muestra en la tabla 10, residuos como Tyr130 se encuentra en la interacción de unión de LSB0-I tanto con T como con GalNac, este se ha reportado previamente como un residuo importante dentro del sitio de unión a carbohidrato, en otras lectinas como *Vicia villosa* B4 (155) y *Arachis hypogaea* (PNA) (156).

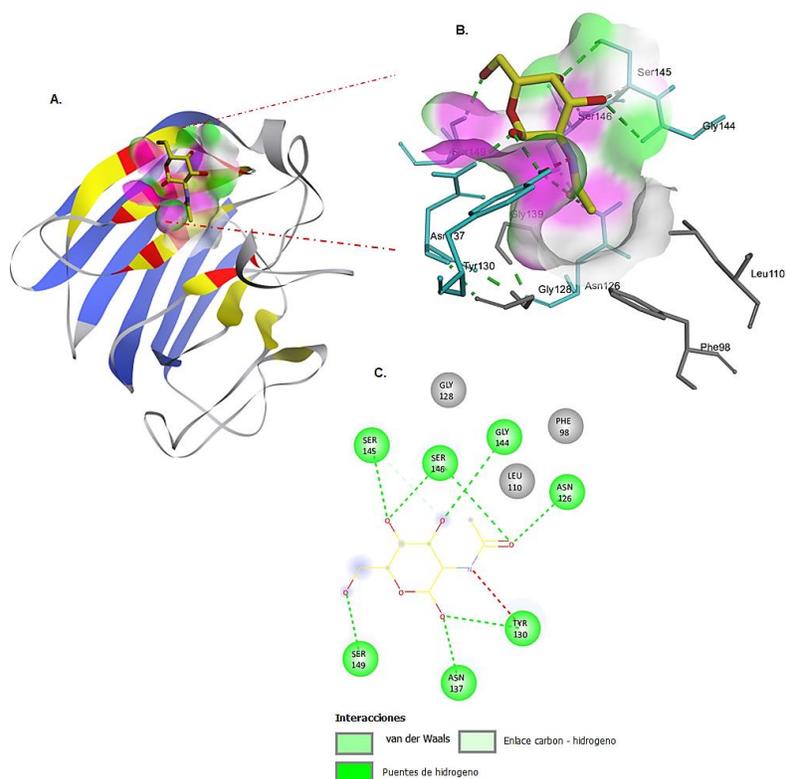


Figura 50. Modelo de acoplamiento entre LSBo-I y GalNAC obtenido por Autodock Vina. (A) Localización de bolsillo en LSBo-I (B) Interacciones receptor – ligando, (Azul claro) aminoácidos que interactúan directamente con el ligando, (gris) aminoácidos que estabilizan el bolsillo (C) Visualización 2D de las diferentes interacciones receptor – ligando determinado por Discovery studio.

Tabla 12. Descripción de las interacciones de los aminoácidos en el sitio de reconocimiento con los carbohidratos GalNAC.

LSBo-I - GalNac		Armstrong (Å)
Asn 126	H de la amida de N126 al O7 del acetilo en GalNac	2,6
Tyr130	H de fenol de Y130 al O1 de GalNac	2,6
Asn137	H de la amida de N137 al O1 de GalNac	2,1
Ser146	H del amino peptídico con O4 de GalNac	2,2
	H del alcohol con O7 acetilo en GalNac	2,6
Ser149	H del alcohol con O6 del GalNac	2,2
Ser145	Del C α al O3 del GalNac	3,4
GalNac	H del O3 al OH de S145	

*En color se muestran las interacciones que son iguales con otro ligando

Tabla 13. Aminoácidos identificados en el sitio de unión de LSBo-I, que presentan interacción con los ligandos ensayados.

Residuo	Glicotipe T	Glicotipe Tn	GalNac
Asn137	x		x
Tyr130	x		x
Asn126	x	x	x
Gly108	x	x	
Ser145	x		x
Asp95	x	x	
Gly144		x	x
Ser146	x	x	x
Gly128	x		
Lys116	x		
Ser149			x

Al realizar la superposición del modelo propuesto en este trabajo con el 71 % de cobertura de la secuencia, respecto al modelo 3D propuesto en trabajos previos (5) con el 62 % de la secuencia de la lectina LSBo-I (Figura 51), se puede observar que aunque el porcentaje de cobertura en 9 %, el modelo 3D es más definido respecto a la estructura característica de las leguminosas, tanto en la hoja frontal, cuyas hebras se ven mejor resueltas (Figura 46-A), como en la posterior (Figura 46-B), observando el plegamiento tipo Leguminosa que presenta la lectina de *S. bogotensis -I*.

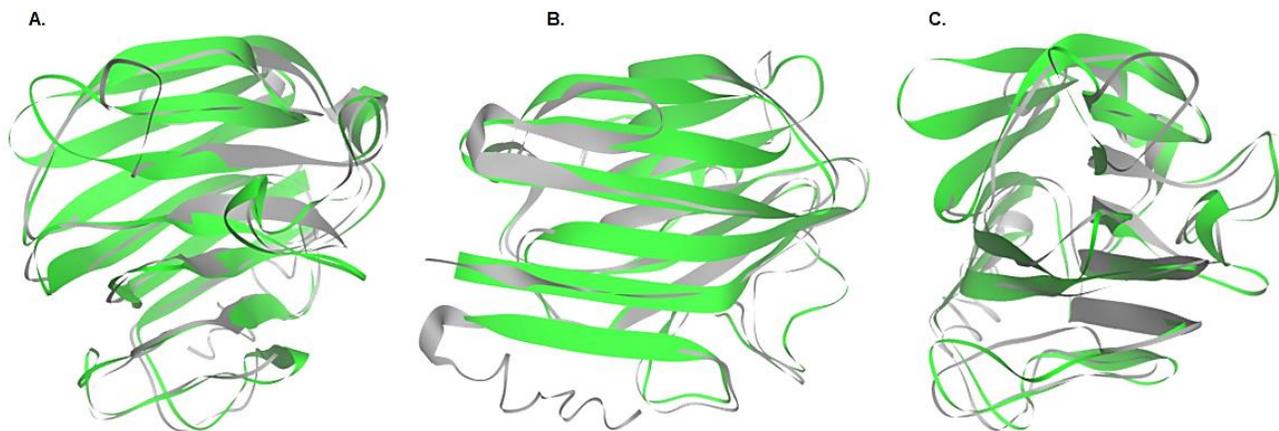


Figura 51. Alineamiento tridimensional de las lectinas de *S. bogotensis-I* propuesta en este trabajo (Verde), con la propuesta en trabajos previos (gris), (A) Vista frontal, (B) Vista posterior, (C) Vista lateral.

6.2. Ensayos preliminares para la obtención de LSBo-I recombinante

La clonación y expresión de lectinas recombinantes han permitido avanzar en los estudios relacionados con su estructura, función, actividad y aplicaciones. La obtención de recombinantes es un proceso de menor dificultad que permite obtener moléculas puras y en cantidad suficiente, con secuencias de aminoácidos definidas y por lo tanto sus propiedades moleculares se pueden definir mejor. Adicional a esto, la expresión en huéspedes heterólogos ha permitido estudiar la influencia de diversas modificaciones postraduccionales en el plegamiento y el direccionamiento de esta clase de proteínas de unión a carbohidratos (91). Algunas de las lectinas específicas que reconocen los glicotopes Tn/T obtenidas por medio de recombinantes son las lectinas de *V. Macrocarpa* (91) y *S. miltiorrhiza* (7) y de *Glechoma hederácea* (*Gleheda*) (3). Para alcanzar mayor cobertura y evaluar la exactitud de la secuencia de la LSBo-I, se realizaron ensayos para la obtención de LSBo-I recombinante; extracción de ARN, obtención de ADNc y estudios de amplificación usando primers basados en la secuencia parcial de LSBo-I.

Se seleccionaron semillas verdes como material de partida para realizar la extracción, puesto que las lectinas de plantas están en mayor proporción en este tejido, en especial las lectinas de Leguminosas dicotiledóneas, debido a que se almacenan durante el proceso de maduración (1). Aunque también se encuentran en otros tejidos vegetales, como tallo, raíz y hojas, pero su abundancia es menor y está regulada durante el proceso de desarrollo de la planta (162). Sin embargo, también se seleccionaron las hojas para realizar la extracción de ARN y comparar

el proceso de extracción en los dos tejidos. Las semillas verdes que se extrajeron de las flores tenían ~ 2 mm de diámetro y color beige fueron separadas del material vegetal y almacenadas en crioviales a -80 °C (Figura 52). Igualmente, se seleccionaron hojas verdes, de aproximadamente 5 cm, se colectaron en tubos cónicos de 50 mL y se almacenaron a -80 °C.



Figura 52. Semillas de *S. bogotensis* en estado inmaduro. Imágenes propias.

6.2.1. Extracción ARN por los métodos de CTAB y Trizol

La extracción de ARN de buena calidad, con relación de absorbancias (260/280) entre 1,85 – 2,0 sin degradación apreciable en geles de agarosa y en cantidades apropiadas, es fundamental para la construcción de bibliotecas de ADNc a partir de semillas de plantas.

Las semillas han sido caracterizadas como órganos de almacenamiento compuestos de embrión y endospermo revestidos por una cubierta de semilla dura. Este último, el endospermo, representa la mayor parte de la semilla, contiene una cantidad significativamente alta de almidón, seguido de otros polisacáridos como celulosa, arabinoxilano, β -glucano, fructanos, proteínas, lípidos y metabolitos secundarios (163), que al oxidarse se unen de forma irreversible a los ácidos nucleicos y pueden llegar a co-precipitar con el material genético generando ARN contaminado para la síntesis de ADNc (164). Por lo tanto, se han desarrollado diferentes metodologías para la extracción de ARN a partir de varios tejidos vegetales, por lo que es de gran importancia seleccionar el método de extracción de ARN a partir de las semillas de *S. bogotensis*.

Se trabajaron dos métodos para la extracción de ARN; CTAB (bromuro de cetiltrimetilamonio) y Trizol. El primero, permite la extracción de ARN en tejidos que contienen altos niveles de polisacáridos y fenoles (165); se ha demostrado que funciona adecuadamente para la extracción de ARN en diversos tejidos. Sin embargo, el protocolo se ha ido modificando debido a las necesidades y requerimientos asociados a las características individuales del material biológico (cantidad de polifenoles, lípidos, polisacáridos y metabolitos secundarios) (166–168).

Por su parte, el segundo método, Trizol, es una solución de fenol e isotiocianato de guanidinio que solubiliza el material biológico (animal y/o vegetal) y desnaturaliza las proteínas simultáneamente (169). La extracción con estos dos métodos se realizó por duplicado usando diferentes cantidades de semillas verdes; para la extracción con Trizol, se ensayaron dos marcas comerciales (ZYMO RESEARCH y Ambion by life) (Anexo F). Paralelamente, se realizó la extracción de ARN a partir de diferentes cantidades de hojas verdes, con el fin de comparar los resultados y seleccionar el mejor material de partida. Las semillas verdes y las hojas fueron maceradas en nitrógeno líquido por separado, teniendo precaución de mantenerlas a baja temperatura, para prevenir la oxidación que induce la degradación del material genético.

La extracción con hojas verdes fue limitada, no se obtuvo material pulverizado homogéneo, debido a la adhesión de las hojas al mortero, por tanto, las semillas fueron más fáciles de manipular para obtener un macerado homogéneo.

Uno de los pasos más importantes durante el proceso de extracción, fue el lavado del precipitado para obtener el material genético y por consiguiente la realización de pipeteos fuertes, que permitió obtener un pellet color blanco, para tener ARN de buena calidad. Sin embargo, en algunas extracciones se obtuvo pellet de color café indicando oxidación del ARN, que se confirmó con la determinación de la relación de absorbancia 260/280, las cuales fueron menores a 1,85 mostrando contaminación con proteínas, remanentes de fenol o simplemente una baja concentración de ARN ($> 10 \text{ ng}/\mu\text{l}$) (170), tanto para semillas como para hojas en Trizol y CTAB.

De acuerdo con los resultados preliminares se repitió la extracción (5 veces más) con CTAB y Trizol (ZYMO RESEARCH), dado que no se observó diferencia en la extracción con ninguna de las dos marcas. Adicionalmente se ensayaron diferentes cantidades de semillas verdes (Tabla 14). Por el método de CTAB, los pellets presentaron coloración más blanca respecto a los obtenidos con Trizol, que presentaron tonalidad café claro. Al cuantificar por espectrofotómetro (NanoDrop, ThermoFisher) los pellets solubilizados, se obtuvo relaciones de absorbancia (260/280) mayores a 1,85, aunque con CTAB, estuvo entre 2,04 y 2,10.

Al realizar el gel de agarosa para confirmar la integridad del ARN (Figura 53-A y B) se observó dos bandas correspondientes a las subunidades 28s y 18s de ARN ribosomal (flechas azules). Estas bandas, se presentan claras y nítidas para las muestras 1, 2, y 3, extraídas con menor cantidad de material vegetal respecto a la muestra 4. Las bandas provenientes del método de CTAB se pueden observar mejor resueltas en el gel (figura 53-A), debido a una relación 260/280 mayor que las obtenidas por el método de Trizol, en cuyo gel (figura 53-B) las bandas características de ARN ribosomal no se resuelven bien.

Tabla 14. Extracción de RNA total de semillas de *S. bogotensis* (SBo)

Método	CTAB				TRIZOL			
Muestra	1	2	3	4	1	2	3	4
mg	350	420	500	780	137	392	436	144
260/280	2,07	2,04	2,10	2,06	1,87	1,86	1,86	0,47
ng/ μ L	1640,6	1012,0	2842,7	4126,7	2263,0	4065,0	837,1	1,0

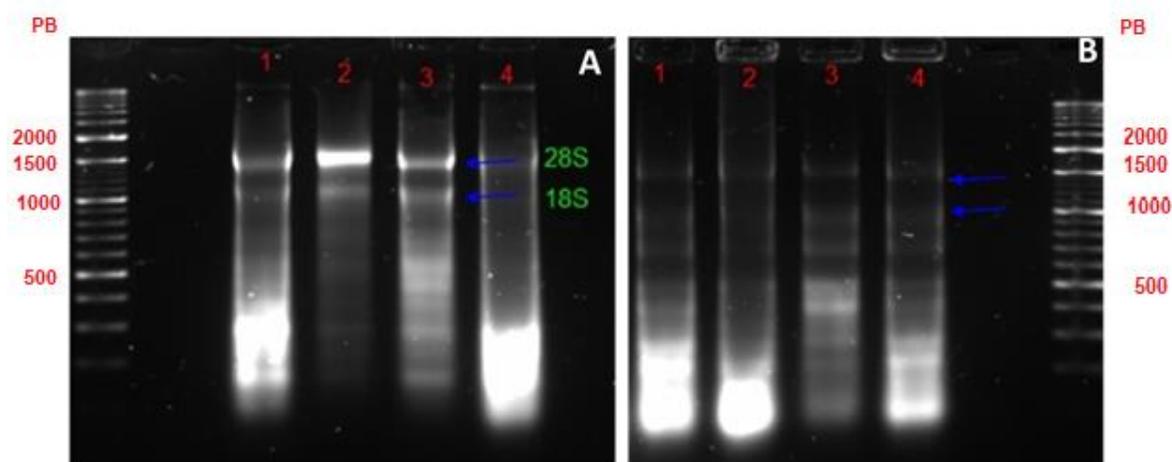


Figura 53. Gel de agarosa 2 %, Tnción con: SYBRr® Safe DNA Gel Stain (Thermo Scientific). Extracción de RNA total de semillas de *S. bogotensis* a partir de (A) Método de CTAB. Carril 1, 2, 3 y 4 corresponden a las muestras 1, 2, 3 y 4 (tabla 14). (B) Método de Trizol. Carril 1, 2, 3, y 4 corresponden a las muestra 1, 2, 3 y 4 (tabla 14). (PB) marador de pares de bases O'RangeRuler 100bp + 500bp (Thermo Fisher).

La mejor extracción de ARN se obtuvo con CTAB y cantidades de 450-500 mg de semillas; con las muestras 2 y 3 se obtuvo relaciones de absorbancia 260/280, mayores o iguales (\geq) a 2.0, sumado a lo anterior, por electroforesis se observó las subunidades del ARN ribosomal 28S y 18S (Tabla 14, Figura 53), las cuales deben encontrarse en una relación (2:1) respectivamente, indicando la calidad del material genético; sin embargo, este requisito puede ser excesivo, por lo que también se ha

demostrado que extracciones con relación de ARNr 28S/18S más baja pueden ser bastante adecuadas para la mayoría de las técnicas de determinación de la expresión génica (171). Por otro lado, se logró determinar que al usar cantidades muy grandes de semillas se generan inconvenientes, ya que se suelen aglomerar en el fondo del tubo cónico impidiendo que los diferentes buffers interactúen con las semillas maceradas provocando su rápida oxidación.

Con el método de Trizol, además de la baja relación de absorbancias, no se observó bandas de RNA. Algunos problemas asociados con esta metodología sugieren que el isotiocianato de guanidina induce la coprecipitación de almidón/polisacáridos dadas las similitudes estructurales con la molécula de ARN. El ARN queda atrapado en los polisacáridos y es arrastrado durante la centrifugación por coprecipitación, de modo que puede ser descartado durante la separación de fases, conduciendo al bajo rendimiento. (163,172). Las semillas maduras de *S. bogotensis*, presentan alto contenido de mucilago (polisacáridos) (146,173,174); sin embargo, esta caracterización no se ha determinado en semillas verdes. Semillas de otras especies de la familia Lamiacea, como *Melissa officinalis*, posee altos niveles de polifenoles y polisacáridos en semillas verdes, que interfieren en la extracción de material genético, por lo tanto, el método de CTAB se estableció como el mejor método de extracción (175).

La mayoría de las lectinas de plantas han sido aisladas a partir de semillas, especialmente de Leguminosas dicotiledóneas, debido a que éstas se almacenan durante el proceso de maduración; constituyen el 10 % del total de la proteína que contiene la semilla; sin embargo, se encuentran en tejidos vegetales, como tallo, raíz y hojas, pero su abundancia es mucho menor y está regulada durante el proceso de desarrollo de la planta (162), de aquí que estos sean los principales órganos seleccionados para la extracción de ARN en plantas para el estudio de genes de estas lectinas.

Para la obtención de lectinas de Lamiácea recombinantes, la extracción de ARN total se realizó de las hojas de *Glechoma hederácea* (*Gleheda*) (3) y en raíces, tallos y hojas para la *S. miltiorrhiza* (2), teniendo en cuenta reportes previos de alta expresión del gen de lectina en estos órganos (176). Se determinó mayor expresión en hojas, y en ambos casos se empleó el método de CTAB, aunque no se mencionaron las dificultades durante la extracción. Otros estudios, se enfocaron en la obtención del RNAm a partir de semillas de *Lepechinia bullata*, con una concentración de 89,7 ng/ μ L y una relación de absorbancias (260/280) de 1,95 (5), empleando kit comerciales.

En este trabajo se seleccionó el método del CTAB para la extracción de ARN a partir de semillas verdes de *S. bogotensis* y, por lo tanto, seguir con el proceso de obtención de ADNc. Además, se hizo tratamiento con DNasa para eliminar contaminación con ADN genómico. Luego de realizar la retro-transcripción y obtener el ADNc, se realizó la cuantificación mediante método de fluorescencia empleando el kit comercial Qubit (Invitrogen), que emplea fluorocromos específicos para diferenciar ADNc de RNA. Una ventaja de este método de cuantificación es su sensibilidad (10 pg/μl) con relación a las medidas por absorbancia (UV), además de eliminar las interferencias que se presentan en la región de 260 nm (177). De esta manera se confirmó la presencia de ADNc con una concentración de 55 μg/mL (muestra 1) y 63,4 μg/mL (muestra 2), lo que permitió iniciar el proceso de amplificación de fragmentos con primers específicos.

6.2.2. Diseño de primers

El uso de primers para la amplificación de ADN mediante reacción en cadena de la polimerasa (PCR), ha sido una técnica ampliamente utilizada en diversos campos de la investigación y si bien el resultado de la PCR puede verse afectado por diferentes factores como la obtención del ADN molde y las condiciones de reacción, el diseño de cebadores es muy importante; ya que deben tener alta especificidad para amplificar una proteína (178). En estudios de amplificación para la obtención de la lectina recombinante de *Lepechina bullata*, el diseño de primers forward degenerados se hizo con base en la secuencia amino terminal de la LSBo-I, dado que ambas son Lamiáceas y un primer reverse universal AUAP (Abridged Universal Amplification Primer), no obstante, los resultados mostraron la amplificación de globulinas o proteínas diferentes a lectinas (5).

6.2.2.1. Traducción Inversa

Para el diseño de los primers, se realizó la traducción inversa (aminoácidos → nucleótidos) de la secuencia parcial de la proteína LSBo-I (5) a la secuencia de nucleótidos (nt), para este fin se usaron los programas bioinformáticos, Sequence Manipulation Suite (http://bioinf.ibun.unal.edu.co/servicios/sms/rev_trans.html) (179) y EMBOSS Backtranseq https://www.bioinformatics.org/sms2/rev_trans.html (179). El último tiene dos herramientas diferentes, una para nucleótidos estándar y otra para nucleótidos ambiguos.

Con el portal Sequence Manipulation Suite, se introdujo la secuencia parcial de la LSBo-I (5) como entrada y se usó la tabla de codones para *S. miltiorrhiza* (Anexo

G) reportada en <http://www.kazusa.or.jp/codon/>, teniendo en cuenta que las dos especies son del mismo género para obtener la secuencia de ADN no degenerada más probable (Anexo H-1). Además de una secuencia ambigua generada a partir de todos los codones que codifican para cada aminoácido (Anexo H-2). Para el portal EMBOSS Backtranseq, se seleccionaron las tablas de codones para *Arabidopsis thaliana* y *Oryza sativa*, especies más cercanas a *S. bogotensis* seleccionadas por el mismo servidor (Anexo H-3), igualmente, se usó el algoritmo Backtranambig que traduce secuencias de proteínas a secuencias de nucleótidos ambiguas, empleando la tabla estándar de codones (Anexo H-4).

La alineación de las cinco secuencias obtenidas se llevó a cabo con Clustal Omega (1.2.4) (Figura 54) para tener una secuencia consenso, que se usó para el diseño de los primers. La secuencia consta de 822 pares de bases y la letra (N) representa los ácidos nucleicos faltantes debido a que en la secuencia proteica consenso aún faltan aminoácidos por determinar.

RT ambigua	ATGGCNACNYTNYTNCARAAYTNYTNACNGCNACNWSNMGNNNNNNNACNYTNYTNGAY	60
Estándar	ATGGCNACNYTNYTNCARAAYTNYTNACNGCNACNWSNMGNNNNNNNACNYTNYTNGAY	60
<i>A. thaliana</i>	ATGGCTACTCTTCTTCAAATCTTCTTACTGCTACTTCTAGANNNNNNACTTCTCTTGAT	60
<i>S. miltiorrhiza</i>	ATGGCTACCCTCCTCCAGAACCTCCTCACCGCTACCAGCAGANNNNNNACCCTCCTCGAT	60
<i>O. sativa</i>	ATGGCCACCCTCCTCCAGAACCTCCTCACCGCCACCTCCCGCNNNNNNACCCTCCTCGAC	60
CONSENSO	ATGGCT TA CCCT CC TCCAG AA CT CC T CA CC CG CT AC CC GA ANNNNNN AC CT CC T CG A T	60
	***** *	
RT ambigua	YTNGAYAAAYACNMGNNNNWSNNNACNYTNYTNACNAYCCNAAYTYAAYGGNGAYYT	120
Estándar	YTNGAYAAAYACNMGNNNNWSNNNACNYTNYTNACNAYCCNAAYTYAAYGGNGAYYT	120
<i>A. thaliana</i>	CTTGATAATACTAGANNNTCTNNNACTTCTTACTCATCCTAATTTAATGGAGATCTT	120
<i>S. miltiorrhiza</i>	CTCGATAACACCAGANNNAGCNNNACCCTCCACCCACCCGAACCTCAACGGCGATCTC	120
<i>O. sativa</i>	CTCGACAACACCCGCNNNTCCNNNACCCTCCTCACCCACCCGAACCTCAACGGCGACCTC	120
CONSENSO	CT CG A T AA CA CC GA AN NT CC NN NA CC CT CC T CA CC CA CC GA AC CT CA AC GG CG A T CT CC	120
	* *	
RT ambigua	GGNCAYGGNCARCAYGGNWSNGGNWSNGGNCARWSNCCNWSNCCNWSNMGNTTYNNNGCN	180
Estándar	GGNCAYGGNCARCAYGGNWSNGGNWSNGGNCARWSNCCNWSNCCNWSNMGNTTYNNNGCN	180
<i>A. thaliana</i>	GGACATGGACAACATGGATCTGGATCTGGACAATCTCCTTCTCCTTAGATTTNNNGCT	180
<i>S. miltiorrhiza</i>	CTCACCGCGCTCTCGCTCAGAAGTGGGCACCGGTGCACCCGAGCCCGAGCAGATTCNNGCT	180
<i>O. sativa</i>	GGCCACGGCCAGCACGGCTCCGGCTCCGGCCAGTCCCGTCCCGTCCCGCTTCNNGGCC	180
CONSENSO	GG CC CA GG CC AG CA CG G CT CC GG CT GC GG CC AG T CC CG TC CC CG T CC CG CT CC CG CT CC CG CT TC NN NG CT	180
	** *	
RT ambigua	YTNACGGNGCNYTNGCNCARAARGTNGGNACNWSNGTNCAYCCNAAYCCNWSNWSNAAY	240
Estándar	YTNACGGNGCNYTNGCNCARAARGTNGGNACNWSNGTNCAYCCNAAYCCNWSNWSNAAY	240
<i>A. thaliana</i>	CTTACTGGAGCTCTTGCTCAAAGGTTGGAACCTTCTGTTTCATCCTAATCCTTCTTCTAAT	240
<i>S. miltiorrhiza</i>	CTCACCGCGCTCTCGCTCAGAAGTGGGCACCGGTGCACCCGAGCCCGAGCAGCAAC	240
<i>O. sativa</i>	CTCACCGCGCCCTCGCCAGAAGTGGGCACCTCCGTGCACCCGAGCCCGTCTCCAAC	240
CONSENSO	CT CA CC GG CG CT CT CG CT CA GA AG T GG GC AC CG GT GC AC CC GA CC CG AG CC CG AG CA CA AC	240
	* *	
RT ambigua	YTNGARGGNGGNGAYGCNTTYGARTTYGTNGGNCNAARNNTTYNNNATHACNNNNNNN	300
Estándar	YTNGARGGNGGNGAYGCNTTYGARTTYGTNGGNCNAARNNTTYNNNATHACNNNNNNN	300
<i>A. thaliana</i>	CTTGAAGGAGGAGATGCTTTTGAATTTGTTGGACCTAAGNNNTTTNNNATTACTNNNNNN	300
<i>S. miltiorrhiza</i>	CTCAGGGCGCGATGCTTTTCGAGTTCTGTTGGGCCCGAAGNNNTTCNNNATCACCNNNNN	300
<i>O. sativa</i>	CTCAGGGCGCGACGCTTCGAGTTCTGTTGGGCCCGAAGNNNTTCNNNATCACCNNNNN	300
CONSENSO	CT CA GG GG CG CG AT G CT TT TC GA GT T CT G TT GG GC CC GA AG NN NT TC NN NAT CA C C NN NN NN	300
	* *	
RT ambigua	NNNNNNNNNNNNNNNGCNGAYGGNNNNGCNNNTTYNNNNNNCCNGTNGGNNNNNNNNN	360
Standard	NNNNNNNNNNNNNNNGCNGAYGGNNNNGCNNNTTYNNNNNNCCNGTNGGNNNNNNNNN	360

<i>A. thaliana</i>	NNNNNNNNNNNNNGCTGATGGANNNGCTNNNTTNNNNNNCCTGTGGANNNNNNNN	360
<i>S. miltiorrhiza</i>	NNNNNNNNNNNNNGCTGATGGCANNNGCTNNNTTNNNNNNCCGGTGGGCNNNNNNNN	360
<i>O. sativa</i>	NNNNNNNNNNNNNGCCGACGGCANNNGCANNNTTNNNNNNCCGGTGGGCNNNNNNNN	360
CONSENSO	NNNNNNNNNNNNNGCTGATGGCANNNGCTNNNTTNNNNNNCCGGTGGGCNNNNNNNN ***** ** *	360
RT ambigua	CCNWSNGGNWSNACNGGNGGNTAYYTNGGNGCNGTNAAYCCNAARGGNNGGNNNNNN	420
Estándar	CCNWSNGGNWSNACNGGNGGNTAYYTNGGNGCNGTNAAYCCNAARGGNNGGNNNNNN	420
<i>A. thaliana</i>	CCTTCTGGATCTACTGGAGGATATCTTGGAGCTGTTAATCCTAAGGGANNNGANNNNNN	420
<i>S. miltiorrhiza</i>	CCGAGCGGCAGCACCGGCGGTACCTCGGCGCTGTGAACCCGAAGGGCANNNGCANNNNNN	420
<i>O. sativa</i>	CCGTCCGGCTCCACCGGCGGTACCTCGGCGCTGTGAACCCGAAGGGCANNNGCANNNNNN	420
CONSENSO	CCGTCCGGCTCCACCGGCGGTACCTCGGCGCTGTGAACCCGAAGGGCANNNGCANNNNNN *	420
RT ambigua	NNNCARGGNGTNGAYGCNGAYYTNAAYGNNNNNGNWSNTAYGNNWSNGGNGGNGGN	480
Estándar	NNNCARGGNGTNGAYGCNGAYYTNAAYGNNNNNGNWSNTAYGNNWSNGGNGGNGGN	480
<i>A. thaliana</i>	NNNCAAGGAGTGTGATGCTGATCTTAATGGANNNGGATCTTATGGATCTGGAGGAGGAGGA	480
<i>S. miltiorrhiza</i>	NNNCAGGGCGTGGATGCTGATCTCAACGGCANNNGCAGCTACGGCAGCGGGCGGGCGGC	480
Oryza	NNNCAGGGCGTGGACGCCGACTCAACGGCANNNGCCTTACGGCTCCGGCGGCGGCGGC	480
CONSENSO	NNNCAAGGAGTGTGATGCTTCAACGGCANNNGCCTTACGGCTCCGGCGGCGGCGGC ***** ** *	480
RT ambigua	AAYCAYGGNTTYTGYGNNWSNGNWSNWSNWSNWSNWSNWSNWSNWSNWSNWSNWSN	540
Estándar	AAYCAYGGNTTYTGYGNNWSNGNWSNWSNWSNWSNWSNWSNWSNWSNWSNWSNWSN	540
<i>A. thaliana</i>	AATCATGGATTTGTGGATCTGGATCTTCTTCTAATCTGAACTTGTCAATCTTCTGGA	540
<i>S. miltiorrhiza</i>	AACCACGGCTTCTGCGCAGCGGCAGCAGCAGCAACAGCGAGCTCGTGACAGCAGCGGC	540
<i>O. sativa</i>	AACCACGGCTTCTGCGCTCCGGCTCCCTCCAACCTCCGAGCTCGTGACGCTCCCGGC	540
CONSENSO	AACCAAGGCTTCTGCGCTCCGGCTCCCTCCAACCTCCGAGCTCGTGACGCTCCCGGC *	540
RT ambigua	YTNGGNCAYCARGARWSNMGNNAHAAYTAYNNNNNNNNNNNNNTAYGARGARYTN	600
Estándar	YTNGGNCAYCARGARWSNMGNNAHAAYTAYNNNNNNNNNNNNNTAYGARGARYTN	600
<i>A. thaliana</i>	CTTGGACATCAAGAATCTAGANNNTAATTAATNNNNNNNNNNNNNTATGAAGAACTT	600
<i>S. miltiorrhiza</i>	CTCGGCCACCAGGAGAGCAGANNNTCAACTACNNNNNNNNNNNNNTACGAGGAGCTC	600
<i>O. sativa</i>	CTCGGCCACCAGGAGTCCCGCANNNTCAACTACNNNNNNNNNNNNNTACGAGGAGCTC	600
CONSENSO	CTCGGCCACCAGGAGTCCCGCANNNTCAACTACNNNNNNNNNNNNNTACGAGGAGCTC *	600
RT ambigua	CARGTNACNGCNGGNNMGNNAARNNNNNNNNNNNWSNTAYGARGCNGARYTNGAYGAY	660
Estándar	CARGTNACNGCNGGNNMGNNAARNNNNNNNNNNNWSNTAYGARGCNGARYTNGAYGAY	660
<i>A. thaliana</i>	CAAGTTACTGCTGGANNAGAAAGNNNNNNNNNNNTCTTATGAAGCTGAAGCTGATGAT	660
<i>S. miltiorrhiza</i>	CAGGTGACCGCTGGCANNAGAAAGNNNNNNNNNNNAGCTACGAGGCTGAGCTCGATGAT	660
<i>O. sativa</i>	CAGGTGACCGCGCANNCGCAAGNNNNNNNNNNNTCCTACGAGGCGGAGCTCGACGAC	660
CONSENSO	CAGGTGACCGCTGGCANNAGAAAGNNNNNNNNNNNACCTACGAGGCTGAGCTCGATGAT *	660
RT ambigua	TTYGTNACNYTNCARGTNACNGTNGGNYTNWSNGCNGCNCNWSNGGNGAYGGNGTNGAR	720
Estándar	TTYGTNACNYTNCARGTNACNGTNGGNYTNWSNGCNGCNCNWSNGGNGAYGGNGTNGAR	720
<i>A. thaliana</i>	TTTGTACTCTTCAAGTTACTGTTGGACTTTCTGCTGCTACTTCTGGAGATGGAGTTGAA	720
<i>S. miltiorrhiza</i>	TTCGTGACCCCTCCAGGTGACCGTGGGCTCAGCGCTGCTACAGCGGCGATGGCGTGGAG	720
Oryza	TTCGTGACCCCTCCAGGTGACCGTGGGCTCCTCCGCCACCTCCGGCGACGGCGTGGAG	720
CONSENSO	TTCGTGACCCCTCCAGGTGACCGTGGGCTCCTCCGCCACCTCCGGCGACGGCGTGGAG *	720
RT ambigua	AARGAYATHNNNNNTGGNNNTTYACNGCANNATGNNNNNNNNNNNNNNNNNNNNNNNN	780
Estándar	AARGAYATHNNNNNTGGNNNTTYACNGCANNATGNNNNNNNNNNNNNNNNNNNNNNNN	780
<i>A. thaliana</i>	AAGGATATTNNNNNTGGNNNTTACTGCTNNATGNNNNNNNNNNNNNNNNNNNNNNNN	780
<i>S. miltiorrhiza</i>	AAGGATATCANNNNNTGGNNNTTACCGCTNNATGNNNNNNNNNNNNNNNNNNNNNNNN	780
<i>O. sativa</i>	AAGGACATCANNNNNTGGNNNTTACCGCANNATGNNNNNNNNNNNNNNNNNNNNNNNN	780
CONSENSO	AAGGATATTNNNNNTGGNNNTTACCGCTNNATGNNNNNNNNNNNNNNNNNNNNNNNN *	780
RT ambigua	NNNNNNNNNNNNNNNGGNNNNNNNATHMGNCARNNNGTN	822
Estándar	NNNNNNNNNNNNNNNGGNNNNNNNATHMGNCARNNNGTN	822
<i>A. thaliana</i>	NNNNNNNNNNNNNNNGGANNNNNNATTAGACAANNNGTT	822
<i>S. miltiorrhiza</i>	NNNNNNNNNNNNNNNGGCNNNNNNATCAGACAGNNNGTG	822
<i>O. sativa</i>	NNNNNNNNNNNNNNNGGCNNNNNNATCCGCCAGNNNGTG	822
CONSENSO	NNNNNNNNNNNNNNNGGCNNNNNNATCAGACAGNNNGTG ***** *	822

Figura 54. Alineamiento múltiple de las 5 secuencias obtenidas de traducción Inversa (Anexo H), RT con codones ambiguos, RT estándar, RT con codones de *A. thaliana*, RT con codones de *S. miltiorrhiza* y RT con codones de *O. sativa*. En asterisco (*) se indican los nucleótidos (nt) conservados en las 5 secuencias. (N) Indica nt no asignados, En la última fila se presenta la secuencia (CONSENSO) obtenida a partir del alineamiento.

10	20	30	40	50	60
ATGGCTACCC	TCCTCCAGAA	CCTCCTCACC	GCTACCTCCA	GANNNNNNAC	CCTCCTCGAT
70	80	90	100	110	120
CTCGATAACA	CCAGANNNTC	CNNNACCCTC	CTCACCCACC	CGAACTTCAA	CGGCGATCTC
130	140	150	160	170	180
GGCCACGGCC	AGCACGGCTC	CGGCTGCGGC	CAGTCCCCGT	CCCCGTCCAG	ATTCNNGGCT
190	200	210	220	230	240
CTCACCGGCG	CTCTCGCTCA	GAAGGTGGGC	ACCTCCGTGC	ACCCGAACCC	GTCCTCCAAC
250	260	270	280	290	300
CTCGAGGGCG	GCGATGCTTT	CGAGTTCGTG	GGCCCGAAGN	NNTTCNNNAT	CACCNNNNNN
310	320	330	340	350	360
NNNNNNNNNN	NNNNNGCTGA	TGGCNNNGCT	NNNTTCNNNN	NNCCGGTGGG	CNNNNNNNNN
370	380	390	400	410	420
CCGTCCGGCT	CCACCGGCGG	CTACCTCGGC	GCTGTGAACC	CGAAGGGCNN	NGGCNNNNNN
430	440	450	460	470	480
NNNCAGGGCG	TGGATGCTGA	TCTCAACGGC	NNNGGCTCCT	ACGGCTCCGG	CGGCGGCGGC
490	500	510	520	530	540
AACCACGGCT	TCTGCGGCTC	CGGCTCCTCC	TCCAACCTCCG	AGCTCGTGCA	GTCCTCCGGC
550	560	570	580	590	600
CTCGGCCACC	AGGAGTCCAG	ANNNATCAAC	TACNNNNNNN	NNNNNNNNTA	CGAGGAGCTC
610	620	630	640	650	660
CAGGTGACCG	CTGGCNNNAG	AAAGNNNNNN	NNNNNNACCT	ACGAGGCTGA	GCTCGATGAT
670	680	690	700	710	720
TTCGTGACCC	TCCAGGTGAC	CGTGGGCCTC	TCCGCTGCTA	CCTCCGGCGA	TGGCGTGGAG
730	740	750	760	770	780
AAGGATATCN	NNNNNTGGNN	NTTACCGCT	NNNATGNNNN	NNNNNNNNNN	NNNNNNNNNN
790	800	810	820		
NNNNNNNNNN	NNNNGGCNNN	NNNATCAGAC	AGNNNGNNNG	TG	

Figura 55. Secuencia consenso propuesta de 822 pb, para la obtención de cDNA de la LSB0-I.

Se llevó a cabo la traducción de la secuencia consenso obtenida (Figura 55) usando el servidor disponible en <https://web.expasy.org/translate/> y se realizó el alineamiento con la secuencia proteica parcial de LSB0-I (15), encontrándose diferencia en solo dos aminoácidos (aa) (Figura 56, flecha roja); Serina (Ser) por Cisteína (Cys) en la posición 49. Ambos aminoácidos son polares sin carga, el otro cambio fue Serina (Ser) por Treonina (Thr) en la posición 213, aminoácidos polares sin carga. Con estos resultados se hizo el diseño de primers que abarcara la secuencia completa



Figura 56. Alineamiento de las secuencias traducida obtenida de la secuencia consenso (Traslate), con la secuencia parcial de LSB0-I (5) (Parcial). Las flechas indica los dos aminoácidos que no se conservan.

6.2.2.2. Diseño de primers

El diseño de los primers se hizo con la herramienta bioinformática PrimerBlast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>), los resultados arrojaron diez (10) pares de primers (Figura 57) (Anexo I), de los cuales se seleccionaron seis (6) teniendo en cuenta el tamaño de los primers, (20 o 21 pb), temperatura de fusión (Tm) (50 – 65 °C), el % de CG (45 - 55 %), con máximo en 60 % (180). En la tabla 15, se resumen las secuencias y las características de los primers. Además, se sintetizaron tres pares adicionales reportados para la expresión del gen de la lectina de *S. miltiorrhiza* (2). El primer par de estos, (SmF1 y SmR1) fue diseñado de acuerdo con la secuencia unigénita (Contig1927) de la biblioteca de ADNc de *S. miltiorrhiza* y se usó para ampliar el gen que codifica para la lectina de *S. miltiorrhiza* (SmL1), el otro par de primers (SmF2 y SmR2) se usó para determinar niveles de expresión de SmL1 en diferentes tejidos (raíz tallos y hojas) usando el marcador SYBR Green II y por último, se sintetizó un par de primers correspondientes a un gen housekeeping, gliceraldehido-3-fostato deshidrogenasa, (SmGAPDH) (CV170251), que fue usado como control (2).

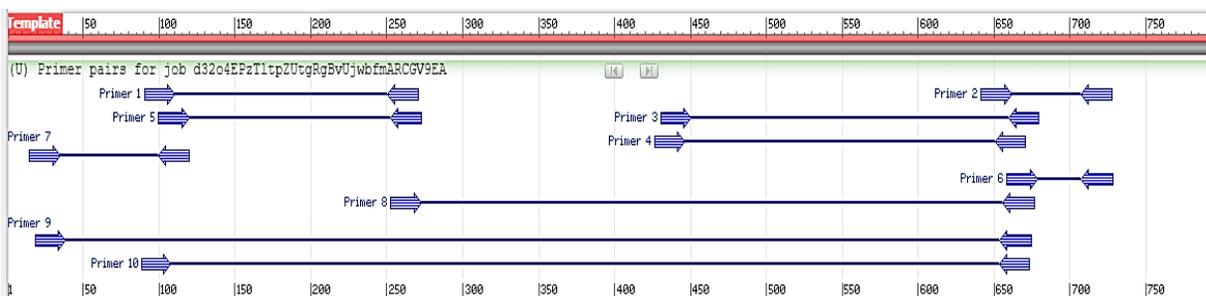


Figura 57. Conjunto de Primers diseñados para la obtención del gen de la LSBo-I.

Tabla 15. Primers seleccionados para obtención del gen de LSBo-I

Nombre*	Secuencia Primer (5'→3')	Empieza	Termina	Tm	%CG	Producto	Nombre PCR
PBSb F1	CTCACCCACCCGAAC TTCAA	91	110	59,89	55	180	A
PBSb R1	CACGAACTCGAAAGCATCGC	270	251	60,25	55		
PBSb F4	GGCGTGGATGCTGATCTCAA	427	446	60,46	55	244	B
PBSb R4	GGGTCACGAAATCATCGAGC	670	551	60,76	55		
PBSb F6	ATTCGTGACCCTCCAGGTG	659	678	58,68	55	70	C
PBSb R6	ATATCCTTCTCCACGCCATCG	728	708	59,73	52		
PBSb F7	CCAGAACCTCCTCACCGCTA	15	34	60,97	55	105	D
PBSb R7	AGATCGCCGTTGAAGTTCGG	119	100	60,74	52		
PBSb F8	GATGCTTTCGAGTTCGTGGG	253	272	58,28	60	424	E
PBSb R8	CCTGGAGGGTCACGAAATCAT	676	656	59,79	55		
PBSb F9	AACCTCCTCACCGCTACCTC	19	38	60,97	60	656	F
PBSb R9	TGGAGGGTCACGAAATCATCG	674	654	60,13	52		
SmF1	ATGGCCAAGCTTCTCCAAAAC	---	---	63	48	886	G
SmR1	GTCGATCGCTTAGTCCTTATTGA	---	---	60	44		
SmF2	CATGACATC GTCTCGTGGTATTTT	---	---	63	46	166	H
SmR2	GATCGCTTAGTCCTTATTGATTTGC	---	---	62	40		
SmGAPF	CCACCGTCCACTCCATCACT	---	---	63	60	--	I
SmGAPR	TG GGA ACTCGGAACGACATAC	---	---	62	52		

*PB (primer Blast), SB (S. bogotensis), F (Forward), R (Reverse), Sm (S. miltiorrhiza) GAP (Gen housekeeping).

6.2.3. PCR

La amplificación de los genes se realizó mediante PCR convencional, empleando los primers (Tabla 15) y modificando diferentes condiciones para la PCR teniendo en cuenta su temperatura de fusión (Tm). Las condiciones variaron de la siguiente manera; desnaturalización inicial: 95 °C por 5 a 10 min, desnaturalización 30 a 35 ciclos a 95 °C por 30s, alineamiento entre 50 °C-70 °C por 30s y extensión a 72 °C por 60s. Extensión final a 72 C por 10 min. Adicionalmente, se ensayaron las condiciones reportadas para la lectina de *S. miltiorrhiza* (SML1) (7), las cuales incluyeron, desnaturalización inicial: 94 °C por 4 min, desnaturalización 30 ciclos a

94 °C por 30s, alineamiento a 72 °C por 80s y extensión final a 72 °C por 10 min) y (desnaturalización inicial: 95 °C por 5 min, desnaturalización 40 ciclos a 95 °C por 10s, alineamiento a 59 °C por 15s y extensión final a 72 °C por 10 min). Otros ensayos que se incluyeron fue la modificación de la concentración de primers de 10 pM a 100 pM y del ADN (entre 1 y 2 µL). La actividad de las enzimas, RevertAid RT y Taq polimerasa, se corroboró por medio de los ensayos de expresión del control positivo (C+) ARN GAPDH y los primers (primer GAPDH forward y primer GAPDH reverse), del kit de síntesis de ADNc (ReverAid First Strand cDNA Synthesis Kit, Thermo Scientific), estos amplificaron 496 pb (Figura 53, Carril C+). A pesar de las diferentes modificaciones realizadas en los ensayos, no fue posible obtener ningún amplificado (Figura 58). Por lo tanto, se hizo una validación para todos los primers, con el servidor PCR Primer Stats (https://www.bioinformatics.org/sms2/pcr_primer_stats.html) (180).

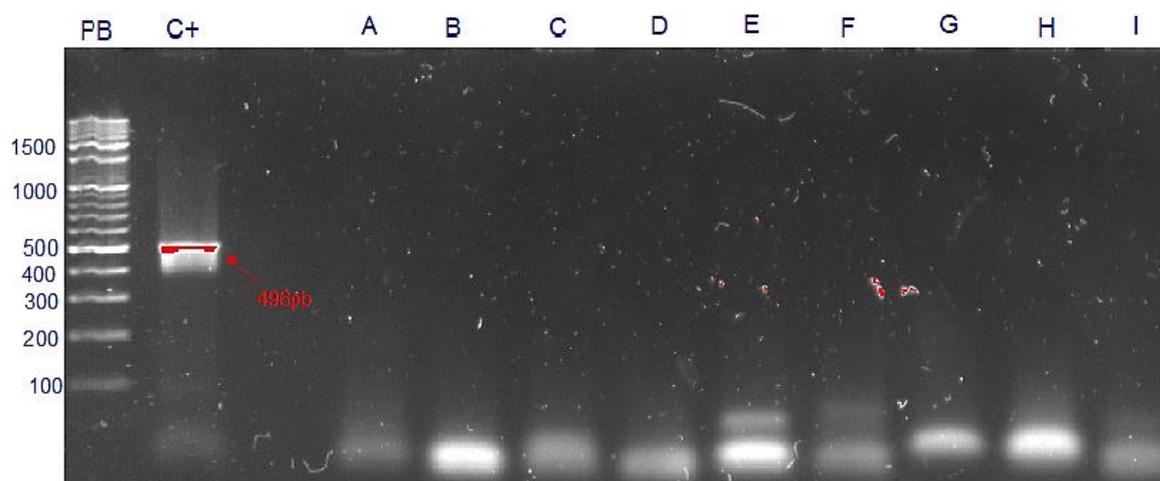


Figura 58. PCR con los primers diseñados. Gel de agarosa 2 %, Tinción con: SYBR® Safe DNA Gel Stain (Thermo Scientific). (PB): marcador de pares de bases, O'RangeRuler 100bp + 500bp (Thermo Fisher), (C+) control positivo Taq polimerasa. Ensayos con primers (A-I) (Tabla 15) Condiciones PCR: Denaturación: 4min (94°C), Amplificación x30Ciclos (30 seg (94 °C), 30 seg (54 °C), 80seg (72 °C) y extensión 10 min (72 °C).

El servidor determinó para todos los primers, que la Tm estaba entre 62 °C y 66 °C (Anexo J); temperaturas más altas que las obtenidas por el programa PrimerBlast. Adicional a esto, para algunos pares de primers señaló otros tipos de advertencias; para el primer PBSb R4 y F8 con más de tres G o C en las últimas 5 bases; para los primers PBSb R6 - R7, SMF1-2 y SmR1 y 2, la presencia de tres bases que se podrían autoanillar. Es claro, que se deben realizar las validaciones del diseño de

primers previa síntesis, dado que no se logró obtener amplificación del gen. Otros estudios de amplificación de lectinas en Lamiácea, como los realizados para la lectina de *L. bullata* (LLB), emplearon el ARNm y primers diseñados a partir de secuencias de lectinas de Fabáceas, aunque, si se obtuvo amplificados, ninguno de los genes que se encontró en el análisis BLAST estuvo relacionado con lectinas, lo que indicó que los primers fueron inespecíficos (5). Para la obtención de la lectina recombinante de *S. miltiorrhiza*, se partió de una biblioteca de genes de ADNc (181), con 30 genes que codifican para lectinas y esto permite diseñar los primers, con base al que se encontraba en mayor abundancia (Conting1927) (2). En este trabajo, se establecieron las condiciones para la extracción de ARN y la obtención de ADNc a partir de semillas de *S. bogotensis* verdes, con el método de CTAB, que permite ser empleado como molde en la síntesis de primera hebra y posteriormente en la amplificación, aunque se debe mejorar el diseño de primers específicos para obtener la secuencia codificante de LSBo-I.

7. Conclusiones y Recomendaciones

7.1 Conclusiones

- Se determinó la secuencia parcial de aminoácidos de la lectina de *Salvia bogotensis* (LSBo-I), proponiéndose 278 residuos, de los cuales 248 aminoácidos fueron ubicados o asignados en los alineamientos; con una cobertura del 71 % del total de los 347 residuos reportados previamente para LSBo-I (4).

```
      10      20      30      40      50      60
MATLLQNLIT ATSMELSRDL QDNQRXSXTL LTHPNFNGDL GHGQHGSGSG QSPSPSRFXA
      70      80      90     100     110     120
LTGALAQKVG TSVHPNPSSN LEGGDAFEFV GPKXFXITXX XXXXXADGXA XFXXPVGXXX
      130     140     150     160     170     180
PSGSTGGYLQ AVNPKGXGXQ XQGVADALNG XGSYSGGGGG NHGFCGSGSS SENSELVQSSG
      190     200     210     220     230     240
LGHQESRXID YEAGSNRLEV RVTAGXSKQG INEYASEVRD FVTLQVTIGF SAATGDWVEV
      250     260     270
HNILSWSFNS SLEINNGKNX XXLMAIGSGI RLVVEPSK
```

- El modelo generado para LSBo-I mediante Phyre², mostró una estructura conformada por una hoja β posterior que consta de seis (6) hebras antiparalelas, una hoja β frontal curva con 6 hebras. Estas hojas forman una estructura aplanada en forma de domo con bucles localizados en la parte superior del monómero formando el dominio de reconocimiento de carbohidratos (CDR) plegamiento característico del tipo leguminosa (L).
- El acople molecular con LSBo-I propone un sitio de unión a carbohidrato en el que están presentes los residuos de Asn126 y Ser146, presentes en la unión con los tres ligandos (T, Tn y GalNAC), también Asn137, Tyr130, Ser145 comunes en la unión con los ligandos T y GalNac; y, por último, Gly108, Asp95, comunes para la unión con T y Tn.
- El presente estudio confirmó la presencia de la lectina de *S. bogotensis* que reconoce manosa (LSBo-II), con peso molecular de 30 kDa, similar a los reportados para lectinas con afinidad por glucosa/manosa pertenecientes a Leguminosas.
- La separación de las dos lectinas se realizó por cromatografía de intercambio iónico sobre DEAE y afinidad por Sepharosa 4B Concanavalina A.

- Se estableció que CTAB fue la mejor alternativa para obtener un ARN de la calidad necesaria para realizar la síntesis de primera hebra; sin embargo, es importante el diseño de primers específicos para obtener el gen de la lectina, y de esta forma obtener la recombinante.

7.2 Recomendaciones

- Hacer digestión de la proteína purificada por cromatografía de afinidad con el anticuerpo con dos enzimas de diferente especificidad, para obtener un grupo de péptidos que permitan realizar superposición y alcanzar el 100 % de la secuencia de la LSBo-I.
- Realizar estudios de solubilidad que permitan cristalizar la proteína y obtener su estructura tridimensional.
- Realizar más estudios enfocados en obtener la lectina recombinante y de esta forma continuar los ensayos de actividad biológica para la detección de los glicotopes T y Tn, citotoxicidad y actividad insecticida entre otras

Referencias

1. Pérez G, Vega N. Lamiaceae lectins. *Funct Plant Sci Biotechnol.* 2007;1:288–99.
2. Hua W, Han L, Wang Z. Molecular cloning and expression of a novel gene related to legume lectin from *Salvia miltiorrhiza* Bunge. *African J Biotechnol.* 2015;14(28):2234–43.
3. Wang W, Peumans WJ, Rougé P, Rossi C, Proost P, Chen J, et al. Leaves of the Lamiaceae species *Glechoma hederacea* (ground ivy) contain a lectin that is structurally and evolutionary related to the legume lectins. *Plant J.* 2003;33(2):293–304.
4. Vega N, Pérez G. Isolation and characterisation of a *Salvia bogotensis* seed lectin specific for the Tn antigen. *Phytochemistry.* 2006;67(4):347–55.
5. Wilches Torres MA. Aproximación a la estructura primaria de lectinas específicas para el antígeno Tn e identificación de nuevas lectinas específicas para glucosa/manosa en semillas de *Salvia bogotensis* y *Lepechinia bullata*. Universidad Nacional de Colombia-Sede Bogotá;
6. da Costa V, van Vliet SJ, Carasi P, Frigerio S, García PA, Croci DO, et al. The Tn antigen promotes lung tumor growth by fostering immunosuppression and angiogenesis via interaction with Macrophage Galactose-type lectin 2 (MGL2). *Cancer Lett.* 2021;
7. Gabba A, Bogucka A, Luz JG, Diniz A, Coelho H, Corzana F, et al. Crystal Structure of the Carbohydrate Recognition Domain of the Human Macrophage Galactose C-Type Lectin Bound to GalNAc and the Tumor-Associated Tn Antigen. *Biochemistry.* 2021;60(17):1327–36.
8. Simplicien M, Barre A, Benkerrou Y, Van Damme EJM, Rougé P, Benoist H. The T/Tn-Specific Helix pomatia Lectin Induces Cell Death in Lymphoma Cells Negative for T/Tn Antigens. *Cancers (Basel).* 2021;13(17):4356.
9. Oinam SD, Senjam SS, Kamei R, Hanjabam JS. The Role of Lectin as Potential Biomarker in Ovarian Cancer. *Curr Pharm Biotechnol.* 2022;23(4):478–85.
10. Hirabayashi J, Tateno H, Shikanai T, Aoki-Kinoshita KF, Narimatsu H. The Lectin Frontier Database (LfDB), and data generation based on frontal affinity chromatography. *Molecules.* 2015;20(1):951–73.
11. Sousa BL, Silva Filho JC, Kumar P, Pereira RI, Łyskowski A, Rocha BAM, et al. High-resolution structure of a new Tn antigen-binding lectin from *Vatairea macrocarpa* and a comparative analysis of Tn-binding legume lectins. *Int J Biochem Cell Biol.* 2015;59:103–10.
12. Wu AM. Lectinochemical studies on the glyco-recognition factors of a Tn (GalNAc α 1 \rightarrow Ser/Thr) specific lectin isolated from the seeds of *Salvia sclarea*. *J Biomed Sci.* 2005;12(1):167–84.
13. Piller V, Piller F, Cartron J-P. Isolation and characterization of an N-acetylgalactosamine specific lectin from *Salvia sclarea* seeds. *J Biol Chem.* 1986;261(30):14069–75.
14. Medeiros A, Bianchi S, Calvete JJ, Balter H, Bay S, Robles A, et al.

- Biochemical and functional characterization of the Tn-specific lectin from *Salvia sclarea* seeds. *FEBS J.* 2000;267(5):1434–40.
15. Lis H, Sharon N. Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem Rev.* 1998;98(2):637–74.
 16. Duk M, Mitra D, Lisowska E, Kabat EA, Sharon N, Lis H. Immunochemical studies on the combining site of the A+ N blood type specific *Moluccella laevis* lectin. *Carbohydr Res.* 1992;236:245–58.
 17. Duarte JJF, Pérez G. Tn-specific lectins production from *Salvia palifolia* and *Hyptis mutabilis* by cellular somaclonal variation. *Rev Fac Ciencias Básicas.* 2013;9(1):134–41.
 18. Manning JC, Romero A, Habermann FA, Caballero GG, Kaltner H, Gabius H-J. Lectins: a primer for histochemists and cell biologists. *Histochem Cell Biol.* 2017;147(2):199–222.
 19. Lam SK, Ng TB. Lectins: production and practical applications. *Appl Microbiol Biotechnol.* 2011;89(1):45–55.
 20. Dan X, Liu W, Ng TB. Development and applications of lectins as biological tools in biomedical research. *Med Res Rev.* 2016;36(2):221–47.
 21. Rüdiger H, Gabius HJ. The history of lectinology. *Sugar Code Fundam Glycosci.* 2009;261–8.
 22. Fitches E, Wiles D, Douglas AE, Hinchliffe G, Audsley N, Gatehouse JA. The insecticidal activity of recombinant garlic lectins towards aphids. *Insect Biochem Mol Biol.* 2008;38(10):905–15.
 23. Kaur M, Singh K, Rup PJ, Kamboj SS, Singh J. Anti-insect potential of lectins from *Arisaema* species towards *Bactrocera cucurbitae*. *J Env Biol.* 2009;30(6):1019–23.
 24. Sattayasai N, Sudmoon R, Nuchadomrong S, Chaveerach A, Kuehnle AR, Mudalige-Jayawickrama RG, et al. *Dendrobium findleyanum* agglutinin: production, localization, anti-fungal activity and gene characterization. *Plant Cell Rep.* 2009;28(8):1243–52.
 25. Yan Q, Jiang Z, Yang S, Deng W, Han L. A novel homodimeric lectin from *Astragalus mongholicus* with antifungal activity. *Arch Biochem Biophys.* 2005;442(1):72–81.
 26. Swanson MD, Winter HC, Goldstein IJ, Markovitz DM. A lectin isolated from bananas is a potent inhibitor of HIV replication. *J Biol Chem.* 2010;285(12):8646–55.
 27. Kobayashi T, Kuroda J, Ashihara E, Oomizu S, Terui Y, Taniyama A, et al. Galectin-9 exhibits anti-myeloma activity through JNK and p38 MAP kinase pathways. *Leukemia.* 2010;24(4):843–50.
 28. Choi SH, Lyu SY, Park WB. Mistletoe lectin induces apoptosis and telomerase inhibition in human A253 cancer cells through dephosphorylation of Akt. *Arch Pharm Res.* 2004;27(1):68–76.
 29. van Leeuwen SS, Schoemaker RJW, Timmer CJAM, Kamerling JP, Dijkhuizen L. Use of *Wisteria floribunda* agglutinin affinity chromatography in the structural analysis of the bovine lactoferrin N-linked glycosylation. *Biochim Biophys Acta (BBA)-General Subj.* 2012;1820(9):1444–55.
 30. Morgan GW, Kail M, Hollinshead M, Vaux DJ. Combined biochemical and cytological analysis of membrane trafficking using lectins. *Anal Biochem.*

- 2013;441(1):21–31.
31. Inoue K, Wada J, Eguchi J, Nakatsuka A, Teshigawara S, Murakami K, et al. Urinary fetuin-A is a novel marker for diabetic nephropathy in type 2 diabetes identified by lectin microarray. *PLoS One*. 2013;8(10):e77118.
 32. Nakajima K, Inomata M, Iha H, Hiratsuka T, Etoh T, Shiraishi N, et al. Establishment of new predictive markers for distant recurrence of colorectal cancer using lectin microarray analysis. *Cancer Med*. 2015;4(2):293–302.
 33. Bertok T, Sediva A, Katrlík J, Gemeiner P, Mikula M, Nosko M, et al. Label-free detection of glycoproteins by the lectin biosensor down to attomolar level using gold nanoparticles. *Talanta*. 2013;108:11–8.
 34. Tsaneva M, Van Damme EJM. 130 years of plant lectin research. *Glycoconj J*. 2020;1–19.
 35. Jain M, Amera GM, Muthukumaran J, Singh AK. Insights into biological role of plant defense proteins: A review. *Biocatal Agric Biotechnol*. 2022;102293.
 36. Nakamura-Tsuruta S, Kominami J, Kuno A, Hirabayashi J. Evidence that *Agaricus bisporus* agglutinin (ABA) has dual sugar-binding specificity. *Biochem Biophys Res Commun*. 2006;347(1):215–20.
 37. Jiang S-Y, Ma Z, Ramachandran S. Evolutionary history and stress regulation of the lectin superfamily in higher plants. *BMC Evol Biol*. 2010;10(1):79.
 38. Cordara G, Egge-Jacobsen W, Johansen HT, Winter HC, Goldstein IJ, Sandvig K, et al. *Marasmius oreades* agglutinin (MOA) is a chimerolectin with proteolytic activity. *Biochem Biophys Res Commun*. 2011;408(3):405–10.
 39. Bonnardel F, Mariethoz J, Salentin S, Robin X, Schroeder M, Perez S, et al. UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. *Nucleic Acids Res*. 2019;47(D1):D1236–44.
 40. Fujimoto Z, Tateno H, Hirabayashi J. Lectin structures: classification based on the 3-D structures. *Lectins*. 2014;579–606.
 41. Gupta RK, Gupta GS. Mannose Receptor Family: R-Type Lectins. In: *Animal Lectins: Form, Function and Clinical Applications*. Springer; 2012. p. 331–47.
 42. Etzler ME, Surolia A, Cummings RD. L-type lectins. *Essentials Glycobiol* 2nd Ed. 2009;
 43. Kilpatrick DC. Animal lectins: a historical introduction and overview. *Biochim Biophys Acta (BBA)-General Subj*. 2002;1572(2–3):187–97.
 44. Loh SH, Park J-Y, Cho EH, Nah S-Y, Kang Y-S. Animal lectins: potential receptors for ginseng polysaccharides. *J Ginseng Res*. 2017;41(1):1–9.
 45. Anderson K, Evers D, Rice KG. Structure and function of mammalian carbohydrate-lectin interactions. *Glycoscience*. 2008;2445.
 46. Matsumoto J, Nakamoto C, Fujiwara S, Yubisui T, Kawamura K. A novel C-type lectin regulating cell growth, cell adhesion and cell differentiation of the multipotent epithelium in budding tunicates. *Development*. 2001;128(17):3339–47.
 47. Walker JR, Nagar B, Young NM, Hiramata T, Rini JM. X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure. *Biochemistry*. 2004;43(13):3783–92.
 48. Isseroff RR, Liu F-T. Galectin-3 regulates intracellular trafficking of epidermal growth factor receptor through Alix and promotes keratinocyte migration. *J*

- Invest Dermatol. 2012;132(12):2828–37.
49. Konno A, Kitagawa A, Watanabe M, Ogawa T, Shirai T. Tracing protein evolution through ancestral structures of fish galectin. *Structure*. 2011;19(5):711–21.
 50. Ruiz FM, Fernández IS, López-Merino L, Lagartera L, Kaltner H, Menéndez M, et al. Fine-tuning of prototype chicken galectins: structure of CG-2 and structure–activity correlations. *Acta Crystallogr Sect D Biol Crystallogr*. 2013;69(9):1665–76.
 51. Sharon N, Lis H. Molecular structure. *Lectins*. 2007;105–74.
 52. Ley K. The role of selectins in inflammation and disease. *Trends Mol Med*. 2003;9(6):263–8.
 53. Selvaraj C, Abhirami R, Vijayakumar R, Alfaiz FA, Singh SK. Immunological insights of selectins in human disease mechanism. *Adv Protein Chem Struct Biol*. 2022;129:163–88.
 54. Martínez Muñoz P. Estudio de las interacciones entre la vicilina y las lectinas con ay Cel-II de la semilla de *Canavalia Ensiformis*. Universidad Nacional de Colombia;
 55. Lima TE, Sartori ALB, Rodrigues MLM. Plant antiherbivore defenses in Fabaceae species of the Chaco. *Brazilian J Biol*. 2017;77(2):299–303.
 56. Wei J, Xu D, Zhou J, Cui H, Yan Y, Ouyang Z, et al. Molecular cloning, characterization and expression analysis of a C-type lectin (Ec-CTL) in orange-spotted grouper, *Epinephelus coioides*. *Fish Shellfish Immunol*. 2010;28(1):178–86.
 57. Van Damme EJM. 35 years in plant lectin research: a journey from basic science to applications in agriculture and medicine. *Glycoconj J*. 2021;1–15.
 58. De Coninck T, Van Damme EJM. The multiple roles of plant lectins. *Plant Sci*. 2021;313:111096.
 59. Corredor ZYC, Montaña EAR, Castro NAV. LEGUME LECTINS DOMAIN: STRUCTURAL CHARACTERISTICS AND INSECTISTATIC AND INSECTICIDAL ACTIVITIES. *Chil J Agric Anim Sci (ex Agro-Ciencia)*. 2016;32(2):157–69.
 60. Gorakshakar AC, Ghosh K. Use of lectins in immunohematology. *Asian J Transfus Sci*. 2016;10(1):12.
 61. Kudelka MR, Ju T, Heimbürg-Molinario J, Cummings RD. Simple sugars to complex disease—mucin-type O-glycans in cancer. In: *Advances in cancer research*. Elsevier; 2015. p. 53–135.
 62. Syed P, Gidwani K, Kekki H, Leivo J, Pettersson K, Lamminmäki U. Role of lectin microarrays in cancer diagnosis. *Proteomics*. 2016;16(8):1257–65.
 63. Akkouh O, Ng TB, Singh SS, Yin C, Dan X, Chan YS, et al. Lectins with anti-HIV activity: a review. *Molecules*. 2015;20(1):648–68.
 64. Santos AFS, da Silva MDC, Napoleão TH, Paiva PMG, Correia MTS, Coelho L. Lectins: Function, structure, biological properties and potential applications. *Curr Top Pept protein Res*. 2014;15:41–62.
 65. Van Holle S, Van Damme EJM. Messages from the past: new insights in plant lectin evolution. *Front Plant Sci*. 2019;10:36.
 66. Tirta Ismaya W, Tjandrawinata RR, Rachmawati H. Lectins from the edible mushroom *Agaricus bisporus* and their therapeutic potentials. *Molecules*.

- 2020;25(10):2368.
67. Carrizo ME, Capaldi S, Perduca M, Irazoqui FJ, Nores GA, Monaco HL. The antineoplastic lectin of the common edible mushroom (*Agaricus bisporus*) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. *J Biol Chem*. 2005;280(11):10614–23.
 68. Asensio JL, Cañada FJ, Siebert H-C, Laynez J, Poveda A, Nieto PM, et al. Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains. *Chem Biol*. 2000;7(7):529–43.
 69. Damme EJM Van, Peumans WJ, Barre A, Rougé P. Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *CRC Crit Rev Plant Sci*. 1998;17(6):575–692.
 70. Van Damme EJM, Lannoo N, Peumans WJ. Plant lectins. In: *Advances in botanical research*. Elsevier; 2008. p. 107–209.
 71. CRUZ PH, CAMPOS EP, Martínez LM, Ortiz B, Martínez G. Las lectinas vegetales como modelo de estudio de las interacciones proteína-carbohidrato. *Rev Educ Bioquímica*. 2005;24(1):21–7.
 72. Itakura Y, Nakamura-Tsuruta S, Kominami J, Tateno H, Hirabayashi J. Sugar-binding profiles of chitin-binding lectins from the hevein family: A comprehensive study. *Int J Mol Sci*. 2017;18(6):1160.
 73. Dang L, Rougé P, Van Damme EJM. Amaranthin-like proteins with aerolysin domains in plants. *Front Plant Sci*. 2017;8:1368.
 74. Santana SS, Gennari-Cardoso ML, Carvalho FC, Roque-Barreira MC, Santiago A da S, Alvim FC, et al. Eutirucallin, a RIP-2 type lectin from the latex of *Euphorbia tirucalli* L. presents proinflammatory properties. *PLoS One*. 2014;9(2):e88422.
 75. Cummings RD, Etzler ME. *Antibodies and lectins in glycan analysis*. 2009;
 76. Lagarda-Diaz I, Guzman-Partida AM, Vazquez-Moreno L. Legume lectins: proteins with diverse applications. *Int J Mol Sci*. 2017;18(6):1242.
 77. Ambrosi M, Cameron NR, Davis BG. Lectins: tools for the molecular understanding of the glycode. *Org Biomol Chem*. 2005;3(9):1593–608.
 78. Lis H, Sharon N. *Moluccella laevis* Lectin. *Trends Glycosci Glycotechnol*. 1994;6(27):65–74.
 79. Cipolla L, Peri F, Airoidi C. Glycoconjugates in cancer therapy. *Anti-Cancer Agents Med Chem (Formerly Curr Med Chem Agents)*. 2008;8(1):92–121.
 80. WILCHES-TORRES A, ROJAS-CARABALLO J, SANABRIA E, REYES-MONTANO E, FERNÁNDEZ-ALONSO J. PURIFICATION AND BIOCHEMICAL CHARACTERIZATION OF AT/TN SPECIFIC LECTIN FROM *LEPECHINIA BULLATA* SEEDS (LAMIACEAE). 2017;
 81. KITAGAKI-OGAWA H, MATSUMOTO I, SENO N, TAKAHASHI N, ENDO S, ARATA Y. Characterization of the carbohydrate moiety of *Clerodendron trichotomum* lectins: Its structure and reactivity toward plant lectins. *Eur J Biochem*. 1986;161(3):779–85.
 82. Wang W, Hause B, Peumans WJ, Smagghe G, Mackie A, Fraser R, et al. The Tn antigen-specific lectin from ground ivy is an insecticidal protein with an unusual physiology. *Plant Physiol*. 2003;132(3):1322–34.
 83. Nasir W, Frank M, Kunze A, Bally M, Parra F, Nyholm P-G, et al. Histo-blood

- group antigen presentation is critical for binding of norovirus VLP to glycosphingolipids in model membranes. *ACS Chem Biol.* 2017;12(5):1288–96.
84. Rougé P, Peumans WJ, Van Damme EJM, Barre A, Singh T, Wu JH, et al. Glycotope structures and intramolecular affinity factors of plant lectins for Tn/T antigens. In: *The Molecular Immunology of Complex Carbohydrates-3.* Springer; 2011. p. 143–54.
 85. Vega NA. Caracterización Bioquímica funcional y biológica de la lectina de *Salvia bogotensis*. PhD Tesis Fac Ciencias Univ Nac Colomb Bogotá. 2004;
 86. Barroso P, Murcia H, Vega N, Pérez G. Purification of IgY against *Salvia bogotensis* lectin. *Biomédica.* 2005;25(4):496–510.
 87. Vega N, Murcia H, Pérez G. CHARACTERIZATION OF *Salvia bogotensis* ANTI-LECTIN IGYS AND THEIR APPLICATION IN IMMUNOCYTOCHEMICAL STUDIES INVOLVING TN ANTIGEN DETECTION. *Rev Colomb Química.* 2009;38(3):363–77.
 88. Torres Romero JC. Purificación y caracterización parcial de mucina citoplasmática utilizando la lectina de *Salvia bogotensis*. *Dep Química.*
 89. Zhao R, Liu X, Wang Y, Jie X, Qin R, Qin W, et al. Integrated glycomic analysis of ovarian cancer side population cells. *Clin Proteomics.* 2016;13(1):1–14.
 90. Bademler S, Zirtiloglu A, Sari M, Ucuncu MZ, Dogru EB, Karabulut S. Clinical significance of serum membrane-bound mucin-2 levels in breast cancer. *Biomolecules.* 2019;9(2):40.
 91. Betge J, Schneider NI, Harbaum L, Pollheimer MJ, Lindtner RA, Kornprat P, et al. MUC1, MUC2, MUC5AC, and MUC6 in colorectal cancer: expression profiles and clinical significance. *Virchows Arch.* 2016;469(3):255–65.
 92. Cazet A, Julien S, Bobowski M, Burchell J, Delannoy P. Tumour-associated carbohydrate antigens in breast cancer. *Breast cancer Res.* 2010;12(3):204.
 93. Ju T, Aryal RP, Kudelka MR, Wang Y, Cummings RD. The Cosmc connection to the Tn antigen in cancer. *Cancer biomarkers.* 2014;14(1):63–81.
 94. Ju T, Otto VI, Cummings RD. The Tn antigen—structural simplicity and biological complexity. *Angew Chemie Int Ed.* 2011;50(8):1770–91.
 95. Kailemia MJ, Park D, Lebrilla CB. Glycans and glycoproteins as specific biomarkers for cancer. *Anal Bioanal Chem.* 2017;409(2):395–410.
 96. Ju T, Wang Y, Aryal RP, Lehoux SD, Ding X, Kudelka MR, et al. Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers. *PROTEOMICS—Clinical Appl.* 2013;7(9–10):618–31.
 97. Lisowska E. Tn antigens and their significance in oncology. *Acta Biochim Pol.* 1995;42(1):11–7.
 98. Mall AS. Analysis of mucins: role in laboratory diagnosis. *J Clin Pathol.* 2008;61(9):1018–24.
 99. Atanasova KR, Reznikov LR. Strategies for measuring airway mucus and mucins. *Respir Res.* 2019;20(1):1–14.
 100. Lakshmanan I, Ponnusamy MP, Macha MA, Haridas D, Majhi PD, Kaur S, et al. Mucins in lung cancer: diagnostic, prognostic, and therapeutic implications. *J Thorac Oncol.* 2015;10(1):19–27.
 101. Malaker SA, Pedram K, Ferracane MJ, Bensing BA, Krishnan V, Pett C, et al. The mucin-selective protease StcE enables molecular and functional analysis

- of human cancer-associated mucins. *Proc Natl Acad Sci*. 2019;116(15):7278–87.
102. Yau T, Dan X, Ng CCW, Ng TB. Lectins with potential for anti-cancer therapy. *Molecules*. 2015;20(3):3791–810.
 103. Mazalovska M, Kouokam JC. Lectins as promising therapeutics for the prevention and treatment of HIV and other potential coinfections. *Biomed Res Int*. 2018;2018.
 104. Rougé P, Peumans WJ, Van Damme EJM, Barre A, Singh T, Wu JH, et al. Structure-function relationships of plant lectins that specifically recognize T and Tn antigens. *Mol Immunol Complex Carbohydrates*, 3rd ed Springer. 2011;157–70.
 105. Sousa BL, Silva-Filho JC, Kumar P, Graewert MA, Pereira RI, Cunha RMS, et al. Structural characterization of a *Vatairea macrocarpa* lectin in complex with a tumor-associated antigen: a new tool for cancer research. *Int J Biochem Cell Biol*. 2016;72:27–39.
 106. Fernández-Alonso JL. ESTUDIOS EN LABIATAE DE COLOMBIA IV. NOVEDADES EN SALVIA Y SINOPSIS DE LAS SECCIONES ANGULATAE Y PURPUREAE/Studies in Colombian Labiatae IV. Novelties in Salvia and synopsis of sections Angulatae and Purpureae. *Caldasia*. 2003;235–81.
 107. Fernández Alonso JL. Salvia guacana, una nueva Labiatae de Colombia con flores resupinadas y sinopsis de Salvia sect. Tubiflorae. *Rev la Acad Colomb Ciencias Exactas, Físicas y Nat*. 2012;36(141):517–33.
 108. Nieto G. Biological activities of three essential oils of the Lamiaceae family. *Medicines*. 2017;4(3):63.
 109. Satyal P, Jones TH, Lopez EM, McFeeters RL, Ali NAA, Mansi I, et al. Chemotypic characterization and biological activity of *Rosmarinus officinalis*. *Foods*. 2017;6(3):20.
 110. Uritu CM, Mihai CT, Stanciu G-D, Dodi G, Alexa-Stratulat T, Luca A, et al. Medicinal plants of the family Lamiaceae in pain therapy: A review. *Pain Res Manag*. 2018;2018.
 111. Martínez-Gordillo M, Bedolla-García B, Cornejo-Tenorio G, Fragoso-Martínez I, García-Peña M del R, González-Gallegos JG, et al. Lamiaceae de México. *Bot Sci*. 2017;95(4):780–806.
 112. Wiederschain GY. *Essentials of glycobiology*. Springer Nature BV; 2009.
 113. Hart GW, Copeland RJ. Glycomics hits the big time. *Cell*. 2010;143(5):672–6.
 114. Smith PK, Krohn RI, Hermanson GT, Mallia a K, Gartner FH, Provenzano MD. Measurement of protein using bicinchoninic acid *Anal Biochem* 150: 76–85. *Anal Biochem*. 1985;150(1):76–85.
 115. Schagger H, Von Jagow G. Tricine-sodium dodecylsulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*. 1987;166(1 987):368–79.
 116. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;227(5259):680–5.
 117. Switzer III RC, Merrill CR, Shifrin S. A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal Biochem*. 1979;98(1):231–7.
 118. Hirohashi S, Clausen H, Yamada T, Shimosato Y, Hakomori S. Blood group A

- cross-reacting epitope defined by monoclonal antibodies NCC-LU-35 and-81 expressed in cancer of blood group O or B individuals: its identification as Tn antigen. *Proc Natl Acad Sci.* 1985;82(20):7039–43.
119. Vega NA. Detección, purificación y caracterización parcial de la lectina presente en las semillas del género *Salvia*. Thesis, Chemistry, Universidad Nacional de Colombia, Bogotá; 1997.
 120. Stoscheck CM. [6] Quantitation of protein. In: *Methods in enzymology*. Elsevier; 1990. p. 50–68.
 121. Baines MG, Thorpe R. Purification of immunoglobulin G (IgG). In: *Immunochemical Protocols*. Springer; 1992. p. 79–104.
 122. J. R. Extracción, purificación y caracterización parcial de la lectina presente en las semillas de *Lepechinia bullata*. 2004.
 123. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, Stanley BA, et al. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Curr Protoc Mol Biol.* 2010;90(1):10–25.
 124. Swaney DL, McAlister GC, Coon JJ. Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods.* 2008;5(11):959–64.
 125. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc mass Spectrom.* 1994;5(11):976–89.
 126. Banerjee R, Das K, Ravishankar R, Suguna K, Surolia A, Vijayan M. Conformation, protein-carbohydrate interactions and a novel subunit association in the refined structure of peanut lectin-lactose complex. *J Mol Biol.* 1996;259(2):281–96.
 127. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5(4):725.
 128. Yang J, Roy A, Zhang Y. BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic Acids Res.* 2012;41(D1):D1096–103.
 129. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc.* 2009;4(3):363–71.
 130. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31(2):455–61.
 131. Schinazi RF, Kohler JJ, Kim B. Reverse transcription. In: *Brenner's Encyclopedia of Genetics*. Elsevier; 2013. p. 224.
 132. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* (80-). 1988;239(4839):487–91.
 133. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10.
 134. Abbas AK, Lichtman AH, Pillai S. *Inmunología celular y molecular*. Elsevier Health Sciences; 2022.
 135. Bergmann-Leitner ES, Mease RM, Duncan EH, Khan F, Waitumbi J, Angov E. Evaluation of immunoglobulin purification methods and their impact on quality and yield of antigen-specific antibodies. *Malar J.* 2008;7(1):129.
 136. Eivazi S, Majidi J. Production and purification of a polyclonal antibody against

- purified mouse IgG2b in rabbits towards designing mouse monoclonal isotyping kits. *Adv Pharm Bull.* 2015;5(1):109.
137. Moser AC, Hage DS. Immunoaffinity chromatography: an introduction to applications and recent developments. *Bioanalysis.* 2010;2(4):769–90.
 138. Baldikova E, Pospiskova K, Ladakis D, Kookos IK, Koutinas AA, Safarikova M, et al. Magnetically modified bacterial cellulose: A promising carrier for immobilization of affinity ligands, enzymes, and cells. *Mater Sci Eng C.* 2017;71:214–21.
 139. Turner AJ. Neprilysin. In: *Handbook of proteolytic enzymes.* Elsevier; 2004. p. 419–26.
 140. Sanabria E. Aislamiento, purificación y caracterización parcial de la lectina presente en las semillas de *Lepechinia bullata*. 1999.
 141. Van Driessche E, Beeckmans S, Dejaegere R, Kanarek L. Thiourea: the antioxidant of choice for the purification of proteins from phenol-rich plant tissues. *Anal Biochem.* 1984;141(1):184–8.
 142. Aoki T. A comprehensive review of our current understanding of red blood cell (RBC) glycoproteins. *Membranes (Basel).* 2017;7(4):56.
 143. Iskratsch T, Braun A, Paschinger K, Wilson IBH. Specificity analysis of lectins and antibodies using remodeled glycoproteins. *Anal Biochem.* 2009;386(2):133–46.
 144. Wu AM, Liu J-H, Singh T, Yang Z. Recognition Roles of Mammalian Structural Units and Polyvalency in Lectin–Glycan Interactions. In: *The Molecular Immunology of Complex Carbohydrates-3.* Springer; 2011. p. 99–116.
 145. Wu AM. Polyvalency of Tn (GalNAc α 1 \rightarrow Ser/Thr) glycotope as a critical factor for *Vicia villosa* B4 and glycoprotein interactions. *FEBS Lett.* 2004;562(1–3):51–8.
 146. Fernandez-Alonso JL, Vega N, Pérez G. Lectin prospecting in Colombian Labiatae. A systematic-ecological approach-III. Mainly exotic species (cultivated or naturalised). *Caldasia.* 2009;31(2):227–45.
 147. Nagano CS, Calvete JJ, Baretino D, Pérez A, Cavada BS, Sanz L. Insights into the structural basis of the pH-dependent dimer–tetramer equilibrium through crystallographic analysis of recombinant Diocleinae lectins. *Biochem J.* 2008;409(2):417–28.
 148. Quintero Moreno M. Elucidación parcial de la estructura primaria de la lectina LGL-P2 y purificación y caracterización parcial de la lectina LGL-P4 presentes en semillas de *Galactia lindenii*. *Dep Química.*
 149. Cavada BS, Marinho ES, Souza EP, Benevides RG, Delatorre P, Souza LAG, et al. Purification, partial characterization and preliminary X-ray diffraction analysis of a mannose-specific lectin from *Cymbosema roseum* seeds. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2006;62(3):235–7.
 150. Perez G, Hernandez M, Mora E. Isolation and characterization of a lectin from the seeds of *Dioclea lehmanni*. *Phytochemistry.* 1990;29(6):1745–9.
 151. Sierra A, Pérez G. Extracción, purificación y caracterización de dos lectinas en semillas de *Dioclea sericea*. *Rev Acad Col Ciencias.* 1999;23:445–54.
 152. Melgarejo LM, Vega N, Pérez G. Isolation and characterization of novel lectins from *Canavalia ensiformis* DC and *Dioclea grandiflora* Mart. ex Benth. seeds. *Brazilian J Plant Physiol.* 2005;17(3):315–24.

153. Wilson IBH. Glycosylation of proteins in plants and invertebrates. *Curr Opin Struct Biol.* 2002;12(5):569–77.
154. Wang Y-Y, Lin S-Y, Chen W-Y, Liao S-L, Wu C-C, Pan P-H, et al. *Glechoma hederacea* extracts attenuate cholestatic liver injury in a bile duct-ligated rat model. *J Ethnopharmacol.* 2017;204:58–66.
155. Babino A, Tello D, Rojas A, Bay S, Osinaga E, Alzari PM. The crystal structure of a plant lectin in complex with the Tn antigen. *FEBS Lett.* 2003;536(1–3):106–10.
156. Ravishankar R, Ravindran M, Suguna K, Surolia A, Vijayan M. Crystal structure of the peanut lectin–T-antigen complex. Carbohydrate specificity generated by water bridges. *Curr Sci.* 1997;855–61.
157. Transue TR, Smith AK, Mo H, Goldstein IJ, Saper MA. Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nat Struct Biol.* 1997;4(10):779–83.
158. Jeyaprakash AA, Rani PG, Reddy GB, Banumathi S, Betzel C, Sekar K, et al. Crystal structure of the jacalin–T-antigen complex and a comparative study of lectin–T-antigen complexes. *J Mol Biol.* 2002;321(4):637–45.
159. Alperin DM, Latter H, Lis H, Sharon N. Isolation, by affinity chromatography and gel filtration in 8 M-urea, of an active subunit from the anti-(blood-group A+ N)-specific lectin of *Moluccella laevis*. *Biochem J.* 1992;285(1):1–4.
160. Morris GM, Lim-Wilby M. Molecular docking. In: *Molecular modeling of proteins.* Springer; 2008. p. 365–82.
161. Forli W, Halliday S, Belew R, Olson A. AutoDock Version 4.2. Citeseer. 2012.
162. Sharon N, Lis H. *Lectins.* Springer Science & Business Media; 2003.
163. Vennapusa AR, Somayanda IM, Doherty CJ, Jagadish S V. A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber. *Sci Rep.* 2020;10(1):1–13.
164. Sangha JS, Gu K, Kaur J, Yin Z. An improved method for RNA isolation and cDNA library construction from immature seeds of *Jatropha curcas* L. *BMC Res Notes.* 2010;3(1):1–6.
165. Chang S, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. *Plant Mol Biol Report.* 1993;11(2):113–6.
166. Iandolo AB, Goes da Silva F, Lim H, Choi H, Williams LE, Cook DR. High-quality RNA, cDNA, and derived EST libraries from grapevine (*Vitis vinifera* L.). *Plant Mol Biol Report.* 2004;22(3):269–78.
167. Gasic K, Hernandez A, Korban SS. RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. *Plant Mol Biol Report.* 2004;22(4):437–8.
168. White EJ, Venter M, Hiten NF, Burger JT. Modified Cetyltrimethylammonium bromide method improves robustness and versatility: the benchmark for plant RNA extraction. *Wiley Online Library;* 2008.
169. Rio DC, Ares M, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harb Protoc.* 2010;2010(6):pdb-prot5439.
170. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids. *JoVE (Journal Vis Exp.* 2010;(45):e2565.
171. Skrypina NA, Timofeeva A V, Khaspekov GL, Savochkina LP, Beabealashvili RS. Total RNA suitable for molecular biology analysis. *J Biotechnol.*

- 2003;105(1–2):1–9.
172. Liu L, Han R, Yu N, Zhang W, Xing L, Xie D, et al. A method for extracting high-quality total RNA from plant rich in polysaccharides and polyphenols using *Dendrobium huoshanense*. *PLoS One*. 2018;13(5):e0196592.
 173. Fernández-Alonso JL, Vega N, Filgueira JJ, Pérez G. Lectin prospecting in Colombian Labiatae. A systematic-ecological approach. *Biochem Syst Ecol*. 2003;31(6):617–33.
 174. Pérez G, Vega N, Fernández-Alonso JL. LECTIN PROSPECTING IN COLOMBIAN LABIATAE. A SYSTEMATIC-ECOLOGICAL APPROACH-II. *Caldasia*. 2006;28(2):179–95.
 175. Ghaffariyan S, Mohammadi SA, Aharizad S. DNA isolation protocol for the medicinal plant lemon balm (*Melissa officinalis*, Lamiaceae). *Genet Mol Res*. 2012;11(2):1049–57.
 176. Li Y, Sun C, Luo H-M, Li X-W, Niu Y-Y, Chen S-L. Transcriptome characterization for *Salvia miltiorrhiza* using 454 GS FLX. *Yao xue xue bao= Acta Pharm Sin*. 2010;45(4):524–9.
 177. O’neill M, McPartlin J, Arthure K, Riedel S, McMillan ND. Comparison of the TLDA with the Nanodrop and the reference Qubit system. In: *Journal of Physics: Conference Series*. IOP Publishing; 2011. p. 12047.
 178. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13(1):1–11.
 179. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. 2000;
 180. Banaganapalli B, Shaik NA, Rashidi OM, Jamalalail B, Bahattab R, Bokhari HA, et al. In Silico PCR. In: *Essentials of Bioinformatics, Volume I*. Springer; 2019. p. 355–71.
 181. Wenping H, Yuan Z, Jie S, Lijun Z. De novo transcriptome sequencing in *Salvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics*. 2011;98(4):272–9.

Anexos

A. ANEXO. Esquema inmunización conejos para obtener anti LSB0-I

Tabla A-1. Protocolo de Inmunización con lectina de *S. bogotensis*¹. Tomado de (85)

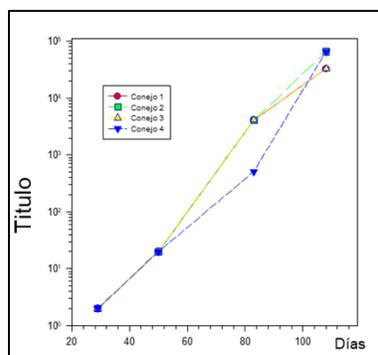
Conejo	Reto inicial	Refuerzo (Día)			
		7	22	37	57
1	0.240	0.134	0.161	0.166	0.217
2	0.153	0.150	0.120	0.132	0.217
3	0.207	0.198	0.138	0.166	0.261
4	0.290	0.160	0.124	0.139	0.261
5	0.00	0.00	0.00	0.00	0.00

¹ 1 mg de lectina de *S. bogotensis* empleados en cada inoculación

Tabla A-2. Obtención de antisueros para evaluar la respuesta al antígeno

Muestra	Días después de la 1 ^a inoculación	mL por Conejo			
		1	2	3	4
1. Primer sangrado parcial	29				
2. Segundo sangrado parcial	50				
3. Tercer sangrado parcial	83				
4. Cuarto sangrado parcial	108				
5. Primera punción	91	15	12	10	12
6. Segunda punción	117	25	34	45	27

Figura A-1. Determinación del título del Antisero anti-lectina LSB0-I.



B. ANEXO. Purificación por Sepharosa 4B- ConA

Tabla B-1. Cuantificación de purificación por Sepharosa 4B- ConA de la fracción Tris-HCl pH 8.3

Fracción	Proteína (mg/mL)*	Vol (μL)	(μg) Totales
FR NaCl 1%-Glucosa 0,2M	0,104	150	15,6
FR Gly -HCl pH 2,8	0,628	175	110
FNR	0,0604	100	60

*Cuantificación por el método del ácido bicinconinico (BCA)

FNR: Fracción no retenida, FR: Fracción retenida

C. ANEXO. Secuencia de péptidos obtenidos por espectrometría de MS/MS

Tabla C-1. Secuencia de péptidos obtenidos por espectrometría de MS/MS, % de cobertura con proteína con dominio tipo Leguminosa y alineamiento.

N°	Secuencia Péptido	% de Cobertura	Proteína												
1	TQGWIDYEAGSNRLEVR	5	Proteína de <i>Malus baccata</i> que contiene el dominio de Lectina_legB												
Alineamiento:															
<p>hypothetical protein DVH24_006186 [Malus domestica] Sequence ID: RXH81361.1 Length: 351 Number of Matches: 1</p> <p>Range 1: 184 to 200</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>59.2 bits(132)</td> <td>1e-08</td> <td>17/17(100%)</td> <td>17/17(100%)</td> <td>0/17(0%)</td> </tr> </tbody> </table> <p>Query 1 TQGWIDYEAGSNRLEVR 17 TQGWIDYEAGSNRLEVR Sbjct 184 TQGWIDYEAGSNRLEVR 200</p>				Score	Expect	Identities	Positives	Gaps	59.2 bits(132)	1e-08	17/17(100%)	17/17(100%)	0/17(0%)		
Score	Expect	Identities	Positives	Gaps											
59.2 bits(132)	1e-08	17/17(100%)	17/17(100%)	0/17(0%)											
2	IGFSASTGDWVEVHNILSW SFNSSLEINNGKNNMGLIIG LVVEPSK	9	Proteína de <i>Acer yangbiense</i> que contiene el dominio LectinlegB												
<p>hypothetical protein EZV62_014624 [Acer yangbiense] Sequence ID: TXG60051.1 Length: 536 Number of Matches: 1</p> <p>Range 1: 240 to 285</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>94.0 bits(232)</td> <td>8e-21</td> <td>Compositional matrix adjust.</td> <td>46/46(100%)</td> <td>46/46(100%)</td> <td>0/46(0%)</td> </tr> </tbody> </table> <p>Query 1 IGFSASTGDWVEVHNILSWSFNSSLEINNGKNNMGLIIGLVVEPSK 46 IGFSASTGDWVEVHNILSWSFNSSLEINNGKNNMGLIIGLVVEPSK Sbjct 240 IGFSASTGDWVEVHNILSWSFNSSLEINNGKNNMGLIIGLVVEPSK 285</p>				Score	Expect	Method	Identities	Positives	Gaps	94.0 bits(232)	8e-21	Compositional matrix adjust.	46/46(100%)	46/46(100%)	0/46(0%)
Score	Expect	Method	Identities	Positives	Gaps										
94.0 bits(232)	8e-21	Compositional matrix adjust.	46/46(100%)	46/46(100%)	0/46(0%)										

3	MEISRLDQDNQQLMAIGSGI R	8	Proteína de <i>Camellia sinensis</i> var. <i>sinensis</i> que contiene el dominio lectin_legB												
<p>hypothetical protein TEA_018348 [<i>Camellia sinensis</i> var. <i>sinensis</i>] Sequence ID: THF98249.1 Length: 280 Number of Matches: 1</p> <p>Range 1: 1 to 21</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>71.9 bits(162)</td> <td>8e-13</td> <td>21/21(100%)</td> <td>21/21(100%)</td> <td>0/21(0%)</td> </tr> </tbody> </table> <p>Query 1 MEISRLDQDNQQLMAIGSGIR 21 MEISRLDQDNQQLMAIGSGIR Sbjct 1 MEISRLDQDNQQLMAIGSGIR 21</p>				Score	Expect	Identities	Positives	Gaps	71.9 bits(162)	8e-13	21/21(100%)	21/21(100%)	0/21(0%)		
Score	Expect	Identities	Positives	Gaps											
71.9 bits(162)	8e-13	21/21(100%)	21/21(100%)	0/21(0%)											
4	NGSSWRLSATLAEPTDD R	2	Dominio de lectina de leguminosa <i>Paenibacillus thiaminolyticus</i>												
<p>L-type lectin-domain containing protein [<i>Paenibacillus thiaminolyticus</i>] Sequence ID: WP_087443418.1 Length: 807 Number of Matches: 1</p> <p>Range 1: 714 to 733</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>66.0 bits(148)</td> <td>8e-11</td> <td>20/20(100%)</td> <td>20/20(100%)</td> <td>0/20(0%)</td> </tr> </tbody> </table> <p>Query 1 NGSSWRLSATLAEPTDDR 20 NGSSWRLSATLAEPTDDR Sbjct 714 NGSSWRLSATLAEPTDDR 733</p>				Score	Expect	Identities	Positives	Gaps	66.0 bits(148)	8e-11	20/20(100%)	20/20(100%)	0/20(0%)		
Score	Expect	Identities	Positives	Gaps											
66.0 bits(148)	8e-11	20/20(100%)	20/20(100%)	0/20(0%)											
5	QVLADVAEGLNYLHHGWD QVVVHRDIKSSNILLDSMR	6	Proteína Quinasa de <i>Populus trichocarpa</i> con dominio de lectina tipo -L												
<p>PREDICTED: L-type lectin-domain containing receptor kinase S.1-like [<i>Populus euphratica</i>] Sequence ID: XP_011046223.1 Length: 674 Number of Matches: 1</p> <p>Range 1: 466 to 503</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>85.5 bits(210)</td> <td>6e-18</td> <td>Composition-based stats.</td> <td>38/38(100%)</td> <td>38/38(100%)</td> <td>0/38(0%)</td> </tr> </tbody> </table> <p>Query 1 QVLADVAEGLNYLHHGWDQVVVHRDIKSSNILLDSMR 38 QVLADVAEGLNYLHHGWDQVVVHRDIKSSNILLDSMR Sbjct 466 QVLADVAEGLNYLHHGWDQVVVHRDIKSSNILLDSMR 503</p>				Score	Expect	Method	Identities	Positives	Gaps	85.5 bits(210)	6e-18	Composition-based stats.	38/38(100%)	38/38(100%)	0/38(0%)
Score	Expect	Method	Identities	Positives	Gaps										
85.5 bits(210)	6e-18	Composition-based stats.	38/38(100%)	38/38(100%)	0/38(0%)										
6	LKLNGVTVGSFVVLVLIAL FSLFLIKR	5	Protein kinase domain-containing protein <i>Gossypium barbadense</i>												
<p>PREDICTED: L-type lectin-domain containing receptor kinase VII.1-like isoform X2 [<i>Gossypium arboreum</i>] Sequence ID: XP_017644138.1 Length: 589 Number of Matches: 1</p> <p>Range 1: 192 to 219</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>88.9 bits(202)</td> <td>3e-18</td> <td>28/28(100%)</td> <td>28/28(100%)</td> <td>0/28(0%)</td> </tr> </tbody> </table> <p>Query 1 LKLNGVTVGSFVVLVLIALFSLFLIKR 28 LKLNGVTVGSFVVLVLIALFSLFLIKR Sbjct 192 LKLNGVTVGSFVVLVLIALFSLFLIKR 219</p>				Score	Expect	Identities	Positives	Gaps	88.9 bits(202)	3e-18	28/28(100%)	28/28(100%)	0/28(0%)		
Score	Expect	Identities	Positives	Gaps											
88.9 bits(202)	3e-18	28/28(100%)	28/28(100%)	0/28(0%)											
7	LSLGEIKLATMGFNQNRVV GEGGSATVYK	4	Protein kinase domain-containing protein												

<p>L-type lectin-domain containing receptor kinase S.6-like [Cucumis sativus] Sequence ID: XP_031742685.1 Length: 694 Number of Matches: 1 Range 1: 358 to 386</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>92.3 bits(210)</td> <td>2e-19</td> <td>29/29(100%)</td> <td>29/29(100%)</td> <td>0/29(0%)</td> </tr> </tbody> </table> <p>Query 1 LSLGEIKLATMGFNQNRVVGEGGSATVYK 29 LSLGEIKLATMGFNQNRVVGEGGSATVYK Sbjct 358 LSLGEIKLATMGFNQNRVVGEGGSATVYK 386</p>				Score	Expect	Identities	Positives	Gaps	92.3 bits(210)	2e-19	29/29(100%)	29/29(100%)	0/29(0%)		
Score	Expect	Identities	Positives	Gaps											
92.3 bits(210)	2e-19	29/29(100%)	29/29(100%)	0/29(0%)											
8	QFVAEVVTMGNIQHRNLVPL LLGYCR	4	Putative L-type lectin-domain containing receptor kinase I.4 Arabidopsis thaliana												
<p>Concanavalin A-like lectin protein kinase family protein [Arabidopsis thaliana] Sequence ID: NP_190128.1 Length: 667 Number of Matches: 1 Range 1: 390 to 414</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>85.9 bits(195)</td> <td>2e-17</td> <td>25/25(100%)</td> <td>25/25(100%)</td> <td>0/25(0%)</td> </tr> </tbody> </table> <p>Query 1 QFVAEVVTMGNIQHRNLVPLLLGYCR 25 QFVAEVVTMGNIQHRNLVPLLLGYCR Sbjct 390 QFVAEVVTMGNIQHRNLVPLLLGYCR 414</p>				Score	Expect	Identities	Positives	Gaps	85.9 bits(195)	2e-17	25/25(100%)	25/25(100%)	0/25(0%)		
Score	Expect	Identities	Positives	Gaps											
85.9 bits(195)	2e-17	25/25(100%)	25/25(100%)	0/25(0%)											
9	QLAVELDTSMNDFDPDGNH VGIDTRSIHPFAAK	5	probable L-type lectin-domain containing receptor kinase S.5 Juglans regia												
<p>probable L-type lectin-domain containing receptor kinase S.5 [Juglans regia] Sequence ID: XP_018812869.1 Length: 668 Number of Matches: 1 Range 1: 149 to 182</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>76.6 bits(187)</td> <td>8e-15</td> <td>Composition-based stats.</td> <td>34/34(100%)</td> <td>34/34(100%)</td> <td>0/34(0%)</td> </tr> </tbody> </table> <p>Query 1 QLAVELDTSMNDFDPDGNHVGIDTRSIHPFAAK 34 QLAVELDTSMNDFDPDGNHVGIDTRSIHPFAAK Sbjct 149 QLAVELDTSMNDFDPDGNHVGIDTRSIHPFAAK 182</p>				Score	Expect	Method	Identities	Positives	Gaps	76.6 bits(187)	8e-15	Composition-based stats.	34/34(100%)	34/34(100%)	0/34(0%)
Score	Expect	Method	Identities	Positives	Gaps										
76.6 bits(187)	8e-15	Composition-based stats.	34/34(100%)	34/34(100%)	0/34(0%)										
10	VQYDGKNISVFVPLVNLST NMNGPVLTLPLNLSDYLP K	6	Serine/threonine protein kinase Trema orientale												
<p>probable L-type lectin-domain containing receptor kinase S.5 [Ziziphus jujuba] Sequence ID: XP_015898912.1 Length: 660 Number of Matches: 1 Range 1: 197 to 236 GenPept Graphics</p> <table border="1"> <thead> <tr> <th>Score</th> <th>Expect</th> <th>Method</th> <th>Identities</th> <th>Positives</th> <th>Gaps</th> </tr> </thead> <tbody> <tr> <td>58.5 bits(140)</td> <td>2e-08</td> <td>Compositional matrix adjust.</td> <td>28/40(70%)</td> <td>32/40(80%)</td> <td>0/40(0%)</td> </tr> </tbody> </table> <p>Query 1 VQYDGKNISVFVPLVNLSTNMNGPVLTLPLNLSDYLPK 40 V+YDGKNISVFV L N T+ M PVL+PL+LSD LPEK Sbjct 197 VEYDGKNISVFVTLTNETAGEMKNPVLTMPLDLSDLLPEK 236</p>				Score	Expect	Method	Identities	Positives	Gaps	58.5 bits(140)	2e-08	Compositional matrix adjust.	28/40(70%)	32/40(80%)	0/40(0%)
Score	Expect	Method	Identities	Positives	Gaps										
58.5 bits(140)	2e-08	Compositional matrix adjust.	28/40(70%)	32/40(80%)	0/40(0%)										
11	IFPNSSKLVFGFSAATGNAV SGHYILWWSFSTGRGSPH GLDISK	6	Protein kinase domain-containing protein Brassica campestris												

L-type lectin-domain containing receptor kinase I.9-like [Brassica napus]					
Sequence ID: XP_022549080.1 Length: 752 Number of Matches: 1					
Range 1: 283 to 326					
Score	Expect	Method	Identities	Positives	Gaps
93.6 bits(231)	1e-20	Compositional matrix adjust.	44/44(100%)	44/44(100%)	0/44(0%)
Query	1	IFPNSSKLFVGFSAATGNAVSGHYILWWSFSTGRGSPHGLDISK		44	
		IFPNSSKLFVGFSAATGNAVSGHYILWWSFSTGRGSPHGLDISK			
Sbjct	283	IFPNSSKLFVGFSAATGNAVSGHYILWWSFSTGRGSPHGLDISK		326	
12	VSKSSKQGINEYASEVR	3	Proteína que contiene el dominio de lectina/gluconasa similar a la concaivalina A [Artemisia annua]		

D. ANEXO. Alineamiento múltiple de la secuencia parcial de LSBo-I con los 12 péptidos obtenidos de MS/MS.

Figura D-1. Múltiple alineamiento realizado en Clustal O (1.2.4) de la secuencia parcial de LSBo-I con los 12 péptidos obtenidos de MS/MS.

```

CLUSTAL O(1.2.4) multiple sequence alignment

Pep2      -----0
Pep11     -----0
Pep12     -----0
Pep9      -----0
Pep6      ---LKLNGVTVGSFVVLVLI-----19
Pep10     ---VQYDGKNIISVFVPLVNLSTNMNGPVLTLP-----30
Pep8      --QFVAEVV-----TMGN-----IQHRNLV-----18
Pep1      -----0
Pep5      -----QVL-----3
Pep7      -----LSLGEIK-----LATMGFNQNRVVG-----21
Pep3      -----MEISRLD-----7
LsB0-I    MATLLQNLLTATSRxxTLLDLNTRxSxTLLTHPNFNGDLGHGQHGSGSQSPSPSRFxA60
Pep4      -----0

Pep2      -----0
Pep11     -----0
Pep12     -----0
Pep9      -----0
Pep6      -----19
Pep10     -----30
Pep8      -----18
Pep1      -----0
Pep5      -----3
Pep7      -----GGSA-----25
Pep3      -----7
LsB0-I    LTGALAQKVGTSVHPNPSSNLEGGDAFEFVGPKxFxITxxxxxxxADGxAxFxxPVGxxx120
Pep4      -----0

```

Pep2	-----	0
Pep11	-----	0
Pep12	-----	0
Pep9	-----	0
Pep6	-----	19
Pep10	-----	30
Pep8	-----PLLGYCR-----	25
Pep1	-----	0
Pep5	-----AD-----	5
Pep7	-----	25
Pep3	-----	7
LsB0-I	PSGSTGGYLGAVNPKGxGxxxQGVADLNGxGSYGGGGGHHGFCGSGSSNSELVQSSG	180
Pep4	-----	0
Pep2	-----IGFSASTGDWVEV	13
Pep11	-----IFPNSSKLFVGFSAATGNAVSG	22
Pep12	-----	0
Pep9	-----QLAVELDTSMNDFDPD	16
Pep6	-----LFSLFLLIKR-----	28
Pep10	-----LNLSDYLPEK-----	40
Pep8	-----	25
Pep1	-----	0
Pep5	-----VAEGLNYL-----	13
Pep7	-----	25
Pep3	-----	7
LsB0-I	LGHQESRxINYxxxxxYEELQVTAGxRKxxxxxSYEAELDDFVTLQVTVGLSAATSGDGVE	240
Pep4	-----	0
Pep11	HYILWWSFSTGRGSPHGLDIS----K-----	44
Pep12	-----VSKSSKQGINEY-----ASEVR-----	17
Pep9	GNHVGIDT-RSIIHPFAK-----	34
Pep6	-----	28
Pep10	-----	40
Pep8	-----	25
Pep1	--TQGWIDYEAGS---NRLE---VR-----	17
Pep5	--HHGWDQVVVHR---DIKSSNILLDSDMR-----	38
Pep7	---TVYK-----	29
Pep3	-----QDNQQLMAIGSGIR-----	21
LsB0-I	KDIxxWxFTAxMxxxxxxxxxxxxxxxxGxxIRQxV----	274
Pep4	GNGSSWRLSATLAEPFTDDR-----	20

E. ANEXO. Validación inicial de los modelos con mejor score de I-Tasser y Phyre²

Figura E-1. Validación predicha del modelo obtenido por I-Tasser empleando MolProbity

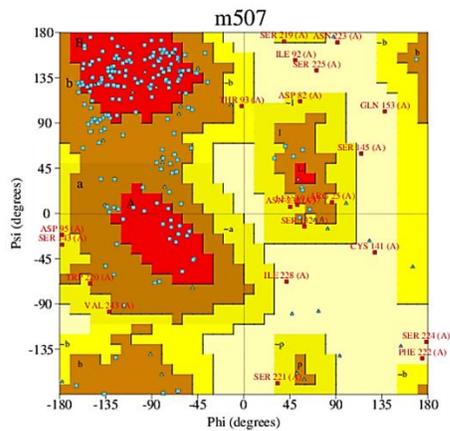
Summary statistics

Protein Geometry	Poor rotamers	28	14.43%	Goal: <0.3%
	Favored rotamers	128	65.98%	Goal: >98%
	Ramachandran outliers	31	12.65%	Goal: <0.05%
	Ramachandran favored	176	71.84%	Goal: >98%
	Rama distribution Z-score	-5.59 ± 0.40		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	12	5.77%	Goal: 0
	Bad bonds:	1 / 1830	0.05%	Goal: 0%
	Bad angles:	53 / 2478	2.14%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	1 / 10	10.00%	Expected: ≤1 per chain, or ≤5%
	Twisted Peptides:	26 / 246	10.57%	Goal: 0
Low-resolution Criteria	CaBLAM outliers	25	10.3%	Goal: <1.0%
	CA Geometry outliers	8	3.29%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/267		

In the two column results, the left column gives the raw count, right column gives the percentage.

Key to table colors and cutoffs here: [?](#)

Figura E-2. Validación predicha del modelo obtenido por I-Tasser empleando Ramachandran.



PROCHECK statistics

1. Ramachandran Plot statistics

	No. of residues	%-tage
Most favoured regions [A,B,L]	112	57.1%*
Additional allowed regions [a,b,l,p]	63	32.1%
Generously allowed regions [-a,-b,-l,-p]	12	6.1%
Disallowed regions [XX]	9	4.6%*
Non-glycine and non-proline residues	196	100.0%
End-residues (excl. Gly and Pro)	2	
Glycine residues	39	
Proline residues	10	
Total number of residues	247	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

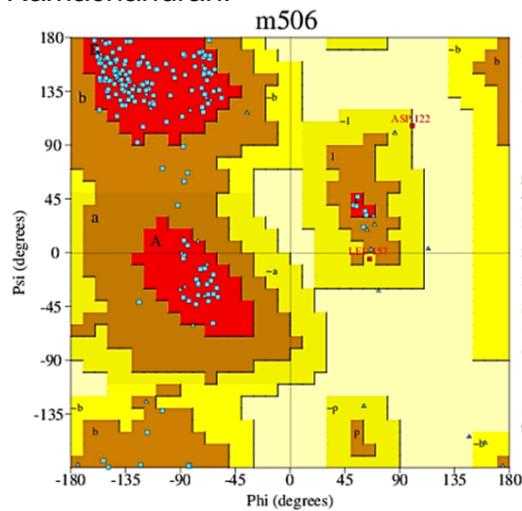
Figura E-3. Validación predicha del modelo obtenido por Phyre² empleando MolProbity

Summary statistics

Protein Geometry	Poor rotamers	2	1.30%	Goal: <0.3%
	Favored rotamers	152	98.70%	Goal: >98%
	Ramachandran outliers	9	4.50%	Goal: <0.05%
	Ramachandran favored	173	86.50%	Goal: >98%
	Rama distribution Z-score	-3.22 ± 0.50		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	24 / 1488	1.61%	Goal: 0%
Bad angles:	22 / 2016	1.09%	Goal: <0.1%	
Peptide Omegas	Cis Prolines:	0 / 9	0.00%	Expected: ≤1 per chain, or ≤5%
	Cis nonProlines:	1 / 192	0.52%	Goal: <0.05%
Low-resolution Criteria	CaBLAM outliers	17	8.6%	Goal: <1.0%
	CA Geometry outliers	5	2.53%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/209		

In the two column results, the left column gives the raw count, right column gives the percentage.
Key to table colors and cutoffs here: [?](#)

Figura E-4. Validación predicha del modelo obtenido por Phyre² empleando Ramachandran.



PROCHECK statistics

1. Ramachandran Plot statistics

	No. of residues	%-tage
Most favoured regions [A,B,L]	138	89.0%*
Additional allowed regions [a,b,l,p]	15	9.7%
Generously allowed regions [-a,-b,-l,-p]	1	0.6%
Disallowed regions [XX]	1	0.6%*
Non-glycine and non-proline residues	155	100.0%
End-residues (excl. Gly and Pro)	2	
Glycine residues	36	
Proline residues	9	
Total number of residues	202	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20.0 a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

F. ANEXO. Cuantificación extracción de RNA de semillas de *S. bogotensis*

Tabla F-1. Extracción de RNA de semillas de *S. bogotensis*

Método	CTAB		Trizol						
	Tubo 1	Tubo 2	Trizol ZYMO RESEARCH				Trizol ambion by life		
Tubo	Tubo 1	Tubo 2	Tubo 1	Tubo 2	Tubo 3	Tubo 4	Tubo 5	Tubo 6	Tubo 7
muestra mg	730	780	180	178	410	380	135	205	415
260/280	1,7	1,6	1,52	1,47	1,114	1,07	1,19	1,56	1,2

G. ANEXO. Tabla de codones de *Salvia miltiorrhiza*

Tabla G-1. Tabla de codones de *S. miltiorrhiza* [gbpln], empleada para la traducción inversa de la secuencia de la LSBo-I.

AA	Codón	Numero/1000		AA	Codón	Numero/1000	
Gly	GGG	72.00	14.86	Leu	TTG	95.00	19.61
Gly	GGA	89.00	18.37	Trp	TGG	50.00	10.32
Gly	GGT	73.00	15.07	End	TGA	7.00	1.44
Gly	GGC	115.00	23.74	Cys	TGT	27.00	5.57
				Cys	TGC	71.00	14.65
Glu	GAG	219.00	45.20	End	TAG	4.00	0.83
Glu	GAA	122.00	25.18	End	TAA	7.00	1.44
				Tyr	TAT	43.00	8.88
Asp	GAT	148.00	30.55	Tyr	TAC	85.00	17.54
Asp	GAC	118.00	24.36				
Val	GTG	146.00	30.13	Ser	TCG	61.00	12.59
Val	GTA	37.00	7.64	Ser	TCA	58.00	11.97
Val	GTT	128.00	26.42	Ser	TCT	51.00	10.53
Val	GTC	78.00	16.10	Ser	TCC	61.00	12.59
				Arg	CGG	35.00	7.22
Ala	GCG	105.00	21.67	Arg	CGA	19.00	3.92
Ala	GCA	89.00	18.37	Arg	CGT	25.00	5.16
Ala	GCT	141.00	29.10	Arg	CGC	25.00	5.16
Ala	GCC	119.00	24.56				
Ser	AGT	34.00	7.02	Gln	CAG	77.00	15.89
Ser	AGC	67.00	13.83	Gln	CAA	41.00	8.46
				His	CAT	35.00	7.22
Lys	AAG	203.00	41.90	His	CAC	48.00	9.91
Lys	AAA	99.00	20.43				
Asn	AAT	87.00	17.96	Leu	CTG	93.00	19.20
Asn	AAC	95.00	19.61	Leu	CTA	40.00	8.26
				Leu	CTT	70.00	14.45
Met	ATG	112.00	23.12	Leu	CTC	112.00	23.12
Ile	ATA	46.00	9.49				
Ile	ATT	91.00	18.78	Pro	CCG	77.00	15.89
Ile	ATC	161.00	33.23	Pro	CCA	55.00	11.35
				Pro	CCT	65.00	13.42
Thr	ACG	57.00	11.76	Pro	CCC	73.00	15.07
Thr	ACA	51.00	10.53				
Thr	ACT	60.00	12.38				
Thr	ACC	68.00	14.04				
Leu	TTA	31.00	6.40				
Phe	TTT	59.00	12.18				
Phe	TTC	118.00	24.36				

H. ANEXO: Secuencias de transcripción inversa

Figura H-1. LSBo-I secuencia de nt más probable usando codones para *S miltiorrhiza*

```
atggctaccctcctccagaacctcctcaccgctaccagcagannnnnacccctcctcgatctcgataaacaccagannnagc
nnnacccctcctcaccaccccgaacttcaacggcgatctcggccacggccagcagcgccagcggccagagcccggagc
ccgagcagattcnnngctctcaccggcgctctcgcctcagaaggtgggaccagcgtgcacccgaacccgagcagcaacctc
gagggcgccgatgctttcgagttcgtgggcccgaagnnnttcnntatcaccnnnnnnnnnnnnnnnnnnngctgatggc
nnngctnnnttcnnnnnccgggtggcnnnnnnnnnccgagcggcagcaccggcgctacctcgcgctgtgaaccgaag
ggcnnngcnnnnnnnncagggcgctggatgctgatctcaacggcnnngcagctacggcagcggcgccggcggaaccac
ggcttctgcccagcggcagcagcagcaacagcagctcgtgcagagcagcggcctcggccaccagagagcagannnnc
aactacnnnnnnnnnnntacgagagctccaggtgaccgctggcnnnagaaagnnnnnnnnnnnagctacgaggt
gagctcgatgatttcgtgacctccaggtgaccgctggcctcagcgtgctaccagcggcgatggcgtggagaaggatctc
nnnnntggnnttaccgctnnnatgnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnngcnnnnnnatc
agacagnnngtg
```

Figura H-2. LSBo-I secuencia ambigua derivada de todos los codones posibles para cada aminoácido

```
atggcnacnytnytnrcaraayytnytnacngcncnacswnmgnnnnnnacnytnytnyngayyngayaayaacnmgnnnnwns
nnnacnytnytnacncayccnaayttaaayggngayynggncayggncarcayggngwsnggncarwsnccnwns
ccnwsnmgnttynnngcnytnacngngcnytnngcncaraargtnggnacnwsngtncayccnaaycncnwsnwsnaayytn
gargngngngaygcnttygarttygtnggncnaarnnttynnnathacnnnnnnnnnnnnnnnnnnngcngaygg
nnngcnnnttynnnnnccngtngggnnnnnnnnccnwsnggngwsnaccngnggntayynggngcngtncayccnaar
ggnnngggnnnnnnnncargngtngaygcnayytnaayggnnnggwsntayggngwsnggnggnggngnaaycay
ggnttygygngwsnggngwsnwsnaaywsngarytngtncarwsnwsnggnytnngncaycargarwsnmgnnnnath
aaytaynnnnnnnnnnnnntaygargarytnrcargtnacngcngggnnnnmgnaarnnnnnnnnnnnwsntaygargcn
garytnyngayytnytnacnytnrcargtnacngtnggnytnwsngcngcncnacswnsgngngayggngtngaraargayath
nnnnntggnnttyacngcnnnatgnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnngggnnnnnnath
mgncarnngtn
```

Figura H-3. Secuencia empleando la tabla de codones para *Arabidopsis thaliana*:

```
TGGCTACTCTTCTTCAAAATCTTCTTACTGCTACTTCTAGANNNNNACTCTTCTTGATCTTGATAATACTAGANNNTCTN
NNACTCTTCTTACTCATCCTAATTTAATGGAGATCTTGGACATGGACAACATGGATCTGGATCTGGACAATCTCCTTCTC
CTTCTAGATTTNNGCTCTTACTGGAGCTCTTGCTCAAAGGTTGGAACCTCTGTTTCATCCTAATCCTTCTTCTAATCTTG
AAGGAGGAGATGCTTTTGAATTTGTTGGACCTAAGNNNTTTNNNATTACTNNNNNNNNNNNNNNNNNNNGCTGATGGAN
NNGCTNNNTTTNNNNNNCCTGTTGGANNNNNNNNCCTTCTGGATCTACTGGAGGATATCTTGGAGCTGTTAATCCTAAGG
GANNNGANNNNNNNNCAAGGAGTTGATGCTGATCTTAATGGANNNGGATCTTATGGATCTGGAGGAGGAGGAAATCATG
GATTTTGTGGATCTGGATCTTCTTCTAATTTCTGAACCTGTTCAATCTTCTGGACTTGACATCAAGAATCTAGANNNATTA
ATTATNNNNNNNNNNNNNTATGAAGAACTTCAAGTTACTGCTGGANNNAGAAAGNNNNNNNNNNNTCTTATGAAGCTG
AACTTGATGATTTTGTACTCTTCAAGTTACTGTTGGACTTTCTGCTGCTACTTCTGGAGATGGAGTTGAAAAGGATATTN
NNNNNTGNNNTTTACTGCTNNNATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGANNNNNNATTA
GACAANNNGTT
```

Figura H-4. Secuencia empleando la tabla de codones para *Oryza sativa*:

```
ATGGCCACCTCCTCCAGAACCTCCTCACC GCCACCTCCCGC NNNNNNACCCTCCTCGACCTCGACAACACCCGCNNNTCC
NNNACCTCCTCACCACCCGAACCTTCAACGGCGACCTCGGCCACGGCCAGCACGGCTCCGGCTCCGGCCAGTCCCCGTC
CCGTCCTCCGCTTCNNGGCCCTCACC GGCGCCCTCGCC CAGAAGTGGGCACCTCCGTGCACCCGAACCCGTCTCCAACCTC
GAGGGCGGCAGCCCTCGAGTTCGTGGGCCCCGAAGNNNTTCNNTATCACCNNNNNNNNNNNNNNNNNNNGCCGACGGC
NNNGCCNNNTTCNNNNNNCCGGTGGGCNNNNNNNNCCGTCCGGCTCCACCGCGGCTACCTCGGCGCCGTGAACCCGAAG
GGC NNNGGC NNNNNNNN CAGGGCGTGGACGCCACCTCAACGGC NNNGGCTCCTACGGCTCCGGCGGGCGGCAACCAC
GGCTTCTGCGGCTCCGGCTCCTCCTCAACTCCGAGCTCGTGCAGTCTCCGGCTCCGGCCACCAGGAGTCCCGC NNNATC
AACTACNNNNNNNNNNNNNTACGAGGAGCTCCAGGTGACCGCCG C NNNCGCAAGNNNNNNNNNNNTCTACGAGGCC
GAGCTCGACGACTTCGTGACCTCCAGGTGACCGTGGGCTCTCCGCCGCCACCTCCGGCGACGGCTGGAGAAGGACATC
NNNNNTGNNNTTCAACGCCNNATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCNNNNNNATC
CGCCAGNNNGTG
```

Figura H-5. Protein Sequence Back-translation con el algoritmo *Backtranambig*.

ATGGCNACNYTNYTNCARAAAYTNYTACNCGCNACNWSNMGNNNNNNNACNYTNYTNGAYYTNGAYAAAYACNMGNNNNWSN
NNNACNYTNYTACNCAAYCCNAAYTTYAAAYGGNGAYYTNGGNCAAYGGNCARCAAYGGNWSNGGNWSNGGNCARWSNCCNWSN
CCNWSNMGNTTYNNNGCNYTACNCGGNGCNYTNGCNCARAARGTNGGNACNWSNGTNCAYCCNAAYCCNWSNWSNAAYYT
GARGGNGGNGAYGCNTTYGARTTYGTNGGNCCNAARNNTTYNNNATHACNNNNNNNNNNNNNNNNNNNNNGCNGAYGGN
NNNGCNNNTTYNNNNNNCCNGTNGGNNNNNNNNNCCNWSNGGNWSNACNCGGNGNTAYYTNGGNCGNGTNAAYCCNAAR
GGNNNGGNNNNNNNNNCARGGNGTNGAYGCNGAYYTNAAYGGNNNGGNWSNTAYGGNWSNGGNGGNGGNGGNAAYCAY
GGNTTYTG YGGNWSNGGNWSNWSNWSNAAYWSNGARYTNGTNCARWSNWSNGGNYTNGGNCA YCARGARWSNMGNNNNATH
AAYTAYNNNNNNNNNNNNNTAYGARGARYTNCARGTACNGCNGGNNNNMGNAARNNNNNNNNNNNWSNTAYGARGCN
GARYTNGAYGAYTTYGTNACNYTNCARGTACNGTNGGNYTNWSNGCNGCNACNWSNGGNAYGGNGTNGARAARGAYATH
NNNNNTGGNNNTTYACNGCNNNATGNNGGNNNNNNATH
MGNCARNNNGTN

I. ANEXO: Pares de primers obtenidos mediante la herramienta bioinformática PrimerBlast.

Tabla I-1. Pares de primers obtenidos mediante la herramienta bioinformática PrimerBlast.

Primer pair 1									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CTCACCCACCCGAACTCAA	Plus	20	91	110	59.89	55.00	5.00	3.00
Reverse primer	CACGAAGCTCGAAAGCATCG	Minus	20	270	251	60.25	55.00	4.00	2.00
Product length	180								
Primer pair 2									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CGAGGCTGAGCTCGATGATT	Plus	20	642	661	59.97	55.00	7.00	1.00
Reverse primer	TATCCTTCTCCACGCCATCG	Minus	20	727	708	59.33	55.00	2.00	2.00
Product length	86								
Primer pair 3									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TGGATGCTGATCTCAACGGC	Plus	20	431	450	60.46	55.00	4.00	2.00
Reverse primer	TCACCTGGAGGGTCACGAAA	Minus	20	679	660	60.76	55.00	4.00	0.00
Product length	249								
Primer pair 4									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GGCGTGGATGCTGATCTCAA	Plus	20	427	446	60.46	55.00	4.00	2.00
Reverse primer	GGGTACAGAAATCATCGAGC	Minus	20	670	651	59.07	55.00	4.00	2.00
Product length	244								
Primer pair 5									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CCGAACCTCAACGGCGATCT	Plus	20	100	119	60.74	55.00	5.00	2.00
Reverse primer	CCCACGAAGCTCGAAAGCATC	Minus	20	272	253	59.28	55.00	4.00	0.00
Product length	173								
Primer pair 6									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	ATTCGTGACCCCTCCAGGTG	Plus	20	659	678	59.68	55.00	4.00	3.00
Reverse primer	ATATCCTTCTCCACGCCATCG	Minus	21	728	708	59.73	52.38	4.00	2.00
Product length	70								
Primer pair 7									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	CCAGAACCTCCTCACCCGCTA	Plus	20	15	34	60.97	60.00	2.00	2.00
Reverse primer	AGATCGCCGTTGAAGTTCGG	Minus	20	119	100	60.74	55.00	5.00	1.00
Product length	105								
Primer pair 8									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	GATGCTTTCGAGTTCGTGGG	Plus	20	253	272	59.28	55.00	4.00	0.00
Reverse primer	CCTGGAGGGTCACGAAATCAT	Minus	21	676	656	59.79	52.38	4.00	3.00
Product length	424								
Primer pair 9									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	AACCTCCTCACCCGCTACCTC	Plus	20	19	38	60.97	60.00	2.00	0.00
Reverse primer	TGGAGGGTCACGAAATCATCG	Minus	21	674	654	60.13	52.38	3.00	2.00
Product length	656								
Primer pair 10									
	Sequence (5'→3')	Template strand	Length	Start	Stop	Tm	GC%	Self complementarity	Self 3' complementarity
Forward primer	TCCTCACCCACCCGAACTT	Plus	19	89	107	60.46	57.89	3.00	3.00
Reverse primer	GGAGGGTCACGAAATCATCG	Minus	20	673	654	58.43	55.00	3.00	2.00
Product length	585								

J. ANEXO. Validación de todos los primers diseñados con el programa PCR Primer Stats

Figura J-1. Resultados obtenidos de la validación de todos los primers diseñados con el programa PCR Primer Stats

Global settings:

- The primers do not have a 5'-phosphate group.
- Combined concentration of K⁺ and Na⁺ in the reaction = 50 millimolar.
- Mg²⁺ concentration in the reaction = 1.5 millimolar.
- Primer concentration in the reaction = 200 nanomolar.

General properties:

Primer name: **PBSb F1**
Primer sequence: CTCACCCACCCGAACTTCAA
Sequence length: 20
Base counts: G=1; A=6; T=3; C=10; Other=0;
GC content (%): 55.00
Molecular weight (Daltons): 5950.91
nmol/A260: 5.49
micrograms/A260: 32.70
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 65.25

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Pass
Hairpin formation: Pass

General properties:

Primer name: **PBSb R1**
Primer sequence: CACGAACTCGAAAGCATCGC
Sequence length: 20
Base counts: G=4; A=7; T=2; C=7; Other=0;
GC content (%): 55.00
Molecular weight (Daltons): 6080.01
nmol/A260: 5.16
micrograms/A260: 31.37
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 64.70

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Pass
Hairpin formation: Pass

General properties:

Primer name: PBSb F4
Primer sequence: GGCGTGGATGCTGATCTCAA
Sequence length: 20
Base counts: G=7; A=4; T=5; C=4; Other=0;
GC content (%): 55.00
Molecular weight (Daltons): 6173.07
nmol/A260: 5.19
micrograms/A260: 32.07
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 65.13

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Pass
Hairpin formation: Pass

General properties:

Primer name: PBSb R4
Primer sequence: GGGTCACGAAATCATCGAGC
Sequence length: 20
Base counts: G=6; A=6; T=3; C=5; Other=0;
GC content (%): 55.00
Molecular weight (Daltons): 6151.06
nmol/A260: 5.02
micrograms/A260: 30.88
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 63.93

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
 Length: Pass
 Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
 GC clamp: Warning: There are more than 3 G's or
C's in the last 5 bases;
 Self-annealing: Pass
 Hairpin formation: Pass

General properties:

Primer name: PBSb F6
Primer sequence: ATTTTCGTGACCCTCCAGGTG
Sequence length: 20
Base counts: G=5; A=3; T=6; C=6; Other=0;
GC content (%): 55.00
Molecular weight (Daltons): 6084.00
 nmol/A260: 5.42
 micrograms/A260: 32.96
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 65.05

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
 Length: Pass
 Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
 GC clamp: Pass
 Self-annealing: Pass
 Hairpin formation: Pass

General properties:

Primer name: PBSb R6
Primer sequence: ATATCCTTCTCCACGCCATCG
Sequence length: 21
Base counts: G=2; A=4; T=6; C=9; Other=0;
GC content (%): 52.38
Molecular weight (Daltons): 6277.12
 nmol/A260: 5.31
 micrograms/A260: 33.34
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 64.62

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass

Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Warning: There are more than 3 self-
annealing bases in a row;
:
:
: GCTACCGCACCTCTTCCTATA
Hairpin formation: Pass

General properties:

Primer name: PBSb F7
Primer sequence: CCAGAACCTCCTCACCCTA
Sequence length: 20
Base counts: G=2; A=5; T=3; C=10; Other=0;
GC content (%): 60.00
Molecular weight (Daltons): 5966.91
nmol/A260: 5.54
micrograms/A260: 33.06
Basic Tm (degrees C): 56
Salt adjusted Tm (degrees C): 51
Nearest neighbor Tm (degrees C): 66.04

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Pass
Hairpin formation: Pass

General properties:

Primer name: PBSb R7
Primer sequence: AGATCGCCGTTGAAGTTCGG
Sequence length: 20
Base counts: G=7; A=4; T=5; C=4; Other=0;
GC content (%): 55.00
Molecular weight (Daltons): 6173.07
nmol/A260: 5.13
micrograms/A260: 31.64
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 65.64

PCR suitability tests (Pass / Warning):

Single base runs: Pass

```
Dinucleotide base runs: Pass
      Length: Pass
      Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
      GC clamp: Pass
      Self-annealing: Warning: There are more than 3 self-
annealing bases in a row;
      :                AGATCGCCGTTGAAGTTCGG
      :                ||||
      :                GGCTTGAAGTTGCCGCTAGA
Hairpin formation: Pass
```

General properties:

```
Primer name: PBSb F8
Primer sequence: GATGCTTTCGAGTTCGTGGG
Sequence length: 20
Base counts: G=8; A=2; T=7; C=3; Other=0;
GC content (%): 55.00
Molecular weight (Daltons): 6195.08
      nmol/A260: 5.32
      micrograms/A260: 32.94
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 64.14
```

PCR suitability tests (Pass / Warning):

```
Single base runs: Pass
Dinucleotide base runs: Pass
      Length: Pass
      Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
      GC clamp: Warning: There are more than 3 G's or
C's in the last 5 bases;
      Self-annealing: Pass
      Hairpin formation: Pass
```

General properties:

```
Primer name: PBSb R8
Primer sequence: CCTGGAGGGTCACGAAATCAT
Sequence length: 21
Base counts: G=6; A=6; T=4; C=5; Other=0;
GC content (%): 52.38
Molecular weight (Daltons): 6455.26
      nmol/A260: 4.85
      micrograms/A260: 31.29
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 65.05
```

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
 Length: Pass
 Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
 GC clamp: Pass
 Self-annealing: Pass
 Hairpin formation: Pass

General properties:

Primer name: PBSb F9
Primer sequence: AACCTCCTCACCGCTACCTC
Sequence length: 20
Base counts: G=1; A=4; T=4; C=11; Other=0;
GC content (%): 60.00
Molecular weight (Daltons): 5917.87
 nmol/A260: 5.80
 micrograms/A260: 34.31
Basic Tm (degrees C): 56
Salt adjusted Tm (degrees C): 51
Nearest neighbor Tm (degrees C): 66.04

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
 Length: Pass
 Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
 GC clamp: Pass
 Self-annealing: Pass
 Hairpin formation: Pass

General properties:

Primer name: PBSb R9
Primer sequence: TGGAGGGTCACGAAATCATCG
Sequence length: 21
Base counts: G=7; A=6; T=4; C=4; Other=0;
GC content (%): 52.38
Molecular weight (Daltons): 6495.29
 nmol/A260: 4.75
 micrograms/A260: 30.83
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 65.31

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass

Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Pass
Hairpin formation: Pass

General properties:

Primer name: SmF1
Primer sequence: ATGGCCAAGCTTCTCCAAAAC
Sequence length: 21
Base counts: G=3; A=7; T=4; C=7; Other=0;
GC content (%): 47.62
Molecular weight (Daltons): 6359.20
nmol/A260: 5.01
micrograms/A260: 31.88
Basic Tm (degrees C): 52
Salt adjusted Tm (degrees C): 47
Nearest neighbor Tm (degrees C): 64.01

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Warning: There are more than 3 self-annealing bases in a row; More than 50 percent of the bases are self-annealing;

: ATGGCCAAGCTTCTCCAAAAC
: ||| ||||| |||
: CAAAACCTCTTCGAACCGGTA

Hairpin formation: Pass

General properties:

Primer name: SmR1
Primer sequence: GTCGATCGCTTAGTCCTTATTGA
Sequence length: 23
Base counts: G=5; A=4; T=9; C=5; Other=0;
GC content (%): 43.48
Molecular weight (Daltons): 7020.63
nmol/A260: 4.61
micrograms/A260: 32.37
Basic Tm (degrees C): 53
Salt adjusted Tm (degrees C): 48
Nearest neighbor Tm (degrees C): 62.78

PCR suitability tests (Pass / Warning):

```

      Single base runs: Pass
    Dinucleotide base runs: Pass
          Length: Pass
        Percent GC: Pass
      Tm (Nearest neighbor): Warning: Tm is greater than 58;
          GC clamp: Pass
        Self-annealing: Warning: There are more than 3 self-
annealing bases in a row;
          :                               GTCGATCGCTTAGTCCTTATTGA
          :                               ||||
          : AGTTATTCTGATTTCGCTAGCTG
      Hairpin formation: Pass

```

General properties:

```

      Primer name: SmF2
    Primer sequence: CATGACATCGTCTCGTGGTATTTTC
    Sequence length: 24
      Base counts: G=5; A=4; T=9; C=6; Other=0;
      GC content (%): 45.83
    Molecular weight (Daltons): 7309.81
          nmol/A260: 4.49
        micrograms/A260: 32.81
      Basic Tm (degrees C): 56
    Salt adjusted Tm (degrees C): 51
    Nearest neighbor Tm (degrees C): 64.55

```

PCR suitability tests (Pass / Warning):

```

      Single base runs: Pass
    Dinucleotide base runs: Pass
          Length: Pass
        Percent GC: Pass
      Tm (Nearest neighbor): Warning: Tm is greater than 58;
          GC clamp: Pass
        Self-annealing: Warning: There are more than 3 self-
annealing bases in a row;
          :
CATGACATCGTCTCGTGGTATTTTC
          :                               ||||
          : CTTTATGGTGCTCTGCTACAGTAC
      Hairpin formation: Pass

```

General properties:

```

      Primer name: SmR2
    Primer sequence: GATCGCTTAGTCCTTATTGATTTGC
    Sequence length: 25
      Base counts: G=5; A=4; T=11; C=5; Other=0;
      GC content (%): 40.00
    Molecular weight (Daltons): 7629.03
          nmol/A260: 4.34
        micrograms/A260: 33.13

```


nmol/A260: 4.76
micrograms/A260: 30.78
Basic Tm (degrees C): 54
Salt adjusted Tm (degrees C): 49
Nearest neighbor Tm (degrees C): 64.96

PCR suitability tests (Pass / Warning):

Single base runs: Pass
Dinucleotide base runs: Pass
Length: Pass
Percent GC: Pass
Tm (Nearest neighbor): Warning: Tm is greater than 58;
GC clamp: Pass
Self-annealing: Pass
Hairpin formation: Pass
