



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Diseño, modelamiento y evaluación *in silico* de péptidos que reconocen antígeno Tn

Daniel Armando Romero Frenchy

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá D.C., Colombia
2019

Diseño, modelamiento y evaluación *in silico* de péptidos que reconocen antígeno Tn

Daniel Armando Romero Frenchy

Tesis o trabajo de investigación presentada como requisito parcial para optar al título de:

Magister en Bioinformática

Director:

MSc., Dr.Sci. Edgar Antonio Reyes Montaña

Línea de Investigación:

Línea de investigación en Lectinas Vegetales.

Grupo de Investigación:

Grupo de Investigación en Proteínas

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería de Sistemas

Bogotá D.C., Colombia

2019

A mi familia, amigos y compañeros, por su apoyo, tiempo y comprensión.

Agradecimientos

Me gustaría agradecer en estas líneas la ayuda que muchas personas y colegas me han prestado durante el proceso de investigación y redacción de este trabajo. En primer lugar, quisiera agradecer a mis padres Aura Alicia Frenchy y Roberto Armando Romero que me han ayudado y apoyado en todo los proyectos en los que me embarco, a mis hermanos Juan Carlos Romero y Diego Fernando Romero por todos los ratos de ocio y buenos momentos que pasamos juntos. A mi abuela María Elsa Frenche y a mi tía Elsa Raquel Pulido y su familia por haberme orientado en todos los momentos que necesité sus consejos.

Así mismo, deseo expresar mi reconocimiento a mis compañeros del grupo de investigación en proteínas por su ayuda desinteresada y buena voluntad, y en especial a los profesores Nohora Vega y Edgar Reyes infinitas gracias no solo por los conocimientos proporcionados sino por su infinita paciencia para conmigo que permitieron llevar a buen terminino este proyecto de investigación.

A mis amigos y compañeros de trabajo que soportaron mis desplantes durante todo este tiempo.

A la Universidad Nacional de Colombia por ser la sede de todo el conocimiento adquirido en estos años.

Resumen

El antígeno Tn (GalNac α -Ser / Thr) fue descrito por primera vez en pacientes con una rara enfermedad hemolítica (síndrome Tn), y desde que se descubrió su presencia en el 90 % de los carcinomas a mediados de la década de 1980 ha sido objeto de muchos estudios ya que se ha encontrado que el mismo, y el mismo se asocia con cánceres más agresivos, además se encuentra en estadios tempranos de la enfermedad en diferentes carcinomas siendo un marcador alternativo para el diagnóstico temprano de los mismos. En el presente estudio se utilizaron varias estructuras primarias de lectinas vegetales, así como las estructuras terciarias de la Isolectina B4 de *Vicia villosa* (PDB 1N47), las lectinas de *Vatairea macrocarpa* (PDB 4XTP), *Psophocarpus tetragonolobus* (PDB 2D3S), *Glycine max* (PDB 4D69) y *Bauhinia fortificata* (PDB 5T5J) para identificar los aminoácidos involucrados en el reconocimiento del antígeno Tn, luego con base en estos resultados se diseñaron y modelaron péptidos los cuales fueron analizados por medio de pruebas de docking molecular utilizando el algoritmo de AutoDock VINA, con el fin de identificar los péptidos más promisorios y ser utilizados en ensayos *in vitro*.

Palabras clave: Lectinas vegetales, Antígeno Tn, Docking molecular, Diseño de péptidos.

Abstract

The Tn antigen (GalNac α -Ser / Thr) was first described in patients with a rare hemolytic anemia (Tn syndrome) and since its presence was discovered in 90% of human carcinomas in the mid-1980s has been the subject of an extensive study, it has been found that a high detection of this antigen is associated with more aggressive cancers, and it appears in the early stages of different carcinomas being an alternative for the early diagnosis of these group of disease. Currently, lectins such as B4 isolectin from *Vicia villosa*, A4 isolectin from *Griffonia simplicifolia*, as well as lectins from *Salvia sclarea* and *Moluccella laevis* among others, have been used for the identification of the Tn antigen. In the present study we use the amino acids sequences of several legume lectins, as well the three-dimensional structures of *Vicia villosa* (PDB 1N47), *Vatairea macrocarpa* (PDB 4XTP), *Psophocarpus tetragonolobus* (PDB 2D3S), *Glycine max* (PDB 4D69) and *Bauhinia fortificata* (PDB 5T5J) to identify the amino acids presented in the Interaction of these lectins with the Tn antigen, then based on the results obtained we design and modeling peptides that were evaluated using molecular docking tests, using the Autodock VINA algorithm, and identify the more promising peptides to be used in *in vivo* essays.

Keywords: Legume lectins, Tn. Antigen, Molecular Docking, Peptide design.

Contenido

	Pág.
Resumen	IX
Abstract.....	X
Lista de figuras	XIII
Lista de tablas	XIV
Lista de Símbolos y abreviaturas	1
Introducción	2
1. Capítulo 1: Análisis de la estructura primaria y terciaria de las lectinas y su interacción con el antígeno Tn.....	8
1.1 Lectinas.....	8
1.1.1 Estudio de las lectinas en Colombia	9
1.1.2 Ingeniería de Lectinas	10
1.1.3 Lectinas que reconocen el antígeno Tn.....	11
1.2 Metodología	12
1.2.1 Búsqueda bibliográfica y de estructuras primarias y terciarias.	12
1.2.2 Análisis de la estructura primaria, secundaria y terciaria e interacción con el antígeno Tn de las proteínas encontradas.	12
1.3 Resultados y discusión.....	13
1.3.1 Búsqueda de proteínas capaces de reconocer el antígeno Tn.....	13
1.3.2 Análisis de la estructura primaria	13
1.3.3 Análisis de la estructura secundaria	15
1.3.4 Análisis de estructura terciaria.	16
2. Capítulo 2: Diseño de péptidos y modelamiento de los péptidos.....	19
2.1 Diseño racional de compuestos.	19
2.1.1 Diseño computacional de proteínas.....	19
2.1.2 Diseño computacional de péptidos.	20
2.2 Metodología	22
2.2.1 Diseño y modelación de los péptidos.	22
2.2.2 Creación de la librería de péptidos.	22
2.2.3 Modelamiento de los péptidos.	23
2.3 Resultados y discusión.....	23
2.3.1 Diseño y modelamiento de los péptidos con base en lectinas leguminoides.	23
2.3.2 Características fisicoquímicas de los péptidos modelados.	26

3. Capítulo 3: Pruebas de Docking molecular.....	29
3.1 Docking molecular.....	29
3.1.1 Docking comparativo.....	29
3.1.2 Docking Local.....	30
3.1.3 Docking global.....	30
3.1.4 Dificultades de las pruebas de docking molecular.....	30
3.2 Metodología.....	31
3.2.1 Preparación de los archivos para la prueba de Docking molecular.....	31
3.2.2 Análisis de los datos de las pruebas de Docking molecular.....	32
3.2.3 Análisis de las conformaciones de los péptidos seleccionados.....	32
3.3 Resultados y discusión.....	33
3.3.1 Análisis de datos de Docking molecular.....	33
3.3.2 Análisis de las interacciones.....	34
3.3.3 Desempeño péptido DARF51.....	36
3.3.4 Desempeño péptido DARF53.....	37
3.3.5 Desempeño péptido DARF62.....	37
3.3.6 Desempeño péptido DARF65.....	37
3.3.7 Desempeño péptido DARF68.....	38
3.3.8 Desempeño péptido DARF80.....	38
3.3.9 Desempeño péptido DARF3.....	39
4. Conclusiones.....	41
4.1 Recomendaciones.....	41
A. Anexo: Scrip Perl mutagenesis.....	43
B. Anexo: Conformaciones de los péptidos mas promisorios.....	46
Bibliografía.....	47

Lista de figuras

Figura 4-1: Síntesis del antígeno Tn: En la Figura se observa a la izquierda la síntesis de la glicoproteína normal. A la derecha se observa la síntesis del antígeno Tn y del antígeno Tn Sialilado. Modificado de R.D. Cummings y colaboradores 2011 (3).	3
Figura 1-1: Clasificación de las lectinas vegetales: En el diagrama se observan los tres grandes grupos en los cuales se clasifican las lectinas vegetales. Modificado de C.S.F. Bath y colaboradores 2013 (31).....	9
Figura 1-2: Alineamiento múltiple de las secuencias de las lectinas: En rojo se encuentran los residuos que interactúan con el antígeno Tn, en azul los aminoácidos que reconocen el ion Mn^{2+} , en verde los residuos que reconocen el ion Ca^{2+} y en púrpura los residuos con anillo aromático en su cadena lateral que estabilizan el anillo alifático. Las flechas verdes indican las hileras de hebras β , y los puntos rojos indican los giros en la estructura.	15
Figura 1-3: Análisis de NMA muestra que las cinco estructuras se comportan de manera similar.....	16
Figura 1-4: La superposición de las estructuras confirma lo conservada que es esta superfamilia de proteínas.....	17
Figura 1-5: Interacciones entre las diferentes lectinas y el antígeno Tn. Diagrama de interacción de las lectinas de <i>Vicia villosa</i> A, <i>Vatairea macrocarpa</i> B y <i>Psophocarpus tetragonolobus</i> C.....	18
Figura 2-1: Péptido DARF1 y péptido DAR51. El primero adquiere una estructura tipo random coil, mientras que el segundo adquiere una estructura similar a una hebra β	28
Figura 3-1: Promedio de los péptidos evaluados, se observa que solo los péptidos creados a partir de 17 residuos lograron superar el umbral de energía libre establecido (línea punteada roja).....	33
Figura 3-2: Interacción entre el péptido DARF65 y el antígeno Tn. A la derecha se observan los diferentes residuos que interactúan con el anillo de piranosa de la estructura y a la izquierda como el antígeno se acopla al bolsillo del péptido.	38
Figura 3-3: Interacción entre el péptido DARF80 y el antígeno Tn. A la derecha se observan los diferentes residuos que interactúan con el anillo de piranosa de la estructura y a la izquierda como el antígeno se acopla al bolsillo del péptido.	39
Figura 3-4: Interacción entre el péptido DARF3 y el antígeno Tn. A la derecha se observan los diferentes residuos que interactúan con el anillo de piranosa de la estructura y a la izquierda como el antígeno se acopla al bolsillo del péptido.	40

Lista de tablas

	Pág.
Tabla 1-1: Lectinas vegetales que reconocen el antígeno Tn. En negrilla se encuentra las lectinas pertenecientes a la familia de las leguminosas.	14
Tabla 2-1: Péptidos diseñados a partir de 20 aminoácidos.	24
Tabla 2-2: Péptidos diseñados a partir de 17 aminoácidos	25
Tabla 2-3: RMSD de cada uno de los péptidos seleccionados	25
Tabla 2-4: Características fisicoquímicas de los péptidos diseñados a partir de 20 aminoácidos	26
Tabla 2-5: Características fisicoquímicas de los péptidos diseñados a partir de 17 aminoácidos	27
Tabla 3-1: Mejores conformaciones de los péptidos analizados.	34

Lista de Símbolos y abreviaturas

Abreviaturas

Abreviatura	Término
PDB	Protein data base
GalNAc	N-Acetilgalactosamina
Ala, A	Alanina
Arg, R	Arginina
Asn, N	Asparagina
Asp, D	Ácido aspártico
Cys, C	Cisteina
Phe, F	Fenialanina
Gly, G	Glicina
Glu, E	Ácido glutámico
Gln, Q	Glutamina
His, H	Histidina
Ile, I	Isoleucina
Leu, L	Leucina
Lys, K	Lisina
Met, M	Metionina
Pro, P	Prolina
Ser, S	Serina
Tyr, Y	Tirosina
Thr, T	Treonina
Trp, W	Triptofano
Val, V	Valina
-NH ₂	Grupo amina
-NH ₃	Grupo amino
-COOH	Grupo carboxilo
-CH ₃	Grupo metil
-SH	Grupo tiol
Å	Amstrong

Introducción

El antígeno Tn fue descrito hace alrededor de unos 50 años en los eritrocitos de algunos pacientes que presentaban distintos grados de aglutinación en presencia de suero de humanos o en presencia de algunas lectinas de origen vegetal (1). También se encuentra en un gran número de carcinomas siendo especialmente predominante en carcinomas de ovario, seno y algunos relacionados con el tracto digestivo (1).

La estructura del mismo fue definida por Dahr y colaboradores., en 1975 por medio de cromatografía líquida (2), y la misma es sintetizada por un grupo de enzimas denominadas N-acetilgalactosiltransferasas, las cuales transfieren el carbohidrato N-Acetilgalactosamina (GalNAc) por medio de un donante UDP GalNAc a un residuo de serina o treonina, siendo este el primer paso para la síntesis de las mismas (3). La enzima que modifica estas estructuras cortas en tejidos normales se denomina β 3 Galactosil transferasa o T sintetasa, la cual transfiere galactosa desde el donante UDP Gal en presencia de cationes divalentes a la estructura GalNAc α 1-O-Ser/Thr formando un enlace β -glicosídico en la estructura primaria (3). Se ha demostrado que la deficiencia en la proteína chaperona Core 1 β 3 galactosil transferasa o COSMC, esta relacionada en la formación del antígeno Tn, ya que participa activamente en el plegamiento de la enzima T sintetasa activa Figura 4-1 (3).

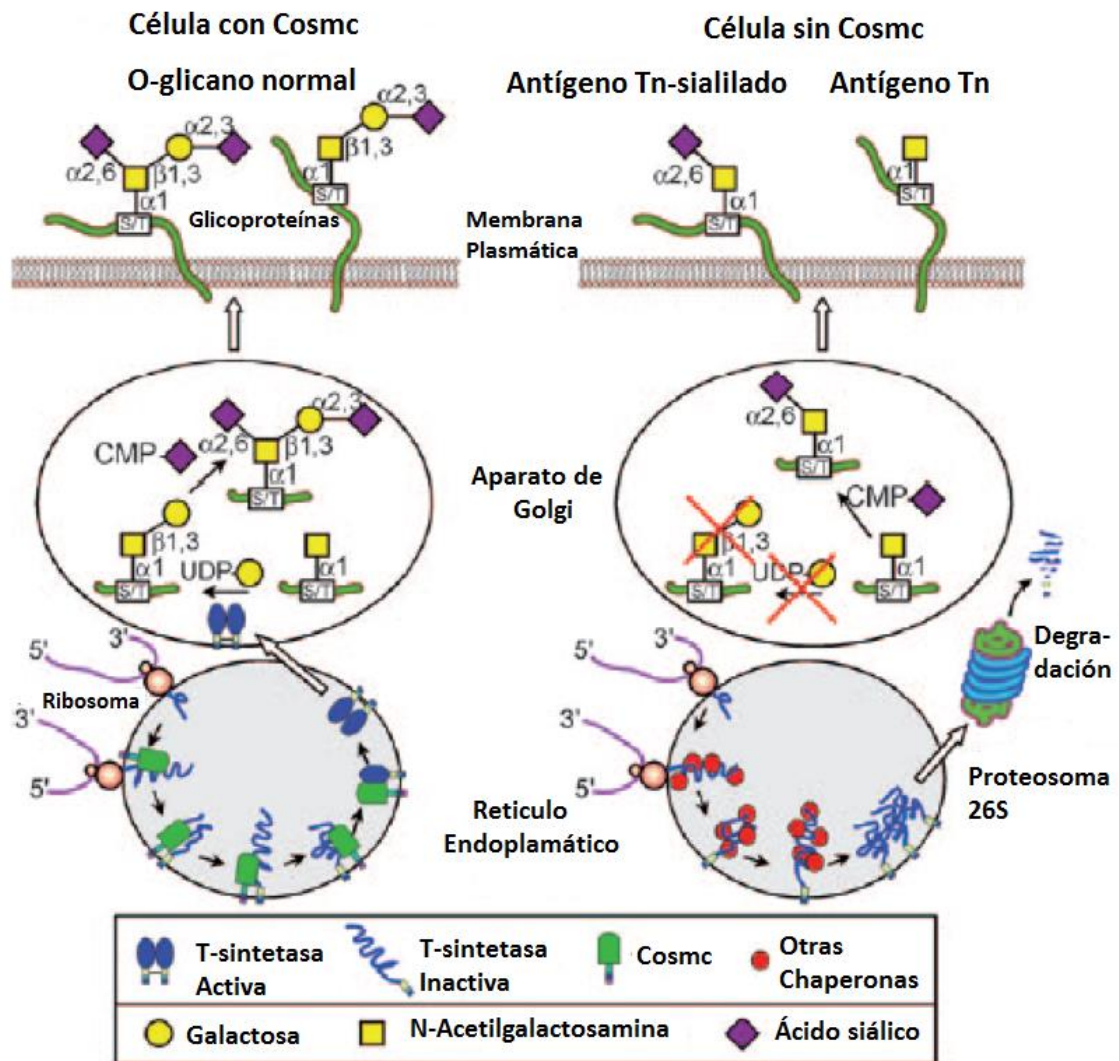


Figura 4-1: Síntesis del antígeno Tn: En la Figura se observa a la izquierda la síntesis de la glicoproteína normal. A la derecha se observa la síntesis del antígeno Tn y del antígeno Tn Sialilado. Modificado de R.D. Cummings y colaboradores 2011 (3).

Debido a la importancia de este antígeno, se han venido buscando proteínas como las lectinas que puedan ser usadas, ya sea como método de identificación y diagnóstico o con fines terapéuticos. En este sentido trabajos como los realizados por Marth y Daxenbichler (4), o el elaborado por Ryder Smith y Rhodes (5), han demostrado que las mismas pueden llegar a tener un efecto en el ciclo celular; ya sea inhibiendo el mismo (4) o como un agente mitogénico para determinados grupos de células (5). Otra de las características que se le atribuyen a este grupo de proteínas, es la capacidad de aumentar la respuesta inmune

celular (6–8), debido a estas características, se han adelantado estudios en donde se utilizan lectinas vegetales para el desarrollo de medicamentos contra el cáncer (9).

Otro tipo de moléculas como los anticuerpos monoclonales y policlonales son capaces de identificar el antígeno, siendo uno de los más reconocidos el anticuerpo MLS128 el cual fue derivado de la inmunización de ratones con células de carcinoma de colon (10) y del cual se han realizado varios estudios en donde se describe su acción sobre células provenientes de carcinoma de colon (11,12) otro de los antígenos más estudiados es el 83D4 el cual fue producido al inmunizar ratones con células en solución de carcinoma de seno (13). Recientemente se han realizado estudios como el de *Kubota et al.*, en el cual produjeron anticuerpos anti-Tn para luego modificarlos con un dominio constante para así usarlos como método terapéutico (14). Por último, estudios como el que presento Osinaga y colaboradores han evidenciado que estos anticuerpos son más efectivos para identificar antígenos con dos o tres enlaces glicosídicos, mientras que las lectinas son más eficientes para identificar antígenos con un solo glicotopes (15).

Recientemente se han empezado a utilizar lectinas vegetales en varios estudios pre clínicos y clínicos como tratamiento para algunos tipos de cáncer las cuales han demostrado su eficiencia en la eliminación del tumor (16). Sin embargo, a la fecha, aun no se han iniciado estudios preclínicos con lectinas que puedan reconocer el antígenoTn; siendo la lectina Morniga G proveniente de *Morus nigra*, capaz de causar muerte celular en células de Jurkat una de las más estudiadas y promisorias en este campo (17). Otros estudios tratan de usar las lectinas vegetales para mejorar la especificidad de la terapia fotodinámica; usando la capacidad para reconocer carbohidratos de estas para que la sustancia fotosensibilizadora sea absorbida tan solo por células tumorales (18,19).

Justificación

El antígeno Tn (GalNac α -Ser/Thr) fue descrito por primera vez en pacientes con una rara anemia hemolítica (síndrome Tn) y desde que se descubrió su presencia en el 90% de los carcinomas humanos a mediados de la década de los 80, ha sido objeto de estudio (1,20). Se ha encontrado que una alta detección de este antígeno se asocia a cánceres más

agresivos y el mismo aparece en etapas tempranas de diferentes carcinomas siendo una alternativa para el diagnóstico temprano de los mismos (21).

En la actualidad, se han desarrollado diferentes métodos para la identificación de este antígeno, inicialmente con anticuerpos monoclonales y policlonales y más recientemente con el uso de lectinas como la isolectina B4 de *Vicia villosa*, la isolectina A4 de *Griffonia simplicifolia*, así como lectinas provenientes de *Salvia sclarea* y *Moluccella laevis* (22) entre otras, siendo estas más específicas para la identificación del antígeno Tn que los anticuerpos monoclonales (22,23). Actualmente, se está estudiado cómo estas proteínas pueden actuar sobre las diferentes etapas del ciclo celular, siendo una de las áreas de interés su influencia en la inducción de la muerte celular (4,5,24), también se están desarrollando estudios como el realizado por Medeiros y colaboradores. en el cual evalúa la capacidad citotóxica de una lectina aislada de hojas de *Myrcine coracea* frente a células que presentan el antígeno Tn (25).

A pesar de su gran utilidad en el reconocimiento del antígeno Tn, actualmente se han caracterizado muy pocas lectinas capaces de reconocer de forma específica el mismo (26,27); por otro lado, recientemente la investigación de anticuerpos capaces de identificar el antígeno Tn se ha centrado en la búsqueda de anticuerpos mas específicos para reconocer este glicotope, pero estos estudios aún se encuentran en etapas de investigación muy tempranas y su utilidad está por ser comprobada (28,29). Esto sumado al hecho que no hay estudios recientes que busquen generar alguna terapia encaminada al reconocimiento del antígeno Tn, utilizando lectinas o anticuerpos, hace necesario en pensar otras alternativas que sirvan para solucionar este problema. Razón por la cual, debido a su utilidad a la hora de reconocer este marcador tumoral, se propone utilizar las lectinas como modelo para la creación de péptidos sintéticos que ayuden a una identificación temprana del antígeno Tn con fines terapéuticos.

Planteamiento del problema

Debido a la expresión del antígeno Tn en un gran número de carcinomas (1), se hace necesaria la búsqueda de métodos que permitan la identificación del mismo, ya sea para su uso diagnóstico o para el diseño de nuevos tratamientos. Es por eso que las lectinas vegetales se presentan como una alternativa para la identificación del mismo.

En la actualidad se han caracterizado algunas lectinas capaces de identificar no solo el antígeno Tn sino también otros glicopéptidos. Por esta razón las mismas están siendo utilizadas como modelo para el diseño de proteínas capaces de identificar este tipo de moléculas (4,5,7,9,24,25). A pesar de estos avances, actualmente existen muy pocos estudios relacionados con el diseño de péptidos con base en lectinas que sea capaz de identificar específicamente el antígeno Tn.

En Colombia, los estudios se han encaminado en la búsqueda de lectinas capaces de identificar dicho antígeno (23,26,27) y no existen actualmente estudios que las utilicen como modelo para el diseño de péptidos sintéticos. Es por eso que se hace necesario explorar esta posibilidad con el fin de generar estrategias alternativas para la identificación del antígeno Tn y su posterior uso en biomedicina.

De lo anteriormente dicho se desprende la pregunta de investigación la cual se define como *¿Cuáles son los aminoácidos que intervienen en la identificación del antígeno Tn en las lectinas vegetales y cómo éstos pueden ser usados para el diseño de péptidos que sean capaces de identificar específicamente el antígeno para el desarrollo futuro de técnicas diagnósticas o posibles tratamientos?*

Objetivos

Objetivo general:

Diseñar diferentes péptidos por medio de herramientas bioinformáticas que sea capaces de reconocer específicamente el antígeno Tn teniendo como base las lectinas vegetales.

Objetivos específicos:

1. Identificar las diferentes lectinas vegetales que se encuentran reportadas que sean capaces de reconocer el antígeno Tn.
2. Reconocer las secuencias y estructuras que se encuentran relacionadas con la interacción del antígeno Tn en las diferentes lectinas vegetales
3. Diseñar péptidos que serán evaluados por medio de herramientas bioinformáticas como las pruebas de Docking molecular.
4. Proponer al menos dos péptidos que presenten una buena interacción *in silico* con el antígeno Tn.

1. Capítulo 1: Análisis de la estructura primaria y terciaria de las lectinas y su interacción con el antígeno Tn.

1.1 Lectinas

Las lectinas son un grupo de proteínas que tienen una alta afinidad por los carbohidratos y que han sido de gran utilidad en el campo de la biomedicina (30), las mismas son ubicuas y se encuentran distribuidas en una gran cantidad de organismos (31). Actualmente, las lectinas vegetales son el grupo de lectinas que más se ha estudiado, debido a que estas se pueden aislar en gran cantidad de semillas. Por otra parte, estas tienen muchas aplicaciones dentro de las cuales destaca la tipificación de grupos sanguíneos y la identificación de marcadores tumorales que se encuentran en la superficie de la membrana celular como el antígeno Tn (1).

Las lectinas vegetales pueden ser clasificadas de acuerdo a su estructura en:

- Merolectinas: las cuales se caracterizan por tener solo un dominio funcional de identificación de carbohidrato por lo que no producen un efecto aglutinador.
- Hololectinas: las cuales poseen dos dominios funcionales capaces de reconocer un carbohidrato en particular.
- Quimerolectinas: las cuales poseen dos dominios de reconocimientos a diferentes tipos de carbohidratos.
- Superlectinas: las cuales poseen una estructura similar a las quimerolectinas pero son capaces de reconocer tres o más carbohidratos distintos (31).

Otras clasificaciones involucran el tipo de carbohidrato que reconocen (manosa, galactosa, lactosa, etc.) y el grupo taxonómico del cual las mismas son aisladas (leguminosas, curcubitáceas, monocotiledóneas, etc) Figura 1-1 (31).

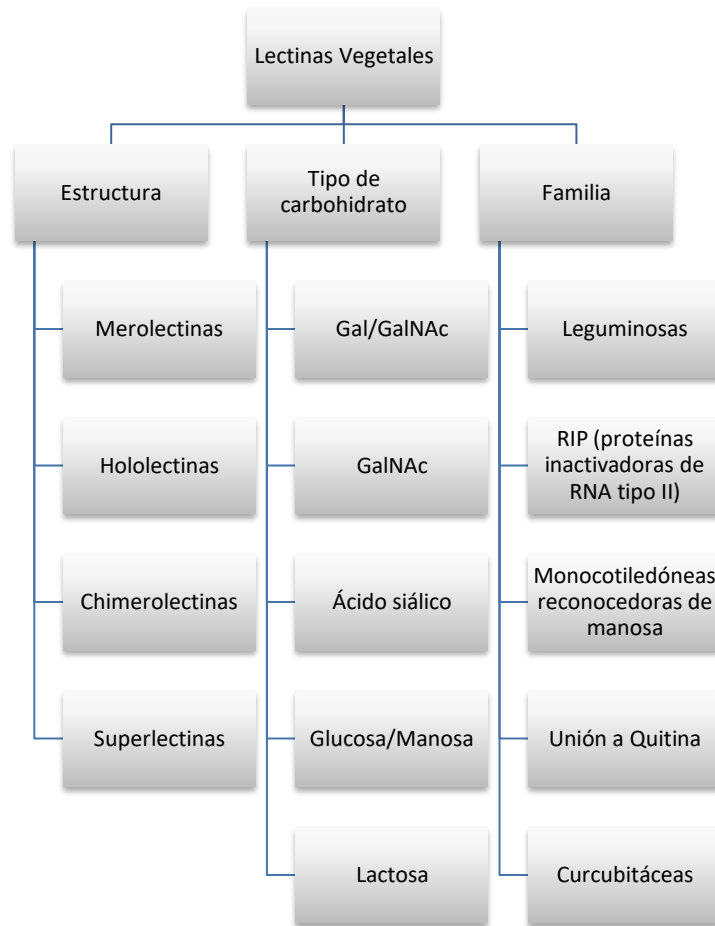


Figura 1-1: Clasificación de las lectinas vegetales: En el diagrama se observan los tres grandes grupos en los cuales se clasifican las lectinas vegetales. Modificado de C.S.F. Bath y colaboradores 2013 (31)

1.1.1 Estudio de las lectinas en Colombia

En Colombia, se han llevado a cabo varios estudios relacionados con lectinas vegetales los cuales están principalmente encaminados al aislamiento de las mismas, dentro de estos se encuentra el realizado por Gerardo Pérez y colaboradores en 1999 en donde se aisló, purificó y caracterizó dos lectinas provenientes de la planta *Dioclea sericea* la cual

se encuentra ampliamente distribuida en nuestro país (23). Recientemente se realizaron estudios de prospección de lectinas en varias especies de Lamiaceas endémicas de Colombia (26) y de algunas especies no endémicas pero naturalizadas en nuestro país (27).

Otros estudios realizados han estado encaminados a la utilización de las lectinas para la identificación de diferentes antígenos; dentro de este tipo de estudios destaca el realizado por Limpias y colaboradores, en donde utilizan la isolectina B4 de *V. villosa* para la identificación del antígeno Tn en células epiteliales de diferentes carcinomas (21); para la identificación de este mismo antígeno también se ha utilizado una lectina aislada de semillas de *Salvia bogotensis* (LSBo) la cual ha demostrado poseer un alto reconocimiento del mismo (32). En otros estudios se ha realizado la obtención y caracterización de inmunoglobulinas de yema de huevo de gallina (IgYs) que reconocen la lectina LSBo proveniente de estas semillas con el fin de que las mismas puedan ser utilizadas en técnicas de inmunohistoquímica para la detección de dicho antígeno (33,34).

1.1.2 Ingeniería de Lectinas

Uno de los campos que recientemente se ha estado explorando es la ingeniería de lectinas. En las que se incorporan pequeñas mutaciones en la secuencia de las proteínas, para que estas aumenten su especificidad frente a determinadas glicoproteínas (35). Existen numerosos ejemplos de este tipo de aplicación; como el estudio realizado por Kuwabara N. et al., en el que evaluaron los cambios en la interacción entre un glicoconjugado y la lectina proveniente del hongo *Agrocybe cylindracea* cuando esta última presenta un cambio en su estructura (36). Otro estudio llevado a cabo por Arango et al., en el que alteraron la lectina de *Erythrina corallodendron* (ECorL) y encontraron que la misma presentaba cambio en su afinidad frente a determinadas glicoproteínas (37). Dentro de este tipo de estudios, también destaca lo realizado por Yamamoto en donde se realizó una proteína quimérica entre una lectina de *Bauhinia purpurea* y *Lens culinaris* la cual logró reconocer glicoconjugados con gran cantidad de manosas en su estructura (38,39).

Otros estudios como el realizado por Elizabeth Jordan e Irwin Goldstein en donde observaron cambios en la especificidad de la lectina LBL extraída del haba de lima (*Phaseolus lunatus*), en la detección del trisacárido presente en células sanguíneas del grupo A humano al realizar un cambio puntual en la estructura de la misma (40) o el que realizó Nishiguchi y colaboradores, en donde lograron alterar la especificidad de la lectina RBL extraída de la corteza de *Robinia pseudoacacia* de tres de las proteínas recombinantes al cambiar el aminoácido ácido aspártico dentro de la estructura de la misma, probando que este es necesario para el reconocimiento del carbohidrato en esta lectina (41).

1.1.3 Lectinas que reconocen el antígeno Tn.

Durante las últimas décadas, el número de lectinas que son capaces de reconocer el antígeno Tn se ha visto incrementado, haciendo que este grupo de proteínas sea considerado para el desarrollo de nuevos métodos para la identificación del mismo (19). Paralelamente, la cantidad de estudios en donde se evidencia la capacidad citotóxica de las mismas en diversos tejidos cancerosos también se ha visto incrementado, ocasionando que las mismas también sean consideradas para fines terapéuticos (19).

Actualmente existen 46 lectinas, aisladas de plantas y hongos, que se encuentran caracterizadas y que son capaces de reconocer el antígeno Tn (19). Las lectinas vegetales que se han encontrado pertenecen a las familias de las Fabaceae en donde se encuentran *B. purpurea*, *G. simplicifolia*, *A. hypogaea*, *G. max*, *V. villosa*, *P. tetraglobonobulos*, *W. floribunda*, Caprifoliaceae como *S. nigra*, Lamiaceae como *S. sclarea*, *S. hominum*, *S. bogotensis*, Euphorbiaceae donde está *R. communis* y Moraceae como *M. pomífera*, *M. nigra* (19).

1.2 Metodología

1.2.1 Búsqueda bibliográfica y de estructuras primarias y terciarias.

La búsqueda bibliográfica de las lectinas se realizó mediante el motor de búsqueda google y del NCBI (42). Una vez identificadas las proteínas se escogieron aquellas que pertenecían a la familia de lectinas leguminoides y se procedió a buscar si estas poseían estructura primaria o terciaria en las bases de datos Uniprot (43) y Protein Data Bank (PDB) (44)

1.2.2 Análisis de la estructura primaria, secundaria y terciaria e interacción con el antígeno Tn de las proteínas encontradas.

Para el análisis de la estructura primaria y secundaria, se procedió a realizar un alineamiento múltiple de las secuencias encontradas con el algoritmo de ClustalOmega (45,46), el cual se analiza con el programa Jalview (47). Para el análisis de la estructura terciaria se escogieron las estructuras que se obtuvieron junto con el antígeno Tn y se descargaron los archivos .pdb de la base de datos PDB (44), los archivos se procesaron con el software MOE (48) para el alineamiento estructural y las figuras de interacción con el antígeno Tn; paralelamente se realizó un análisis de modo normal (NMA), el cual mide la flexibilidad de un grupo de proteínas con el fin de medir las variaciones del mismo cuando estas son sometidas a modelos vibracionales, con el paquete Bio3D (49).

1.3 Resultados y discusión.

1.3.1 Búsqueda de proteínas capaces de reconocer el antígeno Tn.

Se realizó una búsqueda bibliográfica de las diferentes lectinas que son capaces de reconocer el antígeno Tn, logrando identificar 25 lectinas vegetales que son capaces de identificar el mismo Tabla 1-1, de las cuales 13 pertenecían a la familia de proteínas de lectinas leguminoides Tabla 1-1 (19). De estas 13 lectinas encontradas tan solo las lectinas provenientes de *Glechoma hederacea* (Uniprot Q8H218)(50), *Griffonia simplicifolia* (51) y *Dolichus biflorus* (52) poseían secuencia en la base de datos de Uniprot (43), adicionalmente se usaron las secuencias de las lectinas de *Vicia villosa* (PDB 1N47)(53), *Vatairea macrocarpa* (PDB 4XTP)(54), *Psophocarpus tetragonolobus* (PDB 2D3S)(55), *Glycine max* (PDB 4D69)(56) y *Bauhinia fortificata* (PDB 5T5J)(57) tenían estructura terciaria reportada a la fecha (09-08-17) en conjunto con el antígeno Tn en la base de datos del PDB (44).

1.3.2 Análisis de la estructura primaria

El análisis de las estructuras primarias arrojó siete aminoácidos conservados en todas las estructuras que están involucrados en la interacción del carbohidrato y los iones Ca^{+2} y Mn^{+2} ; de los cuales tres están involucrados en el reconocimiento al carbohidrato y cuyas posiciones dentro del alineamiento son Asp87, Gly105, Asn129. También se encontró que los aminoácidos Asn125 y Asp129 en el alineamiento se encuentran involucrados en la interacción de el ion Ca^{2+} y el ion Mn^{2+} respectivamente. Otro residuo importante para el reconocimiento del ion Mn^{2+} solo se encuentra conservado en siete de las ocho estructuras, cambiándose de un Glu a una Lys en la posición 123 del alineamiento; al igual que la Leu213 el cual ha sido reportado como un residuo que permite estabilizar el anillo de galactopiranososa del antígeno Tn con el anillo aromático de la cadena lateral de un aminoácido no conservado que varía entre Phe, His y Trp que se encuentra ubicado en la posición 126 (53) en el sitio de reconocimiento del carbohidrato en la proteína y que se encuentra solo en las estructuras de *V. villosa*, *V. macrocarpa*, *G. max* y *D. Biflorus*. Estos

residuos coinciden con lo que se reporta en las estructuras cristalizadas de Isolectina B4 de *V. villosa* (53) y las lectinas *V. macrocarpa* (54), *P. tetragonolobus* (55), *G. max* (56) y *B. fortificata* (57), así como con lo reportado en otras lectinas que reconocen el carbohidrato GalNac, pero con poca o ninguna afinidad para reconocer el antígeno Tn (58,59). En la (Figura 1-2) se pueden identificar en rojo los residuos que interactúan con el antígeno Tn, en azul los aminoácidos que reconocen el ion Mn^{2+} , en verde los residuos que reconocen el ion Ca^{2+} y en púrpura los residuos con anillo aromático en su cadena lateral que estabilizan el anillo alifático de la galactopiranososa del antígeno Tn en el sitio de reconocimiento del carbohidrato. Las flechas verdes indican las hileras de hebras β , y los puntos rojos indican los giros en la estructura.

Tabla 1-1: Lectinas vegetales que reconocen el antígeno Tn. En negrilla se encuentra las lectinas pertenecientes a la familia de las leguminosas.

Especie de donde se aísla	Bibliografía	Especie de donde se aísla	Bibliografía
<i>Arachis hipogea</i>	(60)	<i>Griffonia simplicifolia</i>	(51)
<i>Amarantus caudatus</i>	(61)	<i>Molucella laevis</i>	(62)
<i>Amarantus leucocarpus</i>	(63)	<i>Morus nigra</i>	(64)
<i>Artocarpus integrifolia</i>	(65)	<i>Myrsine coriacea</i>	(25)
<i>Artocarpus lakoocha</i>	(66)	<i>Ricinus communis</i>	(67)
<i>Bauhinia fortificata</i>	(57)	<i>Salvia bogotensis</i>	(32)
<i>Bauhinia purpurea</i>	(68)	<i>Salvia sclarea</i>	(69)
<i>Laelia autumnalis</i>	(70)	<i>Salvia hominium</i>	(71)
<i>Maclura pomifera</i>	(72)	<i>Sambucus nigra</i>	(73)
<i>Dolichus biflorus</i>	(74)	<i>Vatairea macrocarpa</i>	(54)
<i>Glechoma hederácea</i>	(50)	<i>Vicia villosa</i>	(53)
<i>Glycine max</i>	(56)	<i>Psophocarpus</i>	(55)
		<i>tetraglonobulus</i>	
		<i>Wisteria floribunda</i>	(58)

1.3.3 Análisis de la estructura secundaria

El análisis de la estructura secundaria arrojó que los monómeros de las lectinas analizadas se configuran en un motivo de sándwich tipo β (jelly-roll), el cual se caracteriza por tener una hoja superior conformada por seis hileras de hebras β , una hoja pequeña en la parte superior, y una hoja frontal curva constituida por siete hileras de hebras β ; finalmente una serie de giros α que unen las hojas superiores con las hojas frontales. Siendo las hojas frontales junto con algunos pocos giros los que constituyen el dominio de reconocimiento al carbohidrato (75,76) (Figura 1-2)

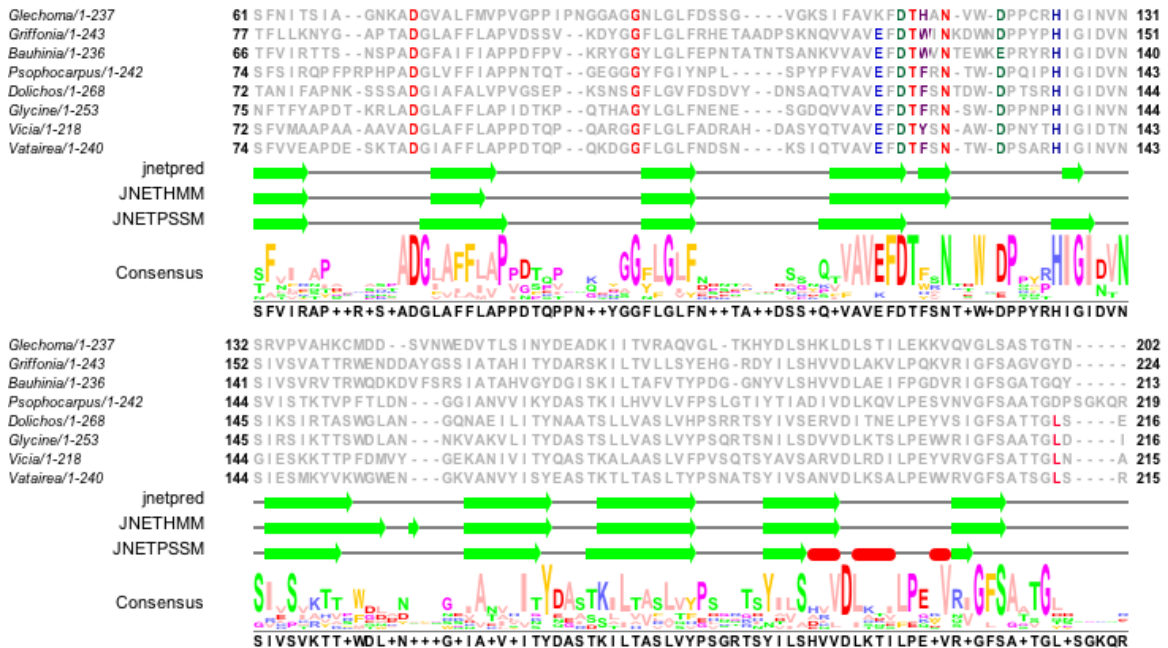


Figura 1-2: Alineamiento múltiple de las secuencias de las lectinas: En rojo se encuentran los residuos que interactúan con el antígeno Tn, en azul los aminoácidos que reconocen el ion Mn²⁺, en verde los residuos que reconocen el ion Ca²⁺ y en púrpura los residuos con anillo aromático en su cadena lateral que estabilizan el anillo alifático. Las flechas verdes indican las hileras de hebras β , y los puntos rojos indican los giros en la estructura.

1.3.4 Análisis de estructura terciaria.

Para el análisis de la estructura terciaria se utilizaron las lectinas de *Vicia villosa* (PDB 1N47)(53), *Vatairea macrocarpa* (PDB 4XTP)(54), *Psophocarpus tetragonolobus* (PDB 2D3S)(55), *Glycine max* (PDB 4D69)(56) y *Bauhinia fortificata* (PDB 5T5J)(57) las cuales tenían estructura terciaria reportada junto con el antígeno Tn a la fecha de realizar el experimento. Se excluyeron del mismo las estructuras terciarias de las lectinas de *G. hederácea*, *G. simplicifolia* y *D. biflorus* debido a que no se encontraban en conjunto con el antígeno Tn. El mismo arrojó que estas lectinas están conformadas por un monómero idéntico que puede agruparse con uno o tres monómeros iguales formando una estructura homodimérica u homotetradimérica, dependiendo de la especie de donde provenga, capaz de identificar diferentes carbohidratos (75,76), la superposición de las estructuras (Figura 1-4), no mostró cambios significativos en ninguna de las cinco estructuras confirmando que todas estas pertenecen a una superfamilia de proteínas. Al mismo tiempo el análisis de NMA (Figura 1-3), indicó que las cinco estructuras se comportan de una manera similar a nivel molecular una vez las mismas sean perturbadas.

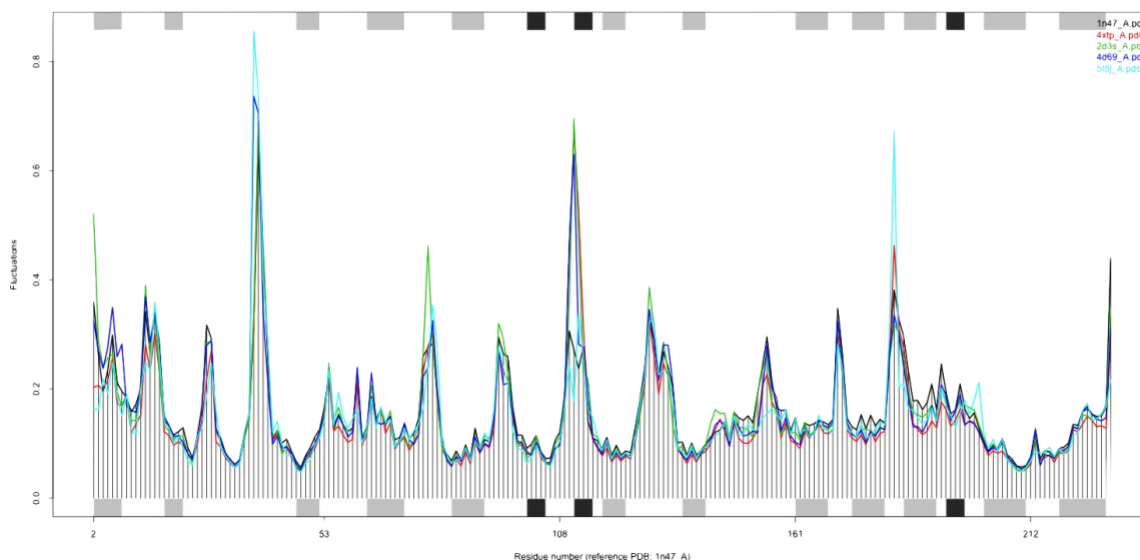


Figura 1-3: Análisis de NMA muestra que las cinco estructuras se comportan de manera similar.

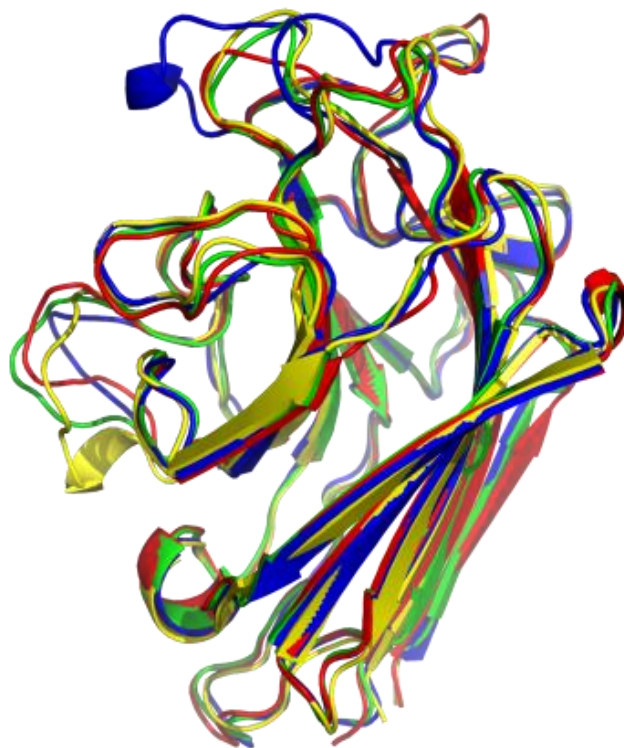


Figura 1-4: La superposición de las estructuras confirma lo conservada que es esta superfamilia de proteínas.

El análisis de la interacción entre la proteína y el antígeno Tn, en donde los aminoácidos de la lectina de *Vicia villosa* que interactúan con el antígeno son el Asp85 con el grupo hidroxilo del carbono 3 y el carbono 4 formando un puente de hidrógeno, Gly103 que interactúa con el grupo carbonilo del carbono 2 formando un puente de hidrógeno, Tyr127 con el grupo amino del aminoácido Thr que está adherido al anillo de galactopiranososa formando un puente de hidrógeno además de una interacción hidrofóbica con el anillo de galactopiranososa, Asn129 que reacciona con el grupo hidroxilo del carbono 4 formando un puente de hidrógeno, Leu213 y Asn214 formando un puente de hidrógeno que interactúan con el grupo hidroxilo del carbono 4 y el carbono 6 respectivamente (53)(Figura 1-5 **A**).

Para la lectina de *Vatairea macrocarpa* los residuos que interactúan son el Asp87 con el grupo hidroxilo del carbono 3 y el carbono 4 formando un puente de hidrógeno, Gly105 con el grupo carbonilo del carbono 2 formando un puente de hidrógeno, Asn129 con el grupo

hidroxilo del carbono 3 y Leu213 y Ser214 con el grupo hidroxilo del carbono 4 y el carbono 6 respectivamente todos ellos formando un puente de hidrógeno (54)(Figura 1-5 **B**).

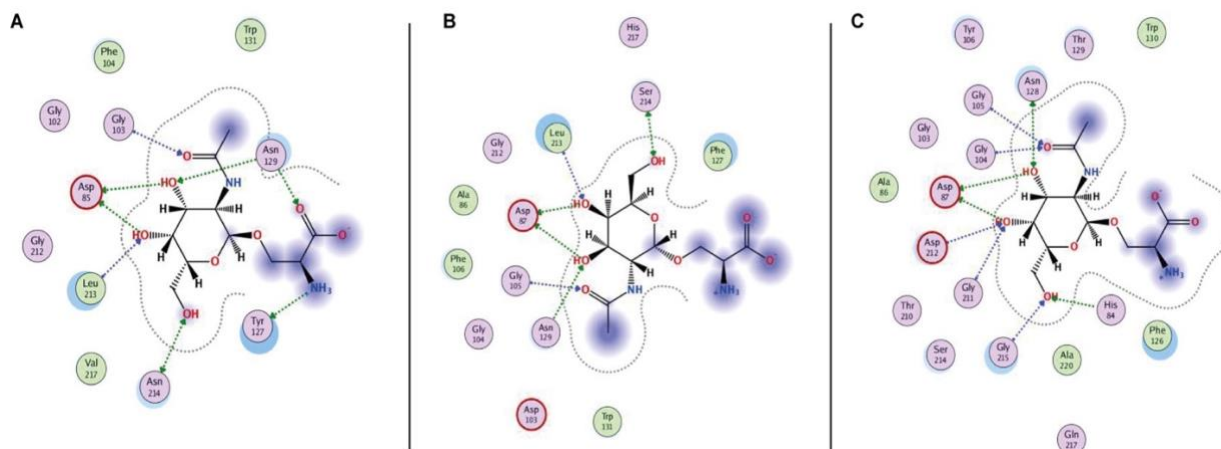


Figura 1-5: Interacciones entre las diferentes lectinas y el antígeno Tn. Diagrama de interacción de las lectinas de *Vicia villosa* **A**, *Vatairea macrocarpa* **B** y *Psophocarpus tetragonolobus* **C**.

Por último para la lectina de *Psophocarpus tetragonolobus* los aminoácidos que interactúan son la His84 con el grupo hidroxilo del carbono 6 formando un puente de hidrógeno, Asp87 con el grupo hidroxilo del carbono 3 y el carbono 4 formando un puentes de hidrógeno, Gly104 y Gly105 con el grupo carbonilo del carbono 2 formando un puentes de hidrógeno, Asn128 con el grupo hidroxilo del carbono 3 formando un puente de hidrógeno, Gly212 y Asp212 con el grupo hidroxilo del carbono 4 y Gly215 con el grupo hidroxilo del carbono 6 formando un puentes de hidrógeno (55)(Figura 1-5 **C**).

Estos datos confirman lo encontrado en el análisis de la estructura primaria; en donde los aminoácidos Asp, Gly, Thr, Asn son los que principalmente se encuentran relacionados con la identificación del antígeno Tn. Adicionalmente se encuentra que los aminoácidos Tyr, Trp y Phe estan reportados en la bibliografía y se encagan principalmente de formar el bolsillo hidrofóbico que mantiene el anillo de galactopiranososa estable dentro del mismo (53–55). Estas interacciones no son evidenciadas en el gráfico de interacción.

2. Capítulo 2: Diseño de péptidos y modelamiento de los péptidos.

2.1 Diseño racional de compuestos.

En las últimas décadas, la industria farmacéutica ha realizado grandes avances en la identificación de moléculas capaces de identificar objetivos terapéuticos específicos (77). Para ello cada vez se están utilizando técnicas computacionales, con el fin de reducir los costos de desarrollo al mismo tiempo que se crean compuestos objetivo-específicos, los cuales presentan ventajas como la reducción de los efectos adversos o la resistencia a los medicamentos frente a los medicamentos convencionales (77)(78). Dentro de este tipo de compuestos, las moléculas que poseen aminoácidos en su estructura; tales como proteínas y los péptidos se han presentado como una alternativa de tratamiento prometedora, sobre todo para objetivos biológicos que actualmente no poseen un tratamiento (77).

2.1.1 Diseño computacional de proteínas.

Las proteínas son macromoléculas capaces de realizar diferentes funciones en un organismo. Estas funciones están dictadas mayormente por la manera en que estas se pliegan por lo que se podría afirmar que la misma depende de su estructura (79). Actualmente existen dos métodos para el diseño de proteínas; el primer método consiste en el uso de una estructura tridimensional existente como plantilla para la modelación de una nueva proteína, en el segundo método se crea una secuencia de aminoácidos y a partir de ella se genera una estructura tridimensional con el fin de generar una proteína completamente nueva (77).

Estudios como el de King y colaboradores en donde se describe la creación de bloques de proteínas con 24 subunidades a partir de proteínas triméricas para la creación de nanomateriales (80), otro como el estudio de Simmons y colaboradores en donde utilizan algoritmos bayesianos capaces de tomar fragmentos de proteínas no relacionadas para crear proteínas que asemejan proteínas nativas (81) que han abierto la posibilidad para que hoy en día existan varios softwares capaces de generar proteínas capaces de identificar objetivos biológicos específicos utilizando algoritmos evolutivos (82).

Aunque el diseño de proteínas *de novo* presenta muchos inconvenientes debido a el gran número de posibilidades que se presentan (20^N donde N se considera el número existente de aminoácidos) (77); estudios como el de Chevalier y colaboradores en donde se logró crear 2698 pequeñas proteínas capaces de identificar la hemaglutinina del virus de la influenza y la neurotoxina B encontrada en la botulina (83) demuestran lo eficiente que puede llegar a ser esta aproximación.

2.1.2 Diseño computacional de péptidos.

Los péptidos son pequeñas estructuras protéicas que usualmente interactúan en pequeñas cavidades de un estructura protéica mas grande inhibiendo su actividad catalítica, característica que hace que los mismos sean usualmente utilizados como objeto de estudio para la obtención de nuevos tratamientos (84). Aunque estudios como el de Roberts y colaboradores (85) demuestran que la idea de modificar este tipo de moléculas no es nueva, el incremento de la resolución en las estructuras que se encuentran en las bases de datos así como el aumento en el poder de procesamiento de los equipos informáticos, han propiciado que recientemente hayan cada vez más estudios en este campo (84).

Actualmente el diseño racional de péptidos para su uso en tejido canceroso busca que las moléculas creadas cumplan al menos una de las siguientes funciones en el tejido objetivo: 1) un componente terapéutico, que incluya nanotecnología y que permita un mejoramiento en el suministro del medicamento y 2) un componente de medición que permita monitoriar fácilmente la evolución del paciente (86).

Para el primer ítem destacan los estudios realizados por Karagiannis y colaboradores en donde crearon una librería con 50 péptidos capaces de transportar pequeños RNAs de interferencia a través de la membrana al interior de la célula sin que generen respuesta inmune de memoria (87) o el estudio Ran Xu y colaboradores, quienes en 2018 crearon unos péptidos basados en los motivos LZ4 y LZ1-3 con el fin de bloquear el factor de calor 1 (HSF1) el cual es responsable de la activación de varios mecanismos de supervivencia en células con procesos cancerosos (88).

En cuanto al segundo ítem, el estudio liderado por Hongwang y colaboradores en donde crearon una nonoesponja que asemeja un lisosoma y podría permitir que moléculas altamente hidrofóbicas penetren el tejido afectado y realicen su cometido al interior de la célula (89). O el estudio de Fang y colaboradores en donde se buscaba crear nanopartículas capaces de competir con ciertos epítopes presentes en algunos venenos (90).

Por último uno de los estudios más promisorios que se tienen en el campo de medición, es el elaborado por el equipo de Yifeng y colaboradores. Quienes crearon un RNA de interferencia unido a una nanopartícula de oro marcada fluorescentemente (GNCs) que les permitió a los investigadores; no solo silenciar el gen NGF que está involucrado en la proliferación de cáncer pancreático, sino que, al mismo tiempo, también les permitió monitorear el estado del paciente sin efectos adversos (91).

En resumen el diseño racional de péptidos terapéuticos es una alternativa útil para el tratamiento de cánceres especialmente agresivos como el de páncreas o colon. El mismo presenta una serie de ventajas como su fácil síntesis, baja toxicidad, alta diversidad química y biológica y alta especificidad, comparado frente a métodos más convencionales como el aislamiento de anticuerpos monoclonales u otras proteínas que participan en la respuesta inmune tumoral (92).

2.2 Metodología

2.2.1 Diseño y modelación de los péptidos.

Con los datos obtenidos sobre los aminoácidos involucrados en el reconocimiento del antígeno, se procedió a diseñar dos péptidos, un primer péptido conformado por 20 aminoácidos y un segundo péptido derivado del primero conformado por 17 aminoácidos. Para ello, se tomaron los tres aminoácidos conservados que se evidenciaron en la fase previa (Asp, Gly, Asn) junto con un residuo que poseía un anillo aromático en su cadena latera (Phe) y se completaron hasta poseer el número de aminoácidos que se quería con residuos de prolina (Pro), para que el péptido adquiriera una forma de bolsillo, un residuo de Glutamato (Glu), un residuo de Serina (Ser) y aminoácidos cuya cadena lateral no interactúa con el antígeno (Ala, Ile, Met, Val).

2.2.2 Creación de la librería de péptidos.

Para la creación de la librería de péptidos, los péptidos iniciales se mutaron utilizando la técnica de mutagénesis simple dirigida; para ello se creó un script en el lenguaje de programación Perl (Anexo A), el cual selecciona al azar un residuo de la cadena de cualquiera de los péptidos y lo cambia por un aminoácido, seleccionado de la siguiente lista: Asn, Gln, Asp, Glu, Ser, Met, Ile, Leu, Gly, Pro, Phe, Tyr, también al azar. Con este script se generaron 50 péptidos de cada uno, para un total de 100 péptidos.

2.2.3 Modelamiento de los péptidos.

Se seleccionaron al azar 20 de cada uno de los dos péptidos y estos fueron modelados utilizando el servidor PepFold (93–96). Al mismo tiempo se determinó su masa molecular, coeficiente de extinción, punto isoeléctrico y carga eléctrica a pH 7 utilizando el servidor PepDraw (<http://www.tulane.edu/~biochem/WWW/PepDraw/index.html>) que se encuentra en el servicio de recursos bioinformáticos ExPASy. Los valores de RMSD se obtuvieron a través del paquete Bio3D (49).

2.3 Resultados y discusión

2.3.1 Diseño y modelamiento de los péptidos con base en lectinas leguminoides.

El péptido creado a partir de 20 aminoácidos se denominó DARF1, el mismo se mutó 49 veces para obtener la librería de péptidos (Tabla 2-1). De esta se eligieron 20 péptidos al azar, los cuales se encuentran resaltados con negrilla. El análisis de RMSD Tabla 2-3 para este péptido muestra que los péptidos DARF37, DARF41 y DARF44 son los que más se parecen estructuralmente al péptido DARF1 siendo el que más difiere de la misma el péptido DARF27.

Tabla 2-1: Péptidos diseñados a partir de 20 aminoácidos.

Péptido	Secuencia	Péptido	Secuencia
DARF1	AIRQGPIPFPGPDGDPRNSV	DARF26	AIRQGPIPFPGPDGDPPNSV
DARF2	AIREGPIPGPGMDGDPRNSV	DARF27	AIRDGPIPFPGPDGDPRNSF
DARF3	VALEGPLPGMGPDGDPVNSA	DARF28	AIRMGPIPFPGPDGDPRNSV
DARF4	AIRQGPIPFPGPDGDPRNSV	DARF29	PIRQGPIPFPGPDGDPRNSV
DARF5	AIRQGSIPFPGPDGDPRNSV	DARF30	AIRQGPIPFPGPDGDPRNSQ
DARF6	AIRQGPIPFPGPDIDPRNSV	DARF31	AIRQGPIPMFPGPDGDPRNSV
DARF7	AIRQGPIPFPGPDMDPRNSV	DARF32	AMRQGPIPFPGPDGDPRNSV
DARF8	AGRQGPIPFPGPDGDPRNSV	DARF33	AIRQGPPPFPGPDGDPRNSV
DARF9	AIRQGPIIFPGPDGDPRNSV	DARF34	AIRQGPIPFPGPDGDPRNGV
DARF10	AIRQGPIDFPGPDGDPRNSV	DARF35	AIRQGPIPFPGPDGDPRNSV
DARF11	AIRQGPIPFPGPDGYPRNSV	DARF36	AIRQGPIPFPGPDGDPYNSV
DARF12	AIRQGPIPFPGPDGDPQNSV	DARF37	AIRQGPIPFPGNDGDPRNSV
DARF13	AIRQGPIPFPGYDGDPRNSV	DARF38	APRQGPIPFPGPDGDPRNSV
DARF14	AIRQGPISFPGPDGDPRNSV	DARF39	AIRQGPMPFPGPDGDPRNSV
DARF15	AIRQGPIPFPGPDGDPNQNQV	DARF40	AIRQGPIPFPGPDGNPRNSV
DARF16	AIRQGPIPFPGPDGDPRNSV	DARF41	AIRQGPIPFPGPDPRNSV
DARF17	AIRQGPIPFPGPDGDPRNSQ	DARF42	AIRQGPIPFPGPDGDPRNSV
DARF18	LIRQGPIPFPGPDGDPRNSV	DARF43	IIRQGPIPFPGPDGDPRNSV
DARF19	AIRQGPIPFPGPDGYPRNSV	DARF44	AIRQGPIPFPGPGGDPRNSV
DARF20	AIRQGPIPFPGPDGDPRMSV	DARF45	FIRQGPIPFPGPDGDPRNSV
DARF21	AIRQGPIPFPGPDGGPRNSV	DARF46	AIRQGPDPFPGPDGDPRNSV
DARF22	AIRQGSPFPGPDGDPRNSV	DARF47	AISQGPIPFPGPDGDPRNSV
DARF23	AIRQNPIPFPGPDGDPRNSV	DARF48	AIRQGSIPFPGPDGDPRNSV
DARF24	AIRFGPIPFPGPDGDPRNSV	DARF49	AIRQGPIPFYPDGDPRNSV
DARF25	AIRDGPIPFPGPDGDPRNSV	DARF50	AIRQGMIPFPGPDGDPRNSV

En cuanto a el péptido creado con 17 aminoácidos, el mismo se le denominó DARF51; y se mutó 49 veces para obtener la librería (Tabla 2-2), seleccionándose los 20 péptidos que aparecen en negrilla. En cuanto al análisis de RMSD (Tabla 2-3) para el péptido DARF51, encontramos que los péptidos que más se parecen en su estructura al péptido DARF51 son DARF60, DARF65 y DARF92 siendo el que más difiere de la misma el péptido DARF71.

Tabla 2-2: Péptidos diseñados a partir de 17 aminoácidos

Péptido	Secuencia	Péptido	Secuencia
DARF51	LQGPLPFMGPNGDPVDS	DARF76	LQGPLPFMGPYGDVDS
DARF52	LQGPLPFMGNQDPVDS	DARF77	LQGPLPFGGPNQDPVDS
DARF53	LQGPLPFQGPNGDPVDS	DARF78	LQGPLQFMGPNGDPVDS
DARF54	LQGPLPFMGENDPVDS	DARF79	LQGPLPFMGPNGDDVDS
DARF55	LQGPLPFMGNIDPVDS	DARF80	LQGPIPFMGPNGDPVDS
DARF56	LQGPLPFMGPNGDPVDD	DARF81	LFGPLPFMGPNGDPVDS
DARF57	LQGPLPFIGPNDPVDS	DARF82	LQGPLPFMGPNGDIVDS
DARF58	LQGPLPFMGPNGDPVDQ	DARF83	LQGPLPFMGPNGDPVDS
DARF59	IQGPLPFMGPNGDPVDS	DARF84	LQGQPFMGPNGDPVDS
DARF60	LFGPLPFMGPNGDPVDS	DARF85	LQGPLPFMLPNDPVDS
DARF61	LQGPLPFMGPNGPPVDS	DARF86	LQGPLPFQGPNGDPVDS
DARF62	LQGPLPFMGPNGDPVDY	DARF87	LQGPLPFGGPNQDPVDS
DARF63	LQGPLPNMGPNGDPVDS	DARF88	LQGPLPFMGLNDPVDS
DARF64	LQGELPFMGPNGDPVDS	DARF89	LQGPLPFMGLNDPVDS
DARF65	LQGPLPFMGPNGDPMDS	DARF90	LQGPLPFLGPNQDPVDS
DARF66	LQGPLPFMGPNGDPVES	DARF91	LQGQLPFMGPNGDPVDS
DARF67	LQGPLPFMGPNGDPVDS	DARF92	LQGPLPFMGPNGDPVDS
DARF68	LQGPLPFMGPNGDGVDS	DARF93	LQGPLPFGGPNQDPVDS
DARF69	LQGPLPFMGNDDPVDS	DARF94	LQGPLPFMGPNGDPLDS
DARF70	LQGPLPFMGPNGDPVDN	DARF95	LQGPLPDMGPNGDPVDS
DARF71	LQGPLPFMGNIDPVDS	DARF96	SQGPLPFMGPNGDPVDS
DARF72	LQGPLPFMGYNGDPVDS	DARF97	LQGQPFMGPNGDPVDS
DARF73	LQGYLPFMGPNGDPVDS	DARF98	LQGPLPFGGPNQDPVDS
DARF74	LQEPLPFMGPNGDPVDS	DARF99	LQGPLPFGGPNQDPVDS
DARF75	LQGPLPFMGPNGDPPDS	DARF100	YQGPLPFGGPNQDPVDS

Tabla 2-3: RMSD de cada uno de los péptidos seleccionados

Péptido	RMSD	Péptido	RMSD	Péptido	RMSD	Péptido	RMSD
DARF1	0.000	DARF25	7.068	DARF51	0.000	DARF77	3.691
DARF2	7.219	DARF27	7.439	DARF53	4.452	DARF79	4.283
DARF3	6.831	DARF31	6.544	DARF58	4.090	DARF80	4.721
DARF5	5.679	DARF34	7.130	DARF60	3.241	DARF81	6.044
DARF8	5.984	DARF37	4.897	DARF62	3.730	DARF85	4.490
DARF10	6.176	DARF39	6.258	DARF64	3.952	DARF86	4.336
DARF12	5.536	DARF41	4.559	DARF65	2.771	DARF88	3.889
DARF15	5.153	DARF44	4.175	DARF68	4.983	DARF90	4.695
DARF18	5.605	DARF46	5.070	DARF70	3.626	DARF92	1.194
DARF21	5.833	DARF47	6.413	DARF71	6.979	DARF96	3.832

2.3.2 Características fisicoquímicas de los péptidos modelados.

Tabla 2-4: Características fisicoquímicas de los péptidos diseñados a partir de 20 aminoácidos

PÉPTIDO	MASA CON AGUA (Dalton)	PUNTO ISOELÉCTRICO (PI)	CARGA	HIDROFOBICIDAD (KCAL*MOL ⁻¹)	COEFICIENTE DE EXTINCIÓN (M ⁻¹ *CM ⁻¹)
DARF1	2089,0672	6,91	0	+21,12	0
DARF2	2033,9921	4,27	-1	+26,03	0
DARF3	1891,8955	2,89	-3	+22,19	0
DARF5	2079,0465	6,91	0	+21,44	0
DARF8	2033,0048	6,91	0	+23,39	0
DARF10	2107,0414	4,21	-1	+24,62	0
DARF12	2061,0247	3,93	-1	+20,08	0
DARF15	2130,0937	6,91	0	+21,43	0
DARF18	2131,1140	6,91	0	+19,37	0
DARF21	2031,0618	10,96	+1	+18,63	0
DARF25	2076,0356	4,21	-1	+23,99	0
DARF27	2076,0356	4,21	-1	+23,99	0
DARF31	2073,0393	6,91	0	+22,91	0
DARF34	2059,0567	6,91	0	+21,81	0
DARF37	2106,0574	6,91	0	+21,83	0
DARF39	2107,0237	6,91	0	+21,57	0
DARF41	2129,0984	6,21	0	+20,11	0
DARF44	2031,0618	10,96	+1	+18,63	0
DARF46	2091,0102	4,21	-1	+25,88	0
DARF47	2019,9982	3,93	-1	+19,77	0

Para los péptidos construidos a partir del péptido de 20 aminoácidos se obtuvo un peso que se encuentra entre los 1800 Dalton y 2131 Dalton, un punto isoeléctrico para la mayoría de los péptidos que va desde 4,20 a 6,9, la carga eléctrica neta es igual a 0 para la gran mayoría de los péptidos, siendo el péptido DARF3 el que posee mayor carga neta negativa (-3), la hidrofobicidad para este grupo de péptidos varía entre +18,63 y +26,03 en la escala Wimley-White, y un coeficiente de extinción para todos los péptidos fue de 0

Tabla 2-4.

Tabla 2-5: Características fisicoquímicas de los péptidos diseñados a partir de 17 aminoácidos

PÉPTIDO	MASA CON AGUA (Dalton)	PUNTO ISOELÉCTRICO (PI)	CARGA	HIDROFOBICIDAD (KCAL*MOL⁻¹)	COEFICIENTE DE EXTINCIÓN (M⁻¹*CM⁻¹)
DARF51	1739,8161	2,94	-2	+15,93	0
DARF53	1739,8342	2,94	-2	+17,37	0
DARF58	1780,8426	2,94	-2	+16,24	0
DARF60	1758,8259	2,94	-2	+13,45	0
DARF62	1815,8373	2,94	-2	+14,76	1490
DARF64	1771,8059	2,87	-3	+19,42	0
DARF65	1771,7882	2,94	-2	+15,72	0
DARF68	1699,7849	2,94	-2	+16,94	0
DARF70	1766,8270	2,94	-2	+16,32	0
DARF71	1795,8785	2,94	-2	+13,66	0
DARF77	1665,7975	2,94	-2	+17,75	0
DARF79	1757,7903	2,84	-3	+19,43	0
DARF80	1739,8161	2,94	-2	+16,06	0
DARF81	1758,8259	2,94	-2	+13,45	0
DARF85	1795,8785	2,94	-2	+13,53	0
DARF86	1736,8342	2,94	-2	+17,37	0
DARF88	1755,8473	2,94	-2	+14,54	0
DARF90	1721,8596	2,94	-2	+15,35	0
DARF92	1739,8161	2,94	-2	+15,93	0
DARF96	1713,7642	2,88	-2	+17,64	0

Para los péptidos construidos a partir del péptido de 17 aminoácidos se obtuvo un peso que se encuentra entre 1665 Dalton y 1815 Dalton, un punto isoelectrico ácido para todos los péptidos que varía 2,87 a 2,94, la carga neta para la mayoría de los péptidos es de -2 siendo los péptidos DARF64 y DARF79 los que poseen un mayor número de residuos cargados negativamente (-3), la hidrofobicidad varía de +13,45 a +19,42 en la escala Wimley-White, y el coeficiente de extinción para todos los péptidos fue de 0 excepto para el péptido DARF62; esto se debe a que, de los péptidos seleccionados fue el único con una tirosina en su estructura

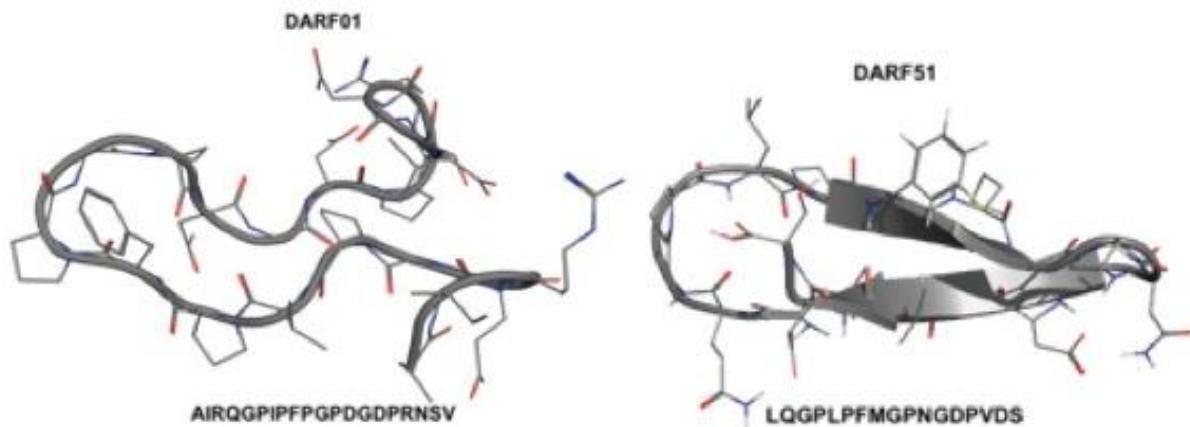


Figura 2-1: Péptido DARF1 y péptido DAR51. El primero adquiere una estructura tipo random coil, mientras que el segundo adquiere una estructura similar a una hebra β .

Los péptidos diseñados a partir de 20 aminoácidos tienden a adquirir una estructura tipo random coil cuando fueron modelados mientras que los péptidos diseñados a partir de 17 aminoácidos tienden a adquirir una estructura similar a una hebra β (Figura 2-1). También se observa que los péptidos diseñados a partir de 17 aminoácidos poseen un mayor número de grupos funcionales con carga negativa que los péptidos diseñados a partir de 20 aminoácidos lo que les confiere una carga neta mas negativa y una mayor probabilidad de que estos interactúen con el antígeno Tn.

3. Capítulo 3: Pruebas de Docking molecular.

3.1 Docking molecular.

Las pruebas de docking molecular han probado ser útiles a la hora de descubrir y diseñar nuevas drogas terapéuticas (97). Actualmente, los programas que se han desarrollado para realizar este procedimiento se dividen en tres categorías: docking comparativo, docking local y docking global los cuales se aproximan al problema de manera diferente y poseen distintos niveles de eficacia al momento de determinar el número de interacciones (97).

3.1.1 Docking comparativo.

El docking comparativo o docking basado en modelos, utiliza estructuras ya conocidas como plantilla y el uso del mismo se recomienda cuando el modelo utilizado es bastante similar a la molécula que se está investigando (97). Usualmente estos programas buscan similitudes entre la plantilla utilizada y la estructura del péptido, estas similitudes pueden ser a nivel estructural, utilizando la secuencia, o a nivel funcional en donde utiliza la información disponible de las interacciones en el archivo .pdb. luego de esto genera las simulaciones teniendo como modelo la optimización de la energía del complejo, lo que permite la flexibilidad de la simulación (98–100).

3.1.2 Docking Local.

El docking local es un método que busca encontrar las interacciones que existen en un espacio definido por el usuario entre el ligando y el receptor utilizando, por lo tanto la exactitud de los datos depende en gran medida de la información que se tenga sobre el sitio de unión (97). Los diferentes métodos disponibles para realizar docking local usualmente utilizan diferentes estrategias para definir el sitio de unión; por ejemplo, algunos algoritmos requieren de una preparación previa de los archivos, como resultado de esto, se pueden obtener mejores resultados siempre y cuando no haya una diferencia mayor a 5 Å entre el esqueleto del péptido y la estructura experimental (101–103), permitiendo el refinamiento de simulaciones anteriores con resultados de calidad media.

3.1.3 Docking global

Esta metodología de docking realiza una búsqueda automática para encontrar la posición del sitio de unión del péptido; la aproximación mas simple para aplicar con esta metodología consiste en poner el ligando y el receptor en una conformación rígida (97). Pero métodos más sofisticados pueden incluir procesos adicionales en el flujo de trabajo, actualmente el flujo de trabajo más empleado cuando se utiliza este tipo de docking consiste en 1) generación de las conformaciones del péptido estudiado; 2) realizar el docking con las estructuras rígidas; y 3) evaluar los modelos y refinar los resultados (97).

3.1.4 Dificultades de las pruebas de docking molecular.

Actualmente se han identificado tres grandes dificultades que poseen las pruebas de docking molecular; 1) el modelamiento de los cambios conformacionales en el receptor y el ligando o problema de flexibilidad; 2) la selección de la conformación mas eficiente de todos modelos generados; 3) la integración de los datos experimentales con los datos obtenidos en el modelo (97).

La exactitud en la predicción de las pruebas de docking es un problema que depende de la cantidad de enlaces rotables que posea tanto el ligando como el receptor (97). En el ligando, esta capacidad depende tanto de su tamaño como de su estructura secundaria, obteniéndose mejores resultados en péptidos con un tamaño entre 7-15 aminoácidos (104,105). En el caso del receptor, el rango de flexibilidad del mismo puede variar de pequeños cambios en las cadenas laterales, a grandes rearrreglos que incluyen el esqueleto de la proteína (97), esta aproximación implica varios desafíos a nivel de procesamiento computacional (106), por lo que se recomienda que el receptor sea rígido (97).

En cuanto al problema en la selección de la mejor conformación, usualmente cada algoritmo califica sus resultados del mejor al de menor calidad; para ello, la gran mayoría de estos utiliza la energía libre que arroja el modelo como un indicativo de calidad (101–103,107), aunque métodos mas refinados estan siendo empleados actualmente (97).

Por último el problema de integración de los datos puede resultar bastante desafiante teniendo en cuenta la ambigüedad que presentan algunos datos experimentales (97). Para ello algoritmos como HADDOCK permiten integrar datos experimentales como restricciones a la hora de realizar el docking (108). Otra aproximación para la interpretación de datos ambiguos ampliamente usada es el uso de varios métodos de docking con el fin de encontrar los mecanismos de unión más comunes a la conformación (97).

3.2 Metodología.

3.2.1 Preparación de los archivos para la prueba de Docking molecular.

Los péptidos seleccionados fueron descargados y preparados utilizando el software AutoDOCK Tools (107) para ser empleados como el ligando. Los enlaces del esqueleto del péptido, los enlaces entre átomos seleccionados y los enlaces activos se configuraron

como enlaces rotables, los demás tipos de enlaces se configuraron como no rotables, para el receptor se utilizó la estructura del antígeno Tn obtenida por Borgert y colaboradores (109) y el mismo se dejó rígido, como control positivo se utilizaron las estructuras de las lectinas obtenidas de *Vicia villosa* (PDB 1N47) (53), *Vatairea macrocarpa* (PDB 4XTP) (54), no se utilizó control negativo.

Para el docking de los péptidos el tamaño del GridBox fue de 20 puntos para los ejes X, Y, Z respectivamente y se ubicó en las coordenadas X=-31.096, Y=-18.484 y Z=-48.698, se realizaron tres simulaciones por cada péptido evaluado.

3.2.2 Análisis de los datos de las pruebas de Docking molecular

La integridad de los datos se revisa con una prueba de anova, se propone un mínimo de energía de de -3.0 kcal/mol para poder considerar el péptido como promisorio; el análisis estadístico se realizó con el software R (110).

3.2.3 Análisis de las conformaciones de los péptidos seleccionados.

Se analizan todas las conformaciones de los péptidos seleccionados teniendo en cuenta los criterios mencionados en el punto anterior, el análisis se realiza con los programas AutoDock Tools (107) y Pymol (111), las imágenes se obtuvieron con el software Pymol (111).

3.3 Resultados y discusión.

3.3.1 Análisis de datos de Docking molecular.

La prueba de Anova verificó la integridad de los datos obtenidos. Se determinó que los péptidos construidos a partir de 20 aminoácidos no interactuaban con el antígeno Tn que se utilizó en las pruebas. Esto se debe a que, la mayoría de los mismos, poseían una carga eléctrica igual a 0; siendo en este grupo el péptido DARF3 el que mejor se desempeñó debido a que el mismo presentaba la carga eléctrica mas baja. En contraste todos los péptidos obtenidos a partir de la secuencia de 17 aminoácidos, obtuvieron un promedio de energía libre inferior a -3 kcal/mol, lo cual refleja lo observado en las propiedades fisicoquímicas de los mismos Figura 3-1.

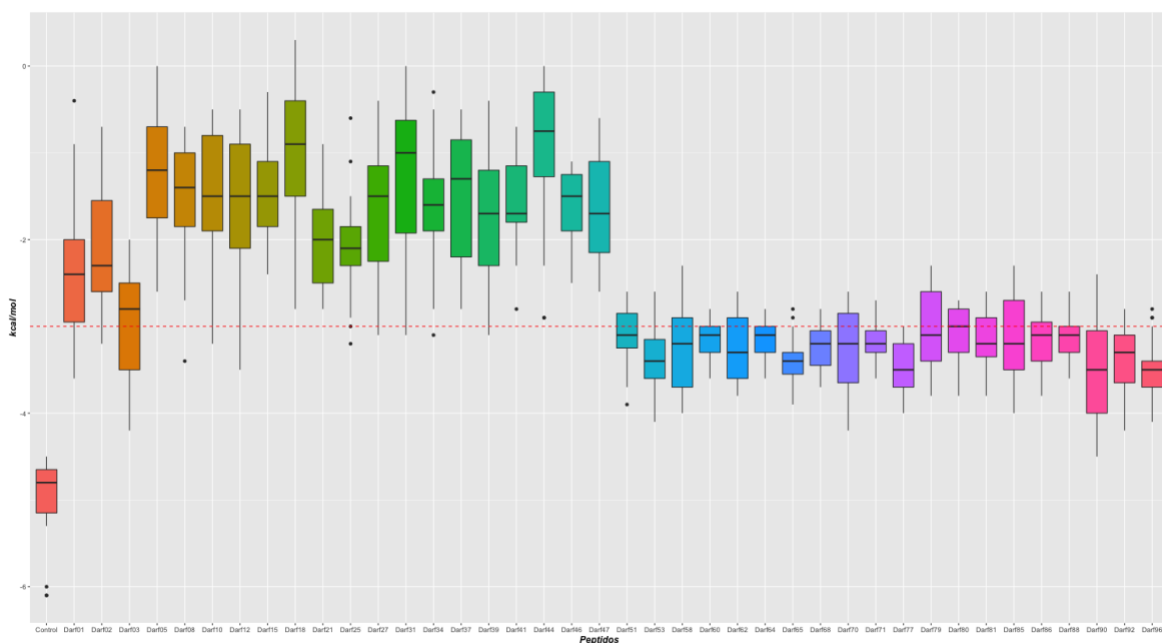


Figura 3-1: Promedio de los péptidos evaluados, se observa que solo los péptidos creados a partir de 17 residuos lograron superar el umbral de energía libre establecido (línea punteada roja).

Con base en los resultados obtenidos se decide realizar el análisis de las interacciones de los péptidos construidos a partir de 17 aminoácidos. Adicionalmente se resuelve analizar el péptido DARF3 con el fin de poder observar el comportamiento de los péptidos provenientes de 20 aminoácidos

3.3.2 Análisis de las interacciones

Para poder determinar los péptidos con mejor rendimiento, se decide revisar todas las conformaciones producto de las pruebas de docking molecular, los datos que se muestran en la Tabla 3-1 corresponden a la mejor conformación adquirida por cada uno de los péptidos.

Se descartan los péptidos DARF51, DARF58, DARF60, DARF64, DARF70, DARF71, DARF77, DARF79, DARF80, DARF85, DARF86, DARF88, DARF90, DARF92 y DARF96 ya que presentaban pocas o ninguna interacción con el antígeno Tn o las mismas no correspondían con los aminoácidos Asp, Gly, Glu, Gln y Asn (Anexo B) los cuales son los que se encuentran reportados en la literatura (53–57). Tampoco se observó interacción hidrofóbica entre los anillos aromáticos de los residuos Tyr, Phe o His y el anillo de galactopiranososa del antígeno en ninguno de los péptidos, esto se puede deber a que el tamaño del péptido no sea suficiente para estos residuos se puedan acomodar de tal manera que se pueda evidenciar dicha interacción.

Tabla 3-1: Mejores conformaciones de los péptidos analizados.

Péptido	Promedio Energía (kcal/mol)	Energía mejor conformación (kcal/mol)	Residuos que intervienen en el péptido	Residuos que intervienen en el antígeno	Tipo de interacción
DARF3	-2,9	-2,7	L3NH ₂ , G5NH, G9NH ₂ , D13CO, D15CO, N18CO, S19COH	L7COOH, O, OHC6, OHC3, OHC6, K8NH ₃ , COH ₃	Puentes de Hidrógeno

Tabla 3-1: Continuación.

Péptido	Promedio Energía (kcal/mol)	Energía mejor conformación (kcal/mol)	Residuos que intervienen en el péptido	Residuos que intervienen en el antígeno	Tipo de interacción
DARF51	-3,1	-2,8	Q2CO, Q2NH ₂ , V15CO, V15NH ₂	OHC3, OHC6, OHC6, OHC6	Puentes de Hidrógeno
DARF53	-3,3	-3,6	N11CO, D13COO, D16COO	OHC6, OHC3, NH	Puentes de Hidrógeno
DARF60	-3,1	-3,4	G3CO, N11COOH, VAL15CO	OHC4, T3OH, OHC3	Puentes de Hidrógeno
DARF62	-3,2	-3,8	Q2CO, G9CO, P10CO, D16COOH	T3OH, OHC4, OHC3, NH	Puentes de Hidrógeno
DARF64	-3,1	-3,6	F7CO, G9CO, S17COO	OHC4, OHC3, K8NH ₃	Puentes de Hidrógeno
DARF65	-3,3	-3,8	P4CO, L5CO, G9CO, N11CO, D13COO, S17CO, S17CO	OHC6, OHC6, K8NH ₃ , K8NH ₃ , OHC4, OHC3, OHC3	Puentes de Hidrógeno
DARF68	-3,2	-3,7	Q2CO, G9CO, D13CO, D16COO	T3OH, OHC4, NH, OHC3	Puentes de Hidrógeno
DARF70	-3,2	-4,1	G9CO, P10CO, D16COO	OHC6, OHC4, OHC3	Puentes de Hidrógeno
DARF77	-3,4	-3,7	F7CO, G8CO, G9CO	OHC6, OHC3, OHC4	Puentes de Hidrógeno
DARF80	-3	-3,7	G3CO, F7CO, M8CO, D11CO, D13CO	NHC2, OHC3, OHC4, K8NH ₃ , OHC6	Puentes de Hidrógeno
DARF81	-3,1	-2,9	G9CO, P14CO, V15CO, V15CO	OHC6, T4OH, T3NH, T3OH	Puentes de Hidrógeno

Tabla 3-1: Continuación

Péptido	Promedio Energía (kcal/mol)	Energía mejor conformación (kcal/mol)	Residuos que intervienen en el péptido	Residuos que intervienen en el antígeno	Tipo de interacción
DARF85	-3,1	-3,2	P6CO, N11CO	OHC3, K8NH ₃	Puentes de Hidrógeno
DARF86	-3,1	-3,6	P4CO, D13CO, S17CO	OHC4, NHC2, T4NH	Puentes de Hidrógeno
DARF88	-3,1	-3,0	G9CO, N11CO, D16CO	OHC6, OHC6, NHC2	Puentes de Hidrógeno
DARF90	-3,4	-3,8	Q2CO, L8CO, D16CO	T4NH, OHC3, OHC6	Puentes de Hidrógeno
DARF92	-3,3	-3,7	L1CO, D13CO	OHC3, OHC4	Puentes de Hidrógeno
DARF96	-3,4	-3,5	Q2COO, P6CO, P10CO, D16CO	OHC3, OHC4, K8NH ₃ , T3CO	Puentes de Hidrógeno

3.3.3 Desempeño péptido DARF51

En promedio este péptido tuvo un total de dos interacciones por conformación. Los residuos que mas interactuaron con el anillo de piranosa del antígeno fueron Gln2, Gly3, Gly9 y Asn11, siendo todos estos residuos reportados en la literatura (53–57) que interactúan con el antígeno. También se destacan las interacciones que tuvieron los aminoácidos Gln2, Gly3, Gly12, Asp16 y Ser17 con los aminoácidos Thr3, Th4 y Lys8 de la cadena lateral del glucido utilizado como receptor, esto es importante porque los datos experimentales muestran una mejor interacción entre las estructuras cuando los aminoácidos de la glicoproteína están involucrados, siendo la interacción entre un Asp de la proteína y una Lys del receptor las que mas contribuyen a esa estabilización.

3.3.4 Desempeño péptido DARF53

Para este péptido se obtuvo en promedio un total de 2 interacciones por conformación, siendo los residuos Gly3, Gln8, Asp13 y Asp16 los que más interactuaron con el anillo de piranosa del antígeno Tn. Se observaron pocas interacciones entre los residuos del péptido y los residuos de la glicoproteína siendo el Asp11 el que más interactuó con la Lys8. Para este péptido se observa que la mayor cantidad de interacciones se da entre los extremos C terminal y N terminal del esqueleto del mismo, sin que las cadenas laterales interactúen.

3.3.5 Desempeño péptido DARF62

Para el péptido DARF62 se observó que los aminoácidos Gln2, Gly9, Asp13 y Tyr17 fueron los que más interactuaron con el antígeno Tn y tuvieron un promedio de 4 interacciones por conformación. Para este péptido destaca la interacción entre el grupo hidroxilo de la cadena lateral en el aminoácido Tyr17 con diferentes carbonos en el anillo de galactopiranososa.

3.3.6 Desempeño péptido DARF65

En términos generales este péptido tuvo un rendimiento superior a otros de los péptidos analizados, con un promedio de 5 interacciones por conformación. Dentro de los residuos involucrados tenemos que la Leu5, Gly9, Asn11, Asp13 y Ser17 están involucrados en diferentes interacciones entre el péptido y los carbonos en el anillo de galactopiranososa del antígeno. Adicionalmente en algunas conformaciones se observa que otros aminoácidos no reportados en la literatura como la Metionina y la prolina interactúan con estos grupos hidroxilo o con algunos residuos de la cadena proteica de la glicoproteína para estabilizar las dos estructuras Figura 3-2.

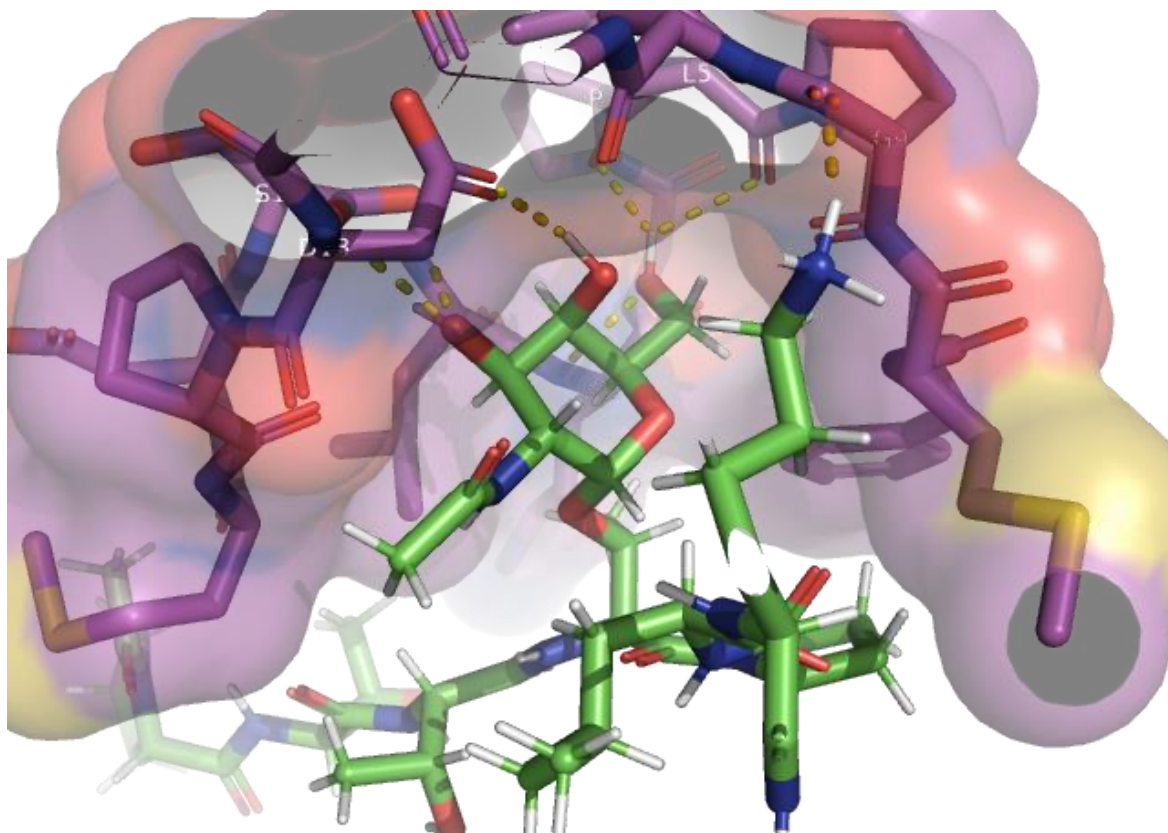


Figura 3-2: Interacción entre el péptido DARF65 y el antígeno Tn. A la derecha se observan los diferentes residuos que interactúan con el anillo de piranosa de la estructura y a la izquierda como el antígeno se acopla al bolsillo del péptido.

3.3.7 Desempeño péptido DARF68

Este péptido tubo un promedio de 3 interacciones por conformación siendo las que mas destacan que involucran los aminoácidos Gln2, Gly9, Asp13 y Asp16 las que mas destacan. Al igual que con los demás péptidos se observa que el mismo interactua con los residuos Thr3, Thr4 y Lys8 de la glicoproteína que se utilizo como ligando.

3.3.8 Desempeño péptido DARF80

Al igual que el péptido DARF65, este péptido fue uno de los que mejor rendimiento presentó en términos de número de interacciones por conformación. Para este péptido destacan las interacciones entre los residuos Gly3, Gly12 y Asp13 con diferentes grupos

hidroxilo en el antígeno, en menor medida otros aminoácidos como Phe7, Met8, Asn11 y la Pro14 contribuyen a que el antígeno se mantenga en el bolsillo de manera estable Figura 3-3.

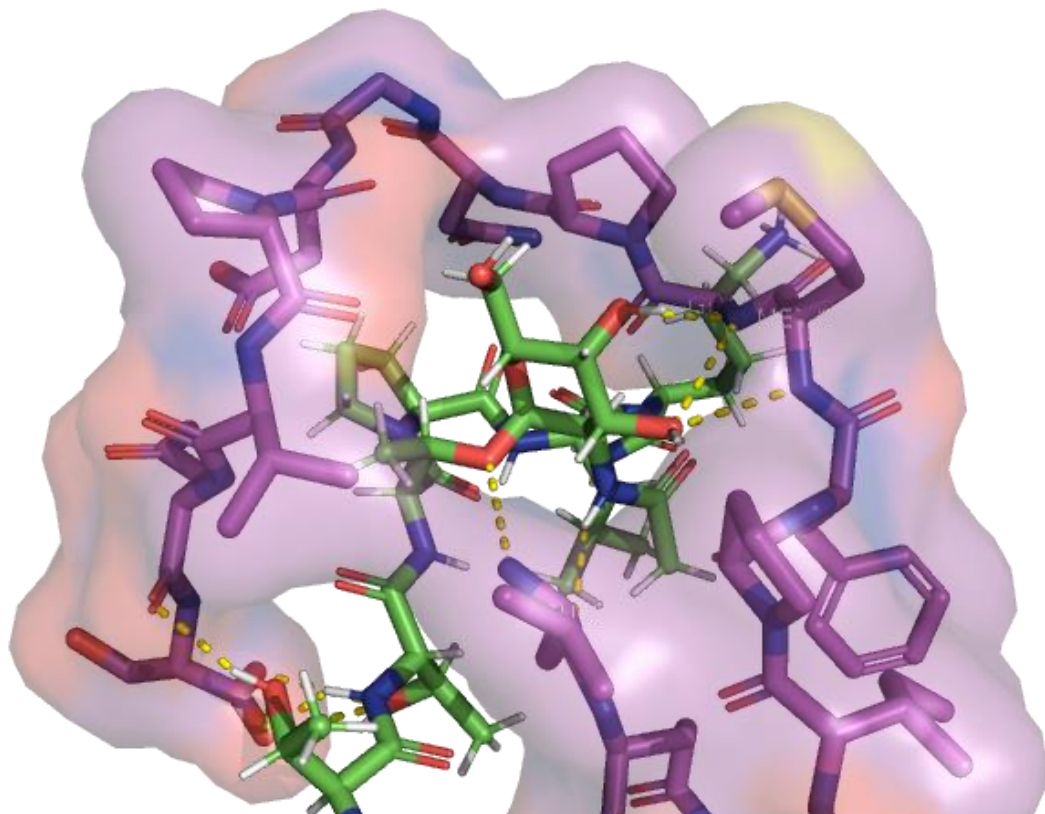


Figura 3-3: Interacción entre el péptido DARF80 y el antígeno Tn. A la derecha se observan los diferentes residuos que interactúan con el anillo de piranosa de la estructura y a la izquierda como el antígeno se acopla al bolsillo del péptido.

3.3.9 Desempeño péptido DARF3

El análisis de las diferentes conformaciones de este péptido se realizó para tener una idea aproximada del comportamiento de los péptidos diseñados a partir de 20 aminoácidos. Sin embargo, al igual que con los péptidos DARF65 y DARF80, este resultó ser uno de los mejores péptidos en términos de la cantidad de interacciones que se encontraron entre el antígeno y él mismo. Dentro de los aminoácidos que interactúan con las diferentes componentes del antígeno Tn utilizado en el estudio destacan Gly5, Gly9, Asp13, Asp15 y Asn18 los cuales interactúan en la mayoría de las conformaciones con el anillo de la

galactopiranososa. Otros residuos como Leu3, Gly9, Ser19 y Ala20 también contribuyen al buen desempeño de este péptido Figura 3-4.

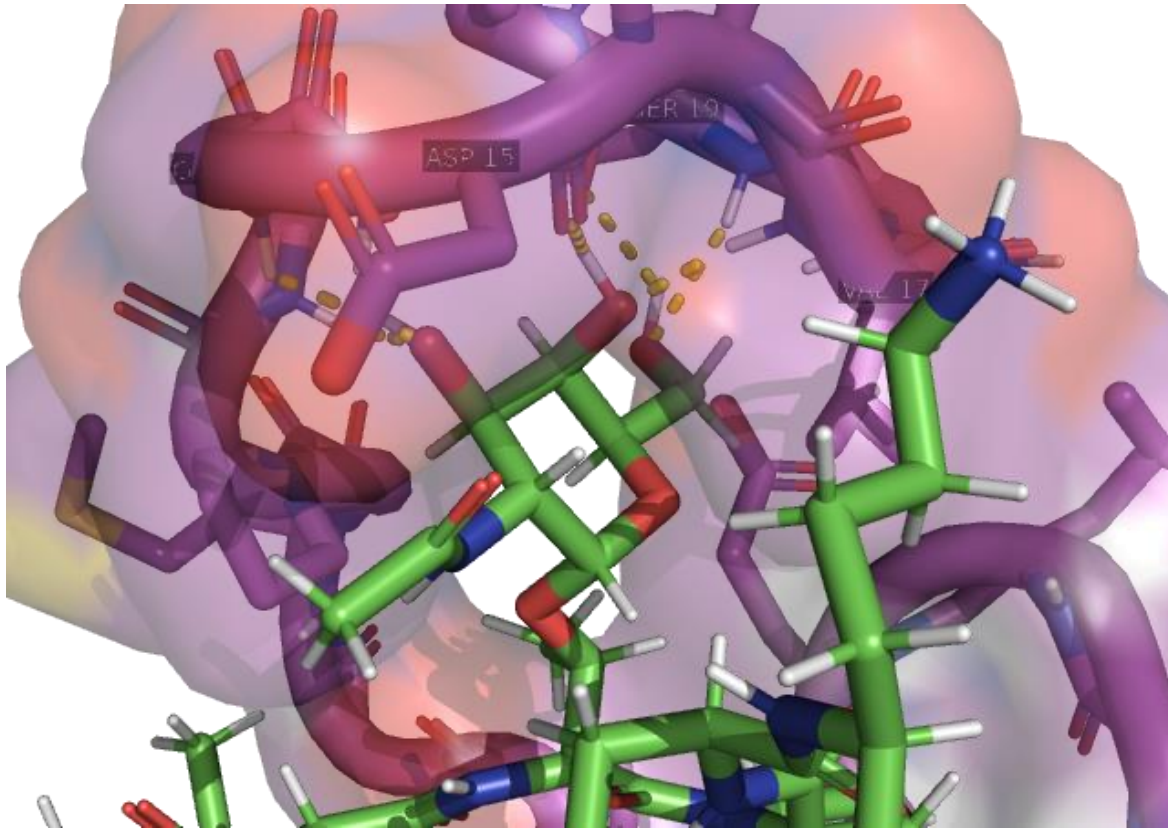


Figura 3-4: Interacción entre el péptido DARF3 y el antígeno Tn. A la derecha se observan los diferentes residuos que interactúan con el anillo de piranososa de la estructura y a la izquierda como el antígeno se acopla al bolsillo del péptido.

4. Conclusiones

De los 20 péptidos diseñados y probados a partir de 20 residuos de aminoácidos ninguno logro superar el umbral de energía libre establecido (-3 kcal/mol) para ser considerado un péptido promisorio; siendo el único péptido que se acerco a este umbral el péptido DARF3.

Aunque no es muy clara la razon por la cual el péptido DARF3 no obtuvo un promedio de energía libre mucho mayor, creemos que esto se debe a como el algoritmo de AutoDock Vina funciona; ya que el mismo selecciona un posicion inicial al azar y a partir de ella empieza a buscar las diferentes conformaciones.

Todos los péptidos creados a partir de 17 residuos de aminoácidos lograron superar el umbral de energia libre que se establecio en el estudio, sin embargo tan solo los péptidos DARF53, DARF62, DARF65, DARF68, DARF80 presentaron interacciones que se asemejan a las interacciones que observan en las proteínas nativas.

Debido a su alto numero de interacciones, se proponen los pépetidos DARF3, DARF65 y DARF80 como los mas promisorios para la identificación del antígeno Tn en pruebas *in vivo*.

4.1 Recomendaciones

Se recomienda realizar esta misma metodología con los péptidos que no fueron probados con el fin de captar otros posibles candidatos para la identificación del antígeno Tn.

Una vez que se sinteticen los péptidos se deben de realizar pruebas in vivo con el fin de saber si los mismos identifican el antígeno Tn.

También se deben crear nuevos péptidos mas pequeños usando la información de la que se dispone; omitiendo esta vez los aminoácidos con anillo aromático, esto se debe a que los mismos no cumplen su función dentro de la estructura de los péptidos y poseen una cadena lateral muy grande. Así mismo es recomendable utilizar un menor número de prolinas en la estructura con el fin de facilitar la síntesis.

A. Anexo: Scrip Perl mutagenesis

```
#!/usr/bin/perl
#Daniel Armando Romero Frenchy
#Maestría en Bioinformática

use strict;
use warnings;
use Fatal qw(open close);

#-----Cuerpo principal -----
#variable que almacena mi cadena Proteínas
my $PROT = 'AIRQGPIPFPGPDGPRNSV'; #se coloca la secuencia que se desee mutar
my $i; #variable inicia mi contador
my $mutant; #variable que almacena mis mutantes
srand(time|$$); #semilla que inicia el proceso de randomizar

open my $OUT1, ">darfall_mutate.txt"; #archivo salida del mutante

#ciclo for que genera los archivos mutantes teniendo como argumento mi cadena PROT y
utilizando la subrutina mutate_better

for ($i=1; $i < 50; $i++){
    $mutant = mutate_better($PROT);
    open my $OUT2, ">darfmutate$i.txt";
```

```

print $OUT2 ">darfmutate$i\n$mutant\n";
print $OUT1 ">darf$i\n$mutant\n";
}
exit;
# salir del programa.

#_____Subrutinas_____

#subrutina que escoge un aminoacido y una posición al azar y la cambia por otro en la
cadena $prot
sub mutate_better {
my ($prot) = @_;
my (@aminoacids) = ('N', 'Q', 'D', 'E', 'S', 'M', 'I', 'L', 'G', 'P', 'F', 'Y');
my ($position) = randomposition($prot); #selecciona una posición
my ($newaa); #selecciona un aminoacido

#ciclo que permite seleccionar un aminoacido al azar luego realiza un control con el fin de
que el mismo no corresponda al aminoacido que se quiere cambiar.
do {
$newaa = randomaminoacid(@aminoacids);
}
until ($newaa ne substr($prot,$position,$i));

substr($prot,$position,1,$newaa);
return $prot;
}

#subrutina que permite seleccionar un aminoacido al azar
sub randomaminoacid {
my @aminoacids = ('N', 'Q', 'D', 'E', 'S', 'M', 'I', 'L', 'G', 'P', 'F', 'Y');
my (@aa) = @_; #arreglo global que va a contener los aminoácidos elegidos al azar

```

```
return $aa[rand @aminoacids];  
}
```

#elige un elemento al azar a partir de la subrutina randomaminoacid

```
sub randomelement {  
my (@nucs) = @_;  
return $nucs[rand @nucs];  
}
```

#elige una posición al azar de la cadena

```
sub randomposition {  
return int(rand(20)); #Se debe cambiar el numero a 17 cuando se muten los otros péptidos  
}
```

B. Anexo: Conformaciones de los péptidos mas promisorios.

Ver Documento excel.

Bibliografía

1. Lisowska E. Tn antigens and their significance in oncology. *Acta Biochim Pol.* 1995;42(1):11–8.
2. Dahr W, Uhlenbruck G, Gunson HH, Van Der Hart M. Molecular Basis of Tn-Polyagglutinability. *Vox Sang.* 1975;29(1):36–50.
3. Ju T, Otto VI, Cummings RD. The Tn antigen-structural simplicity and biological complexity. *Angew Chemie - Int Ed.* 2011;50:1770–91.
4. Marth C, Daxenbichler G. Peanut Agglutinin Inhibits Proliferation of Cultured Breast Cancer Cells. *Oncology.* 1988;45:47–50.
5. Ryder, Stephen D. Smith JA, Rhodes JM. Peanut lectin: A mitogen for normal human colonic epithelium and human HT29 colorectal cancer cells. *J Natl Cancer Inst.* 1992;84(18):1410–6.
6. Kataoka T, Oh-Hashi F, Sakurai Y. BLASTOGENIC POTENCY OF CONCANAVALIN-A-BOUND L1210 LEUKEMIC VACCINE ASSOCIATED WITH ITS IMMUNOGENIC ACTIVITY. *Gann.* 1979;70:155–64.
7. Yoshimoto R, Kondoh N, Isawa M aki, Hamuro J. Plant lectin, ATF1011, on the tumor cell surface augments tumor-specific immunity through activation of T cells specific for the lectin. *Cancer Immunol Immunother.* 1987;25:25–30.
8. Kataoka T, Oh-Hashi F, Akabori Y, Sakurai Y, Okabe M, Gomi K. TUMOR DEPENDENCY OF CONCANAVALIN A-INDUCED POTENTIATION OF TUMOR CELL IMMUNOGENICITY*1. *Gann.* 1983;74(3):412–8.
9. Fu LL, Zhou CC, Yao S, Yu JY, Liu B, Bao JK. Plant lectins: Targeting programmed cell death pathways as antitumor agents. *Int J Biochem Cell Biol.* 2011;43:1442–9.

10. Numata Y, Nakada H, Fukui S, Kitagawa H, Ozaki K, Inoue M, et al. A monoclonal antibody directed to Tn antigen. *Biochem Biophys Res Commun*. 1990;170(3):981–5.
11. Morita N, Yajima Y, Asanuma H, Nakada H, Fujita-Yamaguchi Y. Inhibition of cancer cell growth by anti-Tn monoclonal antibody MLS128. *Biosci Trends*. 2009;3(1):32-37.
12. Zamri N, Masuda N, Oura F, Yajima Y, Nakada H, Fujita-Yamaguchi Y. Effects of two monoclonal antibodies, MLS128 against Tn-antigen and 1H7 against insulin-like growth factor-I receptor, on the growth of colon cancer cells. *Biosci Trends*. 2012;6(6):303–12.
13. Nakada H, Numata Y, Inoue M, Tanaka N, Kitagawa H, Funakoshi I, et al. Elucidation of an essential structure recognized by an anti-GalNAc-Ser(Thr) monoclonal antibody (MLS 128). *J Biol Chem*. 1991;266(19):12402–5.
14. Kubota T, Matsushita T, Niwa R, Kumagai I, Nakamura K. Novel anti-Tn single-chain Fv-Fc fusion proteins derived from immunized phage library and antibody Fc domain. *Anticancer Res*. 2010;30(9):3397–405.
15. Osinaga E, Bay S, Tello D, Babino A, Pritsch O, Assemat K, et al. Analysis of the fine specificity of Tn-binding proteins using synthetic glycopeptide epitopes and a biosensor based on surface plasmon resonance spectroscopy. *FEBS Lett*. 2000;469(1):24–8.
16. Bhutia SK, Panda PK, Sinha N, Praharaj PP, Bhol CS, Panigrahi DP, et al. Plant lectins in cancer therapeutics : Targeting apoptosis and autophagy- dependent cell death. *Pharmacol Res*. 2019;144(April):8–18.
17. Poiroux G, Barre A, Simplicien M, Pelofy S, Bruno S, Damme EJM Van, et al. Morniga-G , a T / Tn-Specific Lectin , Induces Leukemic Cell Death via Caspase and DR5 Receptor-Dependent Pathways. :1–16.
18. Poiroux G, Pitie M, Peumans WJ, Bernadou J, Levade T, Rouge P. Morniga G : A Plant Lectin as an Endocytic Ligand for Photosensitizer Molecule Targeting Toward Tumor-Associated T/Tn Antigens. 2011;370–7.
19. Poiroux G, Barre A, van Damme EJM, Benoist H, Rougé P. Plant lectins targeting

- O-glycans at the cell surface as tools for cancer diagnosis, prognosis and therapy. *Int J Mol Sci.* 2017;18(6).
20. Springer GF. T and Tn, general carcinoma autoantigens. *Science* (80-). 1984;224(4654):1198–206.
 21. Limpias Catalina PG, Vega Nohora ROAJ. DETECCIÓN DEL ANTÍGENO Tn EN TUMORES EPITELIALES CON LA LECTINA DE *Vicia villosa* isolectina B4. *Rev Fac Med.* 2010;58(4):293–305.
 22. Freire T, Osinaga E. Immunological and biomedical relevance of the Tn antigen. *Inmunología.* 2003;22(1):27–38.
 23. Sierra AY, Pérez G. Extracción, purificación y caracterización de dos lectinas en semillas de *Dioclea sericea*. *Rev Acad Colomb Ciencias.* 1999;23(88):445–54.
 24. Chen Y-F, Boland CR, Kraus E, Goldstein IJ. The lectin *Griffonia simplicifolia* I-A4 (GS I-A4) specifically recognizes terminal linked N-acetylgalactosaminyl groups and is cytotoxic to the human colon cancer cell lines LS174t and SW1116. *Int J Cancer.* 1994;57(4):561–7.
 25. Medeiros A, Berois N, Incerti M, Bay S, Franco Fraguas L, Osinaga E. A Tn antigen binding lectin from *Myrsine coriacea* displays toxicity in human cancer cell lines. *J Nat Med.* 2013;67:247–54.
 26. Pérez G, Vega N, Fernández-Alonso JL. LECTIN PROSPECTING IN COLOMBIAN LABIATAE. A SYSTEMATIC-ECOLOGICAL APPROACH -II. Prospección de lectinas en especies de labiadas colombianas. Un enfoque sistemático-ecológico-II. *Caldasia.* 2006;28(2):179–95.
 27. Fernández-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.
 28. Loureiro LR, Carrascal MA, Barbas A, Ramalho JS, Novo C, Delannoy P, et al. Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules.* 2015;1783–809.
 29. Zlocowski N, Grupe V, Garay YC, Nores GA, Ricardo DL, Irazoqui FJ. Purified

- human anti-Tn and anti-T antibodies specifically recognize carcinoma tissues. *Nat Sci Rep*. 2019;(May):1–11.
30. Hernandez Cruz P, Pérez Campos E, Martínez Martínez L, Ortiz B, Martínez G. LAS LECTINAS VEGETALES COMO MODELO DE INTERACCIONES PROTEÍNA CARBOHIDRATO. *REB*. 2005;24(1):21–7.
 31. Fern Bah CS, Fei Fang E, Bun Ng T. Antitumor potential and other emerging medicinal properties of natural compounds. In: Fei Fang E, Bun Ng T, editors. *Antitumor Potential and other Emerging Medicinal Properties of Natural Compounds*. eBook. New York, London: Springer Science+Business Media Dordrecht; 2013. p. 55–74.
 32. Vega N, Pérez G. Isolation and characterisation of a *Salvia bogotensis* seed lectin specific for the Tn antigen. *Phytochemistry*. 2006;67:347–55.
 33. Barroso P, Murcia H, Vega N, Pérez G. Obtención y purificación de IgY dirigidas contra la lectina de *Salvia bogotensis*. *Biomédica*. 2005;25:496–510.
 34. Vega N, Murcia H, Pérez G. CARACTERIZACIÓN DE IGYS ANTI-LECTINA DE *Salvia bogotensis* Y SU APLICACIÓN EN ESTUDIOS CITOQUÍMICOS PARA LA DETECCIÓN DEL ANTÍGENO TN. *Rev Colomb Química*. 2009;38(3):363–77.
 35. Hu D, Tateno H, Hirabayashi J. Lectin engineering, a molecular evolutionary approach to expanding the lectin utilities. *Molecules*. 2015;20:7637–56.
 36. Kuwabara N, Hu D, Tateno H, Makyio H, Hirabayashi J, Kato R. Conformational change of a unique sequence in a fungal galectin from *Agrocybe cylindracea* controls glycan ligand-binding specificity. *FEBS Lett*. 2013;587:3620–5.
 37. Arango R, Rodriguez-Arango E, Adar R, Belenky D, Loontjens FG, Rozenblatt S, et al. Modification by site-directed mutagenesis of the specificity of *Erythrina corallodendron* lectin for galactose derivatives with bulky substituents at C-2. *FEBS Lett*. 1993;330(2):133–6.
 38. Yamamoto K, Konami Y, Osawa T, Irimura T. Alteration of the Carbohydrate-Binding Specificity of the *Bauhinia purpurea* Lectin through the Preparation of a Chimeric Lectin. *J Biochem*. 1992;111(1):87–90.

39. Yamamoto K, Konami Y, Osawa T. A Chimeric Lectin Formed from Bauhinia purpurea Lectin and Lens culinaris Lectin Recognizes a Unique Carbohydrate Structure. *J Biochem.* 2000;127(1):129–35.
40. Jordan ET, Goldstein IJ. Site-directed mutagenesis studies on the lima bean lectin. Altered carbohydrate-binding specificities result from single amino acid substitutions. *J Biochem.* 1995;230:958–64.
41. Nishiguchi M, Yoshida K, Sumizono T, Tazaki K. Studies by site-directed mutagenesis of the carbohydrate-binding properties of a bark lectin from Robinia pseudoacacia. *FEBS Lett.* 1997;403:294–8.
42. Geer LY, Marchler-bauer A, Geer RC, Han L, He J, He S, et al. The NCBI BioSystems database. *Nucleic Acids Res.* 2010;38(October 2009):492–6.
43. Bateman A, Martin MJ, O’Donovan C, Magrane M, Alpi E, Antunes R, et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* 2017;45(D1):D158–69.
44. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000;28(1):235–42.
45. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2014;7(1):539–539.
46. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, et al. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res.* 2010;38(SUPPL. 2):695–9.
47. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 2009;25(9):1189–91.
48. ULC CCG. Molecular Operating Environment (MOE). 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada; 2013.
49. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics.* 2006;22(21):2695–6.

50. 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):293–304.
51. Delbaere LTJ, Vandonselaar M, Prasad L, Quail JW, Wilson KS, Dauter Z. Structure of the Lectin IV of *Griffonia simplicifolia* and its Complex with the Lewis b Human Blood Group Determinant at 2.0 Å Resolution. *J Mol Biol.* 1993;230(3):950–65.
52. Hamelryck TW, Loris R, Bouckaert J, Dao-Thi M-H, Strecker G, Imberty A, et al. Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus* 11 Edited by R. Huber. *J Mol Biol.* 1999;286(4):1161–77.
53. 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;536:106–10.
54. 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.
55. Kulkarni KA, Sinha S, Katiyar S, Surolia A, Vijayan M, Suguna K. Structural basis for the specificity of basic winged bean lectin for the Tn-antigen: A crystallographic, thermodynamic and modelling study. *FEBS Lett.* 2005;579:6775–80.
56. Madariaga D, Martinez-Sáez N, Somovilla VJ, Coelho H, Valero-González J, Castro-Lopez J, et al. Detection of tumor-associated glycopeptides by Lectins: The peptide context modulates carbohydrate recognition. *ACS Chem Biol.* 2015;10(3):747–56.
57. Lubkowski J, Durbin S V., Silva MCC, Farnsworth D, Gildersleeve JC, Oliva ML V., et al. Structural analysis and unique molecular recognition properties of a *Bauhinia forficata* lectin that inhibits cancer cell growth. *FEBS Lett.* 2017;284(3):429–50.

58. Haji-Ghassemi O, Gilbert M, Spence J, Schur MJ, Parker MJ, Jenkins ML, et al. Molecular basis for recognition of the cancer glycobiomarker, LacdiNAc (GalNAc[β 1 \rightarrow 4]GlcNAc), by Wisteria floribunda Agglutinin. *J Biol Chem.* 2016;291(46):24085–95.
59. Elgavish S, Shaanan B. Structures of the Erythrina corallodendron lectin and of its complexes with mono- and disaccharides. *J Mol Biol.* 1998;277(4):917–32.
60. Ultrastructure B. The Purification , Composition , and Specificity Lectin from Peanut (*Arachis hypogaea*) * of the Anti-T. 1975;(21).
61. Wu AM, Wu JH, Yang Z, Singh T, Goldstein IJ, Sharon N. Differential contributions of recognition factors of two plant lectins e *Amaranthus caudatus* lectin and *Arachis hypogaea* agglutinin , reacting with Thomsen-Friedenreich disaccharide (Gal b 1 e 3GalNAc a 1 e Ser / Thr). *Biochimie.* 2008;90(11–12):1769–80.
62. Science E, All PB V. Use of O-glycosylation-defective human lymphoid cell lines and flow cytometry to delineate the specificity of *Moluccella laevis* lectin and monoclonal antibody 5F4 for the Tn antigen. 1993;36:239–43.
63. Debray H, Jaekel H, Garfias Y, Jime C, Herna P, Zenteno E, et al. Chemical characterization of the lectin from *Amaranthus leucocarpus* syn . *hypocondriacus* by 2-D proteome analysis. 2002;321–9.
64. Singh T, Wu JH, Peumans WJ, Roug P, Damme EJM Van, Wu AM. Recognition profile of *Morus nigra* agglutinin (*Morniga G*) expressed by monomeric ligands , simple clusters and mammalian polyvalent glycotopes. 2007;44:451–62.
65. Tachibana K, Nakamura S, Wang H, Iwasaki H, Tachibana K, Maebara K, et al. Elucidation of binding specificity of Jacalin toward O-glycosylated peptides : quantitative analysis by frontal affinity chromatography. 2006;16(1):46–53.
66. Singh T, Chatterjee U, Wu JH, Chatterjee BP, Wu AM. Carbohydrate recognition factors of a T a (Gal b 1 3GalNAc a 1 Ser / Thr) and Tn (GalNAc a 1 Ser / Thr) specific lectin isolated from the seeds of *Artocarpus lakoocha*. 2005;15(1):67–78.
67. Tsai M, Herp A. Lectinochemical studies on the binding properties of a toxic lectin (ricin) isolated from the seeds of *Ricinus communis*. 2005;8966(259):530–42.

68. Wu AM, Wu JH, Liu J, Singh T. Recognition profile of Bauhinia purpurea agglutinin (BPA). 2004;74:1763–79.
69. Wu AM. Lectinochemical studies on the glyco-recognition factors of a Tn (GalNAc a 1 fi Ser / Thr) specific lectin isolated from the seeds of Salvia sclarea. 2005;2118456:167–8.
70. Zenteno E. Purification of a N-acetyl-d-galactosamine specific lectin from the orchid Laelia autumnalis. Phytochemistry. 1995;40(3):651–5.
71. Due MP, Example F. Mixed-Field Polyagglutinability Due to Tn: A Further Example. 1975;15(1).
72. Wu AM. Polyvalent GalNAc a 1 Ser / Thr (Tn) and Gal b 1 3GalNAc a fi Ser / Thr (T a) as the most potent recognition factors involved in Maclura pomifera agglutinin – glycan interactions. 2005;135–6.
73. Svergun DI, Maveyraud L, Niwa H, Konarev P V, Palmer RA, Peumans WJ, et al. Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from Sambucus nigra. 2009;89–103.
74. Hamelryck T, Fernandez EJ, Hamelryck TW, Loris R, Bouckaert J. Carbohydrate binding , quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from Dolichos biflorus. J Mol Biol. 2015;(July).
75. Loris R, Hamelryck T, Bouckaert J, Wyns L. Legume lectin structure. Vol. 1383, Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology. 1998. p. 9–36.
76. Chandra NR, Prabu MM, Suguna K, Vijayan M. Structural similarity and functional diversity in proteins containing the legume lectin fold. Protein Eng. 2001 Nov;14(11):857–66.
77. Farhadi T, Hashemian SMR. Computer-aided design of amino acid-based therapeutics: A review. Drug Des Devel Ther. 2018;12:1239–54.
78. Mócsai A, Kovács L, Gergely P. What is the future of targeted therapy in rheumatology: Biologics or small molecules? BMC Med. 2014;12(1):1–9.
79. Samish I, MacDermaid CM, Perez-Aguilar JM, Saven JG. Theoretical and Computational Protein Design. Annu Rev Phys Chem. 2011;62(1):129–49.

80. King NP, Sheffler W, Sawaya MR, Vollmar BS, John P, André I, et al. Computational design of self-assembling protein nanomaterials with atomic level accuracy. 2012;336(6085):1171–4.
81. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of Protein Tertiary Structures from Fragments with Similar Local Sequences using Simulated Annealing and Bayesian Scoring Functions. 1997;
82. Pearce R, Huang X, Setiawan D, Zhang Y. EvoDesign : Designing Protein – Protein Binding Interactions Using Evolutionary Interface Profiles in Conjunction with an Optimized Physical Energy Function. *J Mol Biol.* 2019;431(13):2467–76.
83. Chevalier A, Silva D, Rocklin GJ, Hicks DR, Vergara R, Murapa P, et al. Massively parallel de novo protein design for targeted therapeutics. *Nat Publ Gr.* 2017;550(7674):74–9.
84. Vanhee P, van der Sloot AM, Verschueren E, Serrano L, Rousseau F, Schymkowitz J. Computational design of peptide ligands. *Trends Biotechnol.* 2011;29(5):231–9.
85. Roberts NA, Martin JA, Kinchington D, Broadhurst A V., Craig JC, Duncan IB, et al. Rational design of peptide-based HIV proteinase inhibitors. *Science (80-).* 1990;248(4953):358–61.
86. Yoo B, Billig A-M, Medarova Z. Guidelines for Rational Cancer Therapeutics. *Front Oncol.* 2017;7(December):1–8.
87. Karagiannis ED, Urbanska AM, Sahay G, Pelet JM, Jhunjunwala S, Langer R, et al. Rational design of a biomimetic cell penetrating peptide library. *ACS Nano.* 2013;7(10):8616–26.
88. Ran X, Burchfiel ET, Dong B, Rettko NJ, Donyak BM, Shao H, et al. Rational design and screening of peptide-based inhibitors of heat shock factor 1 (HSF1). *Bioorganic Med Chem.* 2018;4:1–8.
89. Wang H, Yapa AS, Kariyawasam NL, Shrestha TB, Kalubowilage M, Wendel SO, et al. Rationally designed peptide nanosponges for cell-based cancer therapy. *Nanomedicine Nanotechnology, Biol Med.* 2017;13(8):2555–64.
90. Fang RH, Luk BT, Hu CMJ, Zhang L. Engineered nanoparticles mimicking cell membranes for toxin neutralization. *Adv Drug Deliv Rev.* 2015;90:69–80.

91. Lei Y, Tang L, Xie Y, Xianyu Y, Zhang L, Wang P, et al. Gold nanoclusters-assisted delivery of NGF siRNA for effective treatment of pancreatic cancer. *Nat Commun.* 2017;8:1–15.
92. Marqus S, Pirogova E, Piva TJ. Evaluation of the use of therapeutic peptides for cancer treatment. *J Biomed Sci.* 2017;24(1):1–15.
93. Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: An online resource for de novo peptide structure prediction. *Nucleic Acids Res.* 2010;37(SUPPL. 2):498–503.
94. Maupetit J, Derreumaux P, Tufféry P. A Fast Method for Large-Scale De Novo Peptide and Miniprotein Structure Prediction. *J Comput Chem.* 2010;31(4):726–38.
95. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: An updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* 2012;40(W1):288–93.
96. Shen Y, Maupetit J, Derreumaux P, Tufféry P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J Chem Theory Comput.* 2014;10(10):4745–58.
97. Ciemny M, Kurcinski M, Kamel K, Kolinski A, Alam N, Schueler-Furman O, et al. Protein–peptide docking: opportunities and challenges. *Drug Discov Today.* 2018;23(8):1530–7.
98. Watkins AM, Bonneau R, Arora PS. Modeling Peptide-Protein Interactions. Vol. 1561, *Methods in Molecular Biology.* 2017.
99. Verschueren E, Vanhee P, Rousseau F, Schymkowitz J, Serrano L. Protein-peptide complex prediction through fragment interaction patterns. *Structure.* 2013;21(5):789–97.
100. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. Templates are available to model nearly all complexes of structurally characterized proteins. *Proc Natl Acad Sci U S A.* 2012;109(24):9438–41.
101. Antes I. DynaDock: A now molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins Struct Funct Bioinforma.* 2010;78(5):1084–104.
102. Donsky E, Wolfson HJ. PepCrawler: A fast RRT-based algorithm for high-

- resolution refinement and binding affinity estimation of peptide inhibitors. *Bioinformatics*. 2011;27(20):2836–42.
103. London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O. Rosetta FlexPepDock web server - High resolution modeling of peptide-protein interactions. *Nucleic Acids Res*. 2011;39(SUPPL. 2):249–53.
104. Rentzsch R, Renard BY. Docking small peptides remains a great challenge: An assessment using AutoDock Vina. *Brief Bioinform*. 2015;16(6):1045–56.
105. Hauser AS, Windshügel B. LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *J Chem Inf Model*. 2016;56(1):188–200.
106. Antunes DA, Devaurs D, Kaviraki LE. Understanding the challenges of protein flexibility in drug design. *Expert Opin Drug Discov*. 2015;10(12):1301–13.
107. Morris G, Huey R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J ...*. 2009;30(16):2785–91.
108. Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastiris PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol*. 2016;428(4):720–5.
109. Borgert A, Heimburg-Molinaro J, Song X, Lasanajak Y, Ju T, Liu M, et al. Deciphering Structural Elements of Mucin Glycoprotein Recognition. *ACS Chem Biol*. 2012;7:1031–9.
110. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.;
111. Schrödinger L. The PyMOL Molecular Graphics System.