



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

FACULTAD DE CIENCIAS

Caracterización del desplegamiento térmico de la  
aglutinina de germen de trigo mediante simulación de  
dinámica molecular

**T E S I S**

QUE PARA OBTENER EL TÍTULO DE:

Biólogo

PRESENTA:

**Mateo Calderón Vargas**

TUTOR

**Dr. Luis Fernando Cofas Vargas**

Cd. Mx. 2023





Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

# Agradecimientos

La presente tesis fue realizada en el Departamento de Química de Biomacromoléculas del Instituto de Química de la Universidad Nacional Autónoma de México, bajo la dirección del Dr. Luis Fernando Cofas Vargas y la asesoría del Dr. Enrique García Hernández. Durante la realización de este trabajo, se contó con el apoyo económico de la Dirección General de Asuntos del Personal Académico (DGAPA-UNAM) a través del Proyecto PAPIIT IN206221.

Muchas gracias al Dr. Luis Fernando Cofas, quién ha sido más que un director de tesis para mí. Sin él, la realización de este proyecto no habría sido posible, ya que su disposición y conocimientos fueron fundamentales para este proyecto y mi formación. Al Dr. Enrique García, a quien le debo la entrada al grupo de investigación y al Instituto de Química y quien también nos guió de forma muy cercana durante este proyecto. Muchas gracias a mi familia por apoyarme en mis decisiones y durante este proceso. Por último quisiera agradecerle a mi novia Emilia, quien no sólo fue un apoyo emocional muy importante, sino que en momentos cuando mis ojos ya no podían estar frente a la pantalla, ella escribía mientras yo le dictaba, por lo que parte de este proyecto está escrito por ella.

# Índice general

1. Agradecimientos	2
2. Introducción	5
3. Metodología	17
4. Resultados	22
5. Discusión	43
6. Conclusiones	47
7. Perspectivas	48

## Lista de abreviaturas

Agglutinina de germen de trigo	AGT
Amino terminal	AT
Análisis de componentes principales	PCA
Análisis de componentes principales de ángulos diedros	dPCA
Carboxilo terminal	CT
Componente principal	PC
Modelos de estados de Markov	MSM
Paisaje de energía libre	PEL
Puentes disulfuro	PdS
Puentes de hidrógeno	P-H

## Índice de figuras

Figura	Título	Página
1	Estructura cristalográfica de la AGT	8
2	Curva de estabilidad de proteínas	12
3	Energía potencial respecto a la distancia	14
4	Energía potencial respecto al ángulo	14
5	PEL de los dominios aislados	25
6	Comparación de los estados plegado y desplegado	26
7	RMSD de los dominios a 20° C	28
8	RMSD a la $T_m$ de cada dominio	29
9	$\Delta$ RMSF entre 20° y 120°	32
10	Fracciones de P-H	35
11	Regresión lineal	37
12	RMSD del sitio de unión	38
13	Matrices de contactos	40
14	Valores propios de CP	41
15	RMSD de proteína completa	42
16	Estructuras promedio	43

# Introducción

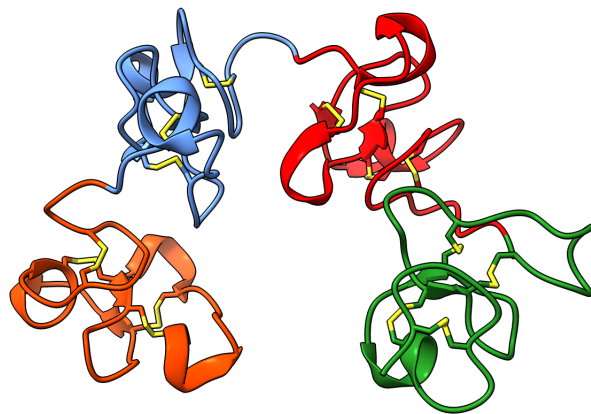
El reconocimiento molecular de carbohidratos por parte de proteínas es un fenómeno recurrente en la naturaleza, el cual permite la realización de varias funciones fisiológicas en diversos organismos. Los anticuerpos, enzimas, proteínas de transporte y lectinas son ejemplos de proteínas con capacidad de interactuar con carbohidratos. En particular las lectinas han captado la atención debido a su capacidad de interactuar con antibióticos, fitohormonas, polifirinas y carbohidratos con una especificidad a nivel micromolar [Leyva et al., 2019]; [Parasuraman et al., 2014]; [Bhattacharyya and Brewer, 1988]. Esta interacción les da la capacidad de aglutinar células, lo que también ha potenciado su uso en la industria e investigación [Cavada et al., 2020]. Las lectinas se encuentran en todo tipo de organismos, desde virus y bacterias hasta plantas y animales [Lis and Sharon, 1998]. En cada uno de estos organismos, las lectinas tienen una función biológica diferente. En plantas tienen un papel importante en germinación, crecimiento, transporte y almacenamiento de carbohidratos, así como en mecanismos de defensa [Kumar et al., 2012]. Microorganismos patógenos cuentan con lectinas en la superficie celular, que actúan como receptores de la célula hospedera y que tienen la finalidad de llevar a cabo la infección [Nizet et al., 2017]. En humanos, estas proteínas forman parte del sistema inmune, ya que pueden adherirse a la membrana celular de bacterias, modificar la porosidad de esas células y provocar lisis celular [Chettri et al., 2021].

En 1880 Peter Hermann Stillmark reportó un factor aglutinante de eritrocitos en el extracto de la planta *Ricinus communis*, el cual fue llamado ricina. Estudios posteriores sobre el extracto de la planta *Abrus precatorius*, reportaron la presencia de otro factor aglutinante llamado abrina [Sharon and Halina, 2004]. Simultáneamente, se encontró que la N-acetilgalactosamina se une a la ricina con gran afinidad, evitando la aglutinación

[Wu et al., 2006]. Un mecanismo similar es el de la abrina con la D-galactosa (Herrman et al., 1980). En 1890, Paul Ehrlich realizó una serie de experimentos en ratones, en los cuales administró dosis letales de ricina, seguidas de D-galactosa, y dosis de abrina, seguidas de N-acetilgalactosamina. Ninguno de los ratones sobrevivió [Sharon and Halina, 2004], lo que indica que la ricina y la abrina no interactúan con las azúcares administradas subsecuentemente. Esto sugiere que las lectinas presentan gran especificidad por ciertos ligandos. En otro estudio se inyectaron varias dosis subletales en modelos animales para cada lectina y se vio que, tras un tiempo, adquirirían inmunidad a la lectina respectiva, lo cual sentó las bases para estudios en inmunología [Sharon and Halina, 2004].

En 1965, [Aub et al., 1965] reportó una proteína con la capacidad de aglutinar células malignas a partir de cambios en la composición del glicocaliz de las células [Chettri et al., 2021]. Se trataba de una lectina proveniente del germen de trigo (*Triticum aestivum*), por lo que posteriormente se le nombró aglutinina de germen de trigo (AGT). Esta lectina tiene como función proteger a la planta contra organismos patógenos e insectos; presenta afinidad por el ácido siálico y N-acetil-D-glucosamina (GlcNAc) [Lis and Sharon, 1998]. Esta última molécula forma parte de la quitina, que es el componente principal del exoesqueleto de insectos y de la membrana celular de hongos filamentosos, bacterias y levaduras [Uehara and Park, 2004]. La unión de la AGT a GlcNAc interrumpe la integridad estructural de la membrana y causa la muerte del organismo [Balčiūnaitė M and Dzikaras, 2021]. La característica aglutinante de estas proteínas proviene de tener múltiples sitios de unión, en los que puede darse el entrecruzamiento de células, causando la precipitación. Esta proteína se encuentra forma de dímero y cada monómero está compuesto por 171 aminoácidos (17 kDa). El plegamiento nativo de estas subunidades forma cuatro regiones de 43 aminoácidos, conocidas como dominios A, B, C y D, respecto al orden de la secuencia en la que se forman, como se muestra en la **Figura 1**.

Los dominios presentan un alto índice de similitud en secuencia, lo que les da la característica de tener un plegamiento idéntico como se muestra en la **Tabla 1**, donde se observa la posición de cada residuo y si se conserva en cada dominio (marcado con \*). Su estructura secundaria consiste en una hélice alfa y dos láminas beta, conectadas por tres puentes disulfuro (PdS), los cuales confieren gran estabilidad [Feige et al., 2018].



**Figura 1. Estructura cristalográfica de la AGT.** (PDB: 2UVO) con la letra correspondiente a cada dominio.

**Tabla 1.** Secuencia de los aminoácidos

de todos los dominios de la AGT. Las cisteínas que forman los PdS se encuentran marcados con amarillo y los residuos que se conservan están marcados con \*.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	
A		Q	R	C	G	E	Q	G	S	N	M	E	C	P	N	N	L	C	C	S	Q	Y	G	Y	C	G	M	G	G	D	Y	C	G	K	G	C	Q	N	G	A	C	W	T	
B	S	K	R	C	G	S	Q	A	G	G	A	T	C	T	N	N	Q	C	C	S	Q	Y	G	Y	C	G	F	G	A	E	Y	C	G	A	G	C	Q	G	G	P	C	R	A	
C	D	I	K	C	G	S	Q	A	G	G	K	L	C	P	N	N	L	C	C	S	Q	W	G	F	C	G	L	G	S	E	F	C	G	G	C	Q	S	G	A	C	S	T		
D	D	K	P	C	G	K	D	A	G	G	R	V	C	T	N	N	Y	C	C	S	K	W	G	S	C	G	I	G	P	G	Y	C	G	A	G	C	Q	S	G	G	C	D	G	
			*	*			*				*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

La AGT cuenta con cuatro sitios de unión a carbohidratos por monómero, los cuáles están formados por 11 residuos: 2 aromáticos, 8 polares neutros y 1 polar cargado, de acuerdo con [Tellez et al., 2011]. La interacción entre el sitio de unión y el carbohidrato está dirigido por puentes de hidrógeno (P-H) e interacciones de van der Waals. El dominio A interacciona con el ligando a través de los residuos Ser<sup>19</sup>, Tyr<sup>21</sup>, Tyr<sup>23</sup>, Tyr<sup>30</sup>, utilizando la enumeración correspondiente a la **Tabla 1**. En el dominio B la interacción es con los residuos Ser<sup>19</sup>, Tyr<sup>21</sup>, Tyr<sup>24</sup>, Ala<sup>28</sup>, Glu<sup>29</sup>, Tyr<sup>30</sup>. Por otro lado, el dominio C interacciona con Ser<sup>19</sup>, Trp<sup>21</sup>, Phe<sup>23</sup>, Asp<sup>29</sup>, mientras que el dominio D con Ser<sup>19</sup>, Trp<sup>21</sup>, Ser<sup>23</sup> y Tyr<sup>30</sup>. Todas las Ser correspondientes a la posición 19 forman P-H, mientras que los residuos aromáticos mantienen interacciones electrostáticas de tipo hidrofóbico [Muraki et al., 2002], [Tellez et al., 2011].

Otra peculiaridad de estos sitios de unión es que son una oquedad ubicada en un lugar estratégico del péptido, dado que queda rodeado por tres PdS [Hernández et al., 1995].



Esto podría tener implicaciones en su especificidad y afinidad, debido a la estabilidad y restricción de movimiento que le confieren al sitio de unión [Parasuraman et al., 2014]. No obstante, esto no ha sido corroborado experimentalmente.

Debido a sus propiedades biofísicas y bioquímicas, la AGT ha sido utilizada en tratamientos contra leucemia [Ryva et al., 2019], ingeniería de tejidos [Teuschl et al., 2014], detección de agentes patógenos en fluidos humanos [Murakami et al., 2014] e identificación temprana de diferentes tipos de cáncer [Neutsch et al., 2014], incluso se ha utilizado como marcador para rastrear anatómicamente el desarrollo de la barrera hematoencefálica de ratones [Gómez-González et al., 2011]. En este último estudio la AGT fue conjugada con la enzima peroxidasa de la *Ammocia rusticana* (AGT-PAR) y su actividad fue captada utilizando 3,3'-diaminobenzidina (DMA). Se utilizó la AGT dado que esta se une a la superficie celular mediante el ácido siálico y GlcNAc. A través de endocitosis, la AGT-PAR logra entrar a la célula cerebral endotelial, donde puede ser reconocida por endosomas y por el aparato de Golgi para después hacer exocitosis y ser liberada en el lado abluminal [Gómez-González et al., 2011]. El producto de la reacción de la AGT-PAR con DMA se encuentra en la hendidura perivascular y macrófagos, dentro de 3-12 hrs después de administración intravenosa [Broadwell et al., 1988].

Se ha propuesto utilizar esta proteína como vehículo para acarreamiento de fármacos suministrados por vía oral debido a su capacidad de adherirse a células intestinales y ser internalizada rápidamente en el citoplasma de células enterocíticas humanas [Wirth et al., 2002]; [Dalla Pellegrina et al., 2005]. [Liu et al., 2011] reportaron un mecanismo de endocitosis y exocitosis vía filamentos de actina y microtúbulos, en el que el proceso de endocitosis culmina en el centro de organización de microtúbulos (MTOC), donde también comienza el proceso de exocitosis. Por otro lado, se ha reportado que esta proteína modificada con nanopartículas de polietilenglicol y poliácido láctico facilitó la ingesta de un rastreador fluorescente en el sistema nervioso central [Gao et al., 2006]. Esto asentó las bases para realizar un estudio en el que el mismo grupo de investigadores demostró que la AGT conjugada con nanopartículas lipídicas y suministrada vía intranasal es capaz de atravesar la barrera hematoencefálica y acarrear péptidos vasoactivos intestinales, con el fin de combatir enfermedades neurodegenerativas [Gao et al., 2007]. También se ha utilizado para el acarreamiento de fármacos poco solubles en agua mediante nanopartículas lipídicas [Liu et al., 2011]. Es importante notar que en todos los estudios que se

han realizado con la AGT como vehículo de administración de fármacos, la proteína está conjugada de manera que el ligando llegue al receptor indicado. La necesidad de conjugar a la AGTco para que pueda ser utilizada como acarreador de fármacos, abre la posibilidad de poder mejorar este mecanismo y hacerlo más práctico y accesible. Por otro lado, no se tienen reportados los mecanismos de estabilidad de esta proteína, los cuales podrían estar involucrados en la eficiencia con la que la AGT lleva a cabo sus funciones fisiológicas y aplicaciones en las que se ha utilizado, además de abrir la posibilidad de proponer un mecanismo de acarreo de fármacos que sea únicamente por interacciones moleculares sin que tenga que llevarse a cabo una conjugación. Esto revelaría un potencial por parte de la AGT que no ha sido explotado.

## Antecedentes

En el grupo de investigación del Dr. García Hernández se han realizado estudios sobre plegamiento y homodimerización de la AGT [Tellez et al., 2011], así como la caracterización biofísica de dicha proteína y sus cuatro dominios [Leyva et al., 2019]. En este último estudio obtuvieron la constante de unión y la firma termodinámica de la unión entre quitotriosa y los cuatro dominios. Este ensayo demostró que la interacción está guiada principalmente por el componente entálpico. Mediante calorimetría diferencial de barrido se observó que la proteína tiene un desplegamiento complejo de varias etapas, debido a que cada dominio posee una temperatura media de fusión específica ( $T_m$ ), como se muestra en la **Tabla 2**.

**Tabla 2.**  $T_m$  correspondiente a cada dominio.

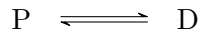
Valores tomados de [Leyva et al., 2019].

Dominios	Temperatura media de fusión (°C)
A	80
B	92
C	86
D	67

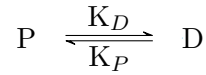
Esta diferencia en termoestabilidad podría estar teniendo un papel importante en las

propiedades biomédicas y biotecnológicas atribuibles a la AGT. Se sabe que la estabilidad en proteínas va asociada con interacciones moleculares como puentes de sal y P-H, así como una tendencia a formar estructura secundaria [Panja et al., 2020]. Sin embargo, el origen molecular de las diferencias de estabilidad entre los dominios de la AGT no ha sido elucidado.

El comportamiento de una proteína en una solución se ve de la siguiente forma:



Donde P indica el estado plegado y D el desnaturalizado. Por lo tanto, existe una constante para cada estado en el que puede estar la proteína, es decir, una constante de desnaturalización y de plegamiento nativo.



Con base en la reacción anterior se define la constante de equilibrio de la siguiente manera:

$$K_{eq} = \frac{K_D}{K_P} = \frac{[\text{Proteína}_{desp}]}{[\text{Proteína}_{pleg}]} \quad (1)$$

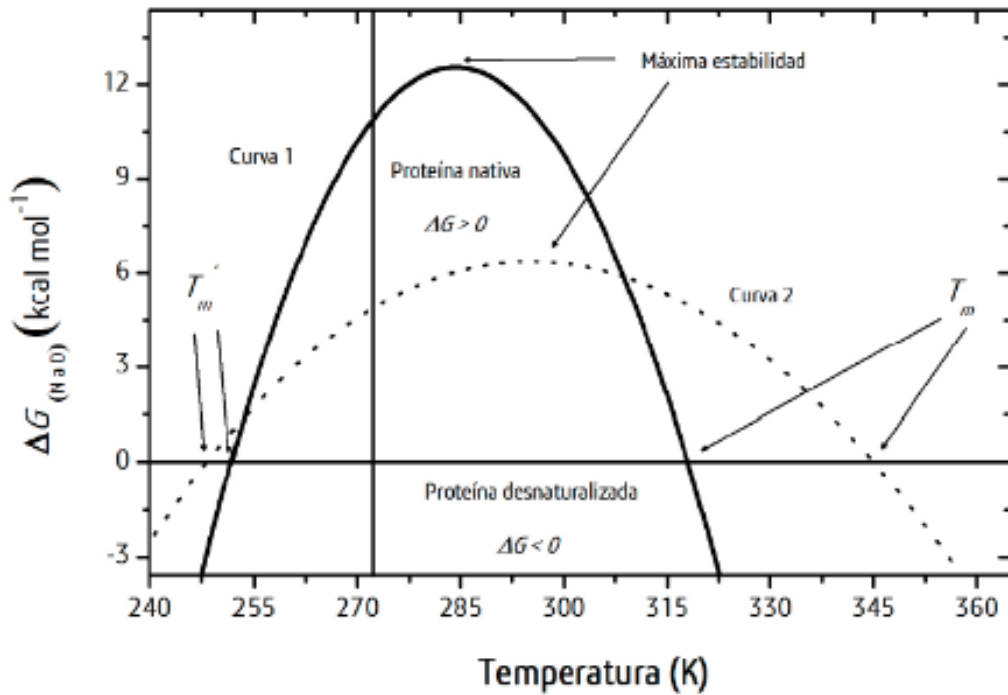
Cuando se habla de estabilidad térmica de proteínas es necesario tomar en cuenta un parámetro que indique la energía del sistema que se esté analizando. Por lo tanto, la estabilidad térmica de una proteína se define cuantitativamente por el cambio de la energía libre de Gibbs del desplegamiento, a través de la ecuación.

$$\Delta G = -RT \ln(K_{eq}) \quad (2)$$

Donde  $R$ , es la constante de Boltzman y  $T$  es la temperatura. Notemos que si la proteína tiene la misma tendencia a desnaturalizarse o mantener su plegamiento, entonces:

$$\begin{aligned} [\text{Proteína}_{despleg}] &= [\text{Proteína}_{pleg}] \\ \implies K_{eq} &= 1 \\ \implies \Delta G &= 0. \end{aligned} \quad (3)$$

La temperatura a la que sucede este fenómeno se le conoce como temperatura media de fusión ( $T_m$ ) y se utiliza como parámetro para describir la termoestabilidad de proteínas [Miotto et al., 2019]. Todas las proteínas presentan un comportamiento similar respecto a



**Figura 2.** Curva de estabilidad de proteínas. Los puntos donde la gráfica toca el eje X indican las dos  $T_m$ , mientras que el punto máximo indica la temperatura de máxima estabilidad.

la temperatura, en el que pueden ser desnaturalizadas con frío o calor, por lo que existen dos valores de  $T_m$  para cada proteína.

Otra forma de describir este parámetro, es a través de la ecuación de la energía libre de Gibbs-Helmholtz dependiente de la temperatura, como se muestra a continuación.

$$\Delta G(T) = \Delta H \left(1 - \frac{T}{T_m}\right) - \Delta C_p \left(T_m - T + T \ln\left(\frac{T}{T_m}\right)\right) \quad (4)$$

Donde  $T$  es la temperatura,  $\Delta H$  es el cambio en la energía desnaturalización y  $\Delta C_p$  es el cambio en la capacidad calorífica. Esta ecuación describe los dos valores de  $T_m$ , además de un tercer parámetro conocido como  $T_{max}$ . Este indica la temperatura a la que la proteína presenta su máxima estabilidad. Curiosamente esta temperatura no corresponde con su máximo rendimiento en función fisiológica. El comportamiento de la proteína se modela a través de la ecuación (4) dado que toma en cuenta las tres temperaturas clave para el comportamiento de una proteína (**Figura 2**).

## Preámbulo de dinámica molecular

Las simulaciones de dinámica molecular son un método computacional que permite describir el movimiento de moléculas, así como sus propiedades energéticas y conformacionales a una escala atómica [Rizzuti, 2022]. Para poder realizar estos cálculos, es necesaria una expresión matemática que describa la dependencia de la energía del sistema ( $U$ ) respecto de la posición de los átomos  $U(r_1, r_2 \dots r_n)$ . Los campos de fuerza son el conjunto de funciones y parámetros utilizados para describir la energía de un sistema en función de sus posiciones [Toofanny and Daggett, 2012]. Entre los componentes de un campo de fuerza típico, se encuentran las interacciones de enlace covalente, las cuales tienen parámetros como: longitud, ángulo y torsión de enlace. Las primeras dos se simulan utilizando la ley de Hooke, ya que el enlace puede comportarse como un resorte debido su capacidad de contraerse o expandirse. La función con la que se simula este comportamiento es:

$$\sum \frac{1}{2}k_b(r - r_0)^2 \quad (5)$$

Donde  $k_b$  es una constante y  $r - r_0$  es la diferencia de longitud del enlace entre el estado de mínima energía ( $r_0$ ) y la distancia actual ( $r$ ).

Para el comportamiento de los ángulos se utiliza la misma ecuación pero tomando en cuenta el cambio de ángulo ( $\theta$ ).

$$\sum \frac{1}{2}k_b(\theta - \theta_0)^2 \quad (6)$$

Donde  $\theta - \theta_0$  es la diferencia entre el ángulo de mínima energía ( $\theta_0$ ) y el ángulo actual ( $\theta$ ). Debido a la geometría de biomoléculas como proteínas o compuestos orgánicos, existen distintos arreglos de átomos unidos por un enlace covalente. Cada arreglo de tres átomos unidos de forma no lineal, forma un plano. El ángulo entre estos planos se define como ángulo diedro. Este parámetro influye en la energía libre de las conformaciones que puede adoptar la molécula, debido al traslape de planos, como se muestra en la **Figura 3**. Para describir este tipo de movimientos se utiliza la función  $\cos()$  de la diferencia de los ángulos de torsión, como se muestra a continuación.

$$\sum k_{\phi,n}[1 + \cos(n\phi - \phi_n)] \quad (7)$$

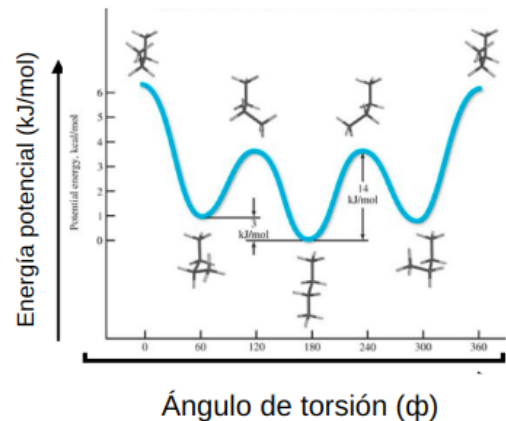
Las interacciones no covalentes que se calculan en el campo de fuerza son interacciones electrostáticas y de van der Waals. Para las interacciones electrostáticas, se toma en cuenta la carga del átomo, la cual depende de la carga elemental de los átomos unidos. Si al interactuar tienen cargas iguales, la energía potencial aumenta conforme se reduce la distancia entre ellos, mientras que si tienen cargas opuestas, la energía potencial disminuye, como se ve en la **figura 4**. Este tipo de interacciones se simulan usando la ley de Coulomb, con la siguiente ecuación.

$$U(r) = \frac{q_i q_j}{r} \quad (8)$$

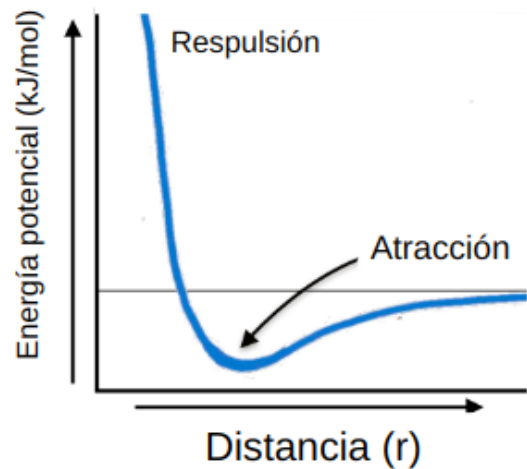
Por último, en las interacciones de van der Waals también consideran la distancia entre dos átomos pero independientemente de las cargas. Cuando dos átomos se encuentran demasiado cerca uno de otro existe una repulsión, mientras que si los átomos se encuentran lejos hay cierta atracción. Esto indica que debe haber una distancia de separación específica con un mínimo de energía, como se ve en la **Figura 4**. Estas interacciones se calculan mediante el potencial Lennard-Jones [Adams, 2001], como se muestra a continuación.

$$U(r) = \epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (9)$$

A pesar de que existen distintos campos de fuerza con variaciones en los parámetros que dependen del sistema a simular, un campo de fuerza general se ve de la siguiente manera:



**Figura 3.** Aumento o disminución de energía respecto a la distancia de los átomos.



**Figura 4.** Energía potencial respecto al ángulo de torsión de enlace.

$$\begin{aligned}
U(r) = & \sum \frac{k_b}{2}(r - r_0)^2 \\
& + \sum \frac{k_a}{2}(\theta - \theta_0)^2 \\
& + \sum k_{\phi,n}[1 + \cos(n\phi - \phi_n)] \\
& + \frac{q_i q_j}{r} \\
& + \epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^6]
\end{aligned} \tag{10}$$

El principio de cada simulación consiste en la resolución de las ecuaciones de movimiento de Newton para un número determinado de átomos [Beck and Daggett, 2004]. Para determinar la evolución de un sistema respecto al tiempo, es necesario saber las posiciones, velocidades y aceleraciones de cada átomo en el sistema. Las dos ecuaciones principales que describen una simulación de dinámica molecular son:

$$\begin{aligned}
F_i &= \frac{\partial U}{\partial r_i} \\
a_i &= \frac{F_i}{m_i}
\end{aligned} \tag{12}$$

Donde  $F_i$  es la fuerza que actúa sobre el átomo  $i$ ,  $U$  es la energía potencial del sistema,  $r_i$ ,  $a_i$  y  $m_i$  son la posición, aceleración y masa del átomo  $i$ , respectivamente.

Dado que estas ecuaciones no pueden ser resueltas de forma analítica para varios átomos, se utiliza el algoritmo de Verlet ([Becker and Karplus, 2006], generando una aproximación numérica para las posiciones  $r$ , velocidades  $v$  y aceleraciones  $a$ , en el tiempo  $t$ . El algoritmo calcula las nuevas posiciones en el tiempo  $(t + \Delta t)$  de la siguiente manera.

$$r(t + \Delta t) = r(t) + v(t)\Delta t + (\frac{1}{2})a(t)(\Delta t)^2 \tag{13}$$

Donde  $a$  está determinada por las posiciones de los átomos obtenidas utilizando la ecuación 2, y a partir de esas nuevas posiciones, se calculan las nuevas aceleraciones  $(t + \Delta t)$ . Una vez conocidas las posiciones, velocidades y aceleraciones, se repite el proceso y se calcula la siguiente configuración.

Una de las aplicaciones de la dinámica molecular es el estudio del desplegamiento térmico de proteínas. Con el fin de comprender los determinantes de estabilidad de la

AGT, en este estudio realizaremos la caracterización del desplegamiento térmico de la proteína completa y de los cuatro dominios aislados, mediante simulaciones de dinámica molecular. Los resultados serán contrastados con el comportamiento observado experimentalmente, con el fin de lograr una descripción unificada del desplegamiento de esta proteína [Toofanny and Daggett, 2012].

## Planteamiento del problema

La proteína AGT tiene complejos mecanismos de termoestabilidad [Hernández et al., 1995], los cuales han sido estudiados mediante calorimetría de titulación isotérmica, barrido diferencial, dicroísmo circular y resonancia magnética nuclear [Leyva et al., 2019], las cuales han aportado con alta precisión valores de entalpía, entropía, energía libre y  $T_m$  de la proteína y de los dominios aislados. El desplegamiento de esta lectina es un proceso complejo de pasos múltiples, en donde cada dominio se despliega a una temperatura diferente [Leyva et al., 2019]. Una forma de comprender las bases moleculares del mecanismo de termoestabilidad de cada dominio aislado y en la proteína completa, requiere ser estudiado mediante simulaciones de dinámica molecular, las cuales permitirán obtener un mejor conocimiento del comportamiento de la proteína bajo condiciones a las que se ha sometido experimentalmente.

## Justificación

Con base en los descubrimientos previos que se han hecho respecto a esta proteína y el rumbo que está tomando la investigación biomédica y biotecnológica, es importante llevar a cabo una caracterización detallada a nivel atómico de los mecanismos de termoestabilidad y desnaturalización de la AGT, con el fin de establecer las bases para ser utilizadas en estudios posteriores como vehículo para la administración de fármacos vía oral.



## Objetivos

### Objetivo general

- Comprender los determinantes moleculares de la estabilidad de la aglutinina de germen de trigo.

### Objetivos particulares

- Estudiar *in silico* a cada dominio aislado y la proteína completa a diferente temperatura y correlacionar con los resultados experimentales del desplegamiento térmico de cada dominio.
- Estudiar el monómero de la proteína para elucidar los efectos de la condición multidominio de los contactos en la estabilidad de cada uno de los dominios y de la proteína completa.

# Metodología

## Preparación de la estructura

La estructura cristalográfica de la AGT fue tomada del *Protein Data Bank* [Berman, 2000], (código de PDB: 2UVO). Los ligandos y otros heteroátomos fueron eliminados y los dominios fueron aislados, utilizando el programa de visualización y análisis molecular UCSF Chimera 1.16 [Pettersen et al., 2004]. Los átomos faltantes en los dominios aislados y en la proteína completa, así como la protonación a pH 4.7 fueron realizados con la aplicación PDBFixer [Eastman et al., 2017].

## Dinámica molecular

Las simulaciones de los dominios aislados y de la proteína completa se hicieron con el software GROMACS 19.6 [Bekker et al., 1993], utilizando el campo de fuerza CHARMM36m [Huang and MacKerell Jr, 2013], y con solvente explícito, utilizando el modelo de agua TIP3P [Price and Brooks III, 2004]. La carga del sistema fue neutralizada agregando contraiones de cloruro y sodio según fuera requerido. Se utilizó una caja dodecaédrica, con una distancia inicial de 1.5 Å entre la proteína y cada lado de la caja, aplicando condiciones periódicas de frontera. Se utilizaron sitios virtuales para disminuir el costo computacional, permitiendo el uso de un tiempo de integración de 4 fs. Los sitios virtuales consisten en restricciones explícitas a los enlaces de los átomos de hidrógeno, lo cual permite obviar el movimiento de dichos átomos y omitir sus cálculos [Larsson et al., 2020]. El sistema se minimizó por medio de una optimización geométrica con un número máximo de 5000 ciclos, realizada con el algoritmo de descenso escarpado (*steepest-descent*), con el fin de ajustar la orientación del solvente y eliminar choques estéricos entre los átomos de la proteína y el solvente. Cada trayectoria comenzó con una equilibración de 1.5 ns para la asigna-

ción de velocidades, utilizando una distribución de Maxwell. Dicha equilibración se llevó a cabo en cinco pasos: cuatro etapas utilizando el ensamble canónico (NVT), en el cual se mantiene constante el número de moléculas (N), volumen (V) y temperatura (T), con calentamiento gradual desde 273 K hasta llegar a la temperatura respectiva de análisis, seguidas por una etapa utilizando el ensamble isobárico-isotérmico (NPT), en donde se mantiene constante la presión (P), al igual que N y T. Para la producción de las simulaciones, se estableció un criterio de corte *cutoff* de 1 nm para interacciones electrostáticas y de van der Waals, mientras que para calcular interacciones de larga distancia se utilizó el método de Ewald mesh (Essman et al., 1995). Para mantener la temperatura y la presión constantes se utilizaron el termostato y barostato, *rscale* y *Parinello-Rahman* [Parinello and Rahman, 1981], respectivamente.

Las producciones de los dominios aislados y de la proteína completa se corrieron en un ensamble NPT a diversas temperaturas. Los dominios se corrieron a su propia  $T_m$  y a cada una de las  $T_m$ s de los demás dominios, además de las temperaturas control, que fueron: 20° y 120° C para observar cómo se comporta la proteína en su estado nativo y desnaturizado. Se corrieron cinco réplicas de 500 ns cada una de las temperaturas por cada dominio aislado y el complejo de la proteína completa a las mismas temperaturas, dando un total de tiempo de simulación de 2.5  $\mu$ s por dominio. Las dinámicas se corrieron en

## Alineamiento

Para alinear la proteína al centro del sistema y eliminar artefactos causados por las condiciones periódicas de frontera, se utilizó el comando:

```
gmx trjconv -s file.tpr -f file.xtc -o file.xtc -pbc mol -center
```

Donde la función *trjconv* con la opción *-pbc mol -center* es para mantener mantener a todas las moléculas alineadas con respecto a la estructura cristalográfica después de centrarlas. Para eliminar los movimientos traslacionales y rotacionales se utilizó nuevamente el siguiente comando, nuevamente con la función *trjconv*.

```
gmx trjconv -s md_0_1.tpr -f md_0_1_noPBC.xtc  
-o md_0_1_ok.xtc -fit rot+trans
```

Por último, se quitaron las moléculas de agua en la trayectoria y en la topología.

```
gmx trjconv -s md_0_1.tpr -f md_0_1_ok.xtc -o md_0_1_dry.xtc
gmx convert-tpr -s md_0_1.tpr -o md_0_1_dry.tpr
```

## RMSD y RMSF

La desviación media cuadrática (RMSD por sus siglas en inglés) fue utilizada para evaluar la similitud estructural entre cada conformero obtenido en las trayectorias con respecto a la estructura cristalográfica. La magnitud del valor de RMSD está mediada por la desviación total que presente la estructura respecto a la de referencia. No obstante, esta desviación puede cambiar en ciertas partes de la estructura, por lo que si se acota a un número determinado de residuos este valor puede disminuir. Debido a esto, no se recomienda como única medida para la inferencia de resultados [Knapp et al., 2011]. La ecuación que utiliza el programa para realizar el cálculo es la siguiente:

$$RMSD(t_1, t_2) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_2) - r_i(t_1)\|^2 \right]^{\frac{1}{2}} \quad (14)$$

Donde  $r_i(t)$  es la posición del átomo  $i$  en el tiempo  $t$  y  $M = \sum_{i=1}^N m_i \|r_i(t)\|$ .

Para determinar el movimiento que presenta el Carbono alfa ( $C_\alpha$ ) de los residuos durante las dinámicas a diferente temperatura, se calculó un índice de fluctuación conocido como fluctuación media cuadrática (RMSF por sus siglas en inglés). La ecuación con la que se calculó dicho índice es la siguiente:

$$RMSF_i = \left[ \frac{1}{T} \sum_{t_j} |r_i(t_j) - r_i^{ref}|^2 \right]^{\frac{1}{2}} \quad (15)$$

El programa calcula este índice a partir de la desviación estándar de las posiciones atómicas a lo largo de la trayectoria, al indicar un marco de referencia de fluctuación, ya sea entre átomos, cadena principal,  $C_\alpha$ , etc. A diferencia de la RMSD, este índice no es en función del tiempo sino de cada residuo de la proteína que se está analizando. Esto da un panorama general de las partes más móviles.

## Cálculo de puentes de hidrógeno

El cálculo de los puentes de hidrógeno (P-H) se hizo con la función *hbonds* de CPP-TRAJ [Roe and Cheatham III, 2013], utilizando el rango de residuos [8-35], el cual no

incluye los amino y carboxilo terminales (AT y CT, respectivamente), como se describe en resultados. El cálculo se hizo sobre todas las réplicas de cada temperatura, por lo que el tiempo total para calcular la fracción de presencia fueron 2.5  $\mu$ s.

## Cálculo de contactos

El cálculo de contactos consistió en cuantificar la fracción de tiempo de cualquier interacción entre residuos, ya fuera de tipo polar o hidrofóbico. Para esto se utilizó la función `nativecontacts` de CPPTRAJ, especificando que no se tomaran en cuenta los sitios virtuales. Al igual que el cálculo de los P-H, esto se hizo sobre el tiempo total de simulación.

## Análisis de componentes principales

El principio de Análisis de componentes principales (PCA), consiste en la descomposición de un conjunto de datos correlacionados en componentes principales (PC) ortogonales que contengan la mayor cantidad de varianza posible. A través de este análisis es posible describir los movimientos esenciales que lleva a cabo la proteína durante su simulación y la obtención de un histograma de probabilidad, a partir del cual se puede hacer el cálculo del paisaje de energía libre (PEL). Con base en esto, se realizó un PCA de ángulos diedros (dPCA), con el fin de eliminar los movimientos generales de la proteína y tomar en cuenta únicamente los movimientos internos [Altis et al., 2007]. El cálculo de los ángulos diedros se hizo con la función *multihedral* de CPPTRAJ en el mismo rango de residuo especificado anteriormente. Los ángulos se convirtieron a coordenadas métricas para después calcular la matriz de covarianza  $2N$  y calcular los vectores y valores propios. Una vez calculados los componentes principales (PC), se calcularon los PEL's al proyectar los dos primeros PC, utilizando PyEMMA. Este cálculo se hace utilizando la ecuación:

$$F(x) = -k_B T \ln[P(x)] \quad (16)$$

Donde  $F(x)$  es la energía libre correspondiente a la coordenada  $x$ ,  $k_B$  es la constante de Boltzman,  $T$  es la temperatura y  $P(x)$ , es la distribución de probabilidad de la coordenada  $x$  obtenida a partir del histograma que se obtiene al proyectar  $PC_1$  vs  $PC_2$ .

## Modelos de Estados de Markov

Este método sirve para modelar las dinámicas de un sistema como una serie de eventos probabilísticos independientes. El modelo consiste en agrupar los estados conformacionales y construir una matriz de probabilidades de transición entre cada estado. Para formar los modelos de Markov se tomó en cuenta la energía cinética de cada conformero y con el algoritmo K-means se agruparon los conformeros que tuvieran valores de energía similar. Con base en la energía cinética se formó la matriz de transición entre cada estado.

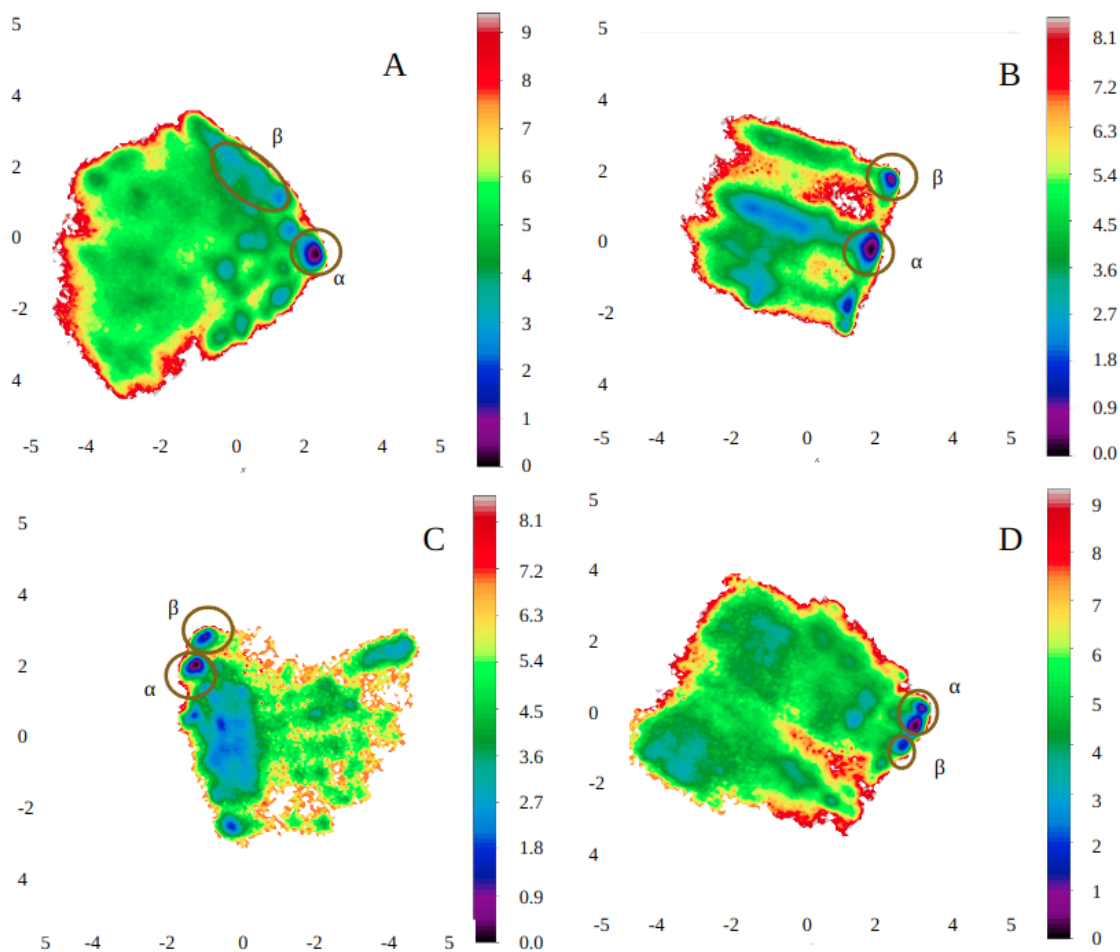
# Resultados

## Definición de estados plegado y desplegado

Se realizó un análisis de componentes principales de ángulos diedros (dPCA), con el fin de reducir el número de dimensiones en las simulaciones, conteniendo la varianza de los datos en PC's y capturando la dinámica esencial de la proteína y sus dominios aislados. Utilizando los conjuntos de datos de los PC<sub>1</sub> y PC<sub>2</sub> se construyeron modelos de estados de Markov (MEM), con el fin de agrupar los conformeros de cada trayectoria en macroestados, a partir de los cuales se puede observar cuales eran los estados más poblados. Se utilizaron estos modelos dado que brindan una descripción cuantitativa de la termodinámica y cinética de las proteínas al definir sus estados estructurales y la tasa de transición entre ellos [Zimmerman et al., 2017]. Esto permitió tener una primera aproximación de los estado plegado y desplegado de los dominios y la proteína completa para poder comparar con resultados posteriores. Los cálculos se realizaron concatenando todas las trayectorias de 20°, T<sub>m</sub> respectiva y 120° en el caso de los dominios aislados, ya que así se aseguraba encontrar los estados con plegamiento nativo y desnaturalizados, además de ver la aportación de la T<sub>m</sub> al paisaje conformacional. Para la proteína completa se concatenaron todas las trayectorias de las T<sub>m</sub>'s de los dominios, más las temperaturas control. En la **Figura 5** se observa el PEL de los cuatro dominios aislados, donde se indican los mínimos de energía que corresponderían a los estado plegado ( $\alpha$ ) y desplegado ( $\beta$ ). El hecho de que en cada PEL haya sólo una cuenca de atracción con una energía de 0.0 kT es congruente con el hecho de el estado plegado es el de menor energía. Los dominios A y D son los que más exploración conformacional presentan debido a la amplitud de la gráfica. En el caso de B y C, los PEL son menos amplios y ambos presentan zonas diferentes a las cuencas de atracción con energías entre 1.8 y 2.7 kT, lo que indica que lograban adoptar

conformaciones de menor energía que los dominios A y D. Al analizar las conformaciones más habitadas de cada cuenca y comparar con la estructura cristalográfica (**Figura 6**), se observó que el conformero  $\alpha$  es prácticamente idéntico a la estructura cristalográfica. Las estructuras correspondientes a las zonas  $\beta$  presentan cambios conformacionales en el CT de los dominios A, C y D, mientras que en el dominio B, el AT es el que cambia más su plegamiento. En todos los dominios, el cambio conformacional respecto a la estructura de referencia provoca la pérdida de la hélice alfa del respectivo terminal. Los dominios en la proteína completa tienen un comportamiento similar, ya que en el estado plegado el rango de residuos correspondientes a cada dominio conservaron un plegamiento similar a la estructura cristalográfica. No obstante, la similitud en plegamiento de la proteína completa respecto al cristal sí se observaron cambios en las azas interdominio, las cuales se compactan, haciendo que la proteína adquiriera una forma de U. En el caso del estado desplegado de la proteína, el plegamiento de los dominios se ve más afectado, especialmente en el carboxilo terminal de cada uno. Se tomaron en cuenta dos métricas sobre el grado de desnaturalización. La primera consistió en contabilizar la cantidad total de P-H que tenía el conformero perteneciente a cada estado y determinar si existe una diferencia en cuanto a dichas interacciones (**Tabla 3 y 4**). Por otro lado, se calculó la (RMSD) entre los dos conformeros correspondientes a  $20^\circ$  y  $120^\circ$  (datos nos mostrados). La diferencia en la cantidad de P-H de ambos estados en todos los dominios aislados es de  $5.5 \pm 1.3$ . El dominio que pierde más interacciones es el dominio D, donde el mayor rompimiento de interacciones ocurre en los segmentos terminales de éste. Esto podría deberse a que es el dominio que más se afecta su plegamiento al desnaturalizarse, lo cual se ve reflejado en un RMSD de  $6.3 \text{ \AA}$ . Un patrón que tienen en común los dominios A, C y D al desnaturalizarse es que CT es el que presenta el mayor cambio conformacional, mientras el núcleo y el AT se mantienen prácticamente con el mismo plegamiento (**Figura 6**). Por otro lado, en el dominio B los primeros 12 residuos del AT cambian de plegamiento, mientras que el CT se mantiene plegado. Esto permite explicar por qué los P-H de dichas zonas perduren por más tiempo.





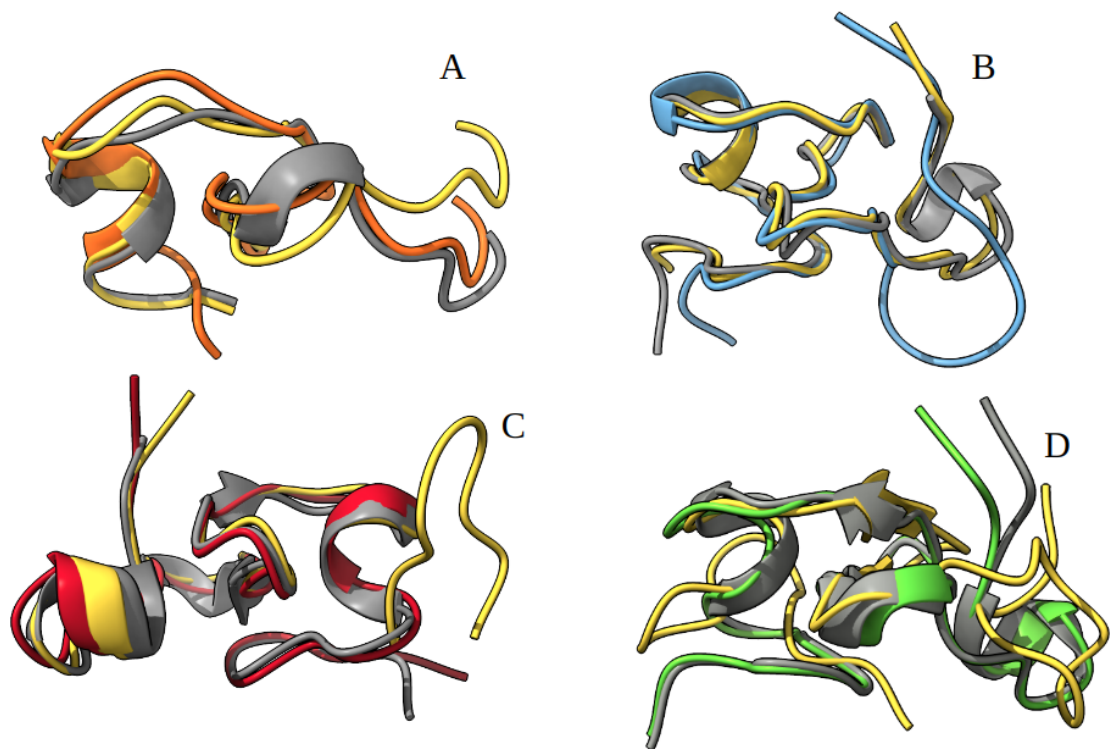
**Figura 5.** PEL de los dominios aislado de las trayectorias concatenadas de  $20^\circ$ , Tm y  $120^\circ$ . Los círculos denotan los pozos de atracción donde las conformaciones transitan más. El pozo  $\alpha$  corresponde al más poblado y el  $\beta$  el segundo más poblado.

entre el terminal que más se despliega se vuelve menos notoria respecto a cuando están aislados, ya que ambos se desnaturalizan en la misma proporción. Los valores de RMSD son de  $6.5 \text{ \AA}$ ,  $4.7 \text{ \AA}$ ,  $6.5 \text{ \AA}$  de los dominios A, B, C y D, respectivamente.

El número de P-H que presentan los dominios cuando están en la proteína es menor que cuando están aislados, aunque la pérdida de interacciones entre el estado plegado y desplegado es menor como se muestra en la tabla 2, ya que estos pierden en promedio  $4.3 \pm 1.7$  interacciones. En este caso, el dominio B mantiene el mismo número de interacciones en ambos estados, a pesar de que la RMSD es de  $4.0$ , entre los dos estados. En los demás dominios en la proteína la diferencia

**Tabla 3.** Cantidad de P-H de los dos estados de los dominios de la AGT.

Dominio	Plegado	Desplegado
A	18	14
B	15	9
C	16	11
D	17	11



**Figura 6.** Comparación de los estados plegado y desplegado (amarilla) vs la estructura cristalográfica (gris) de los dominios aislados. El conformero plegado se encuentra en color diferente para cada dominio con el fin de resaltar la estructura.

**Tabla 4.** Cantidad de P-H en los dominios en la proteína.

Dominio	Plegado	Desplegado
A	10	7
B	13	13
C	11	9
D	13	8

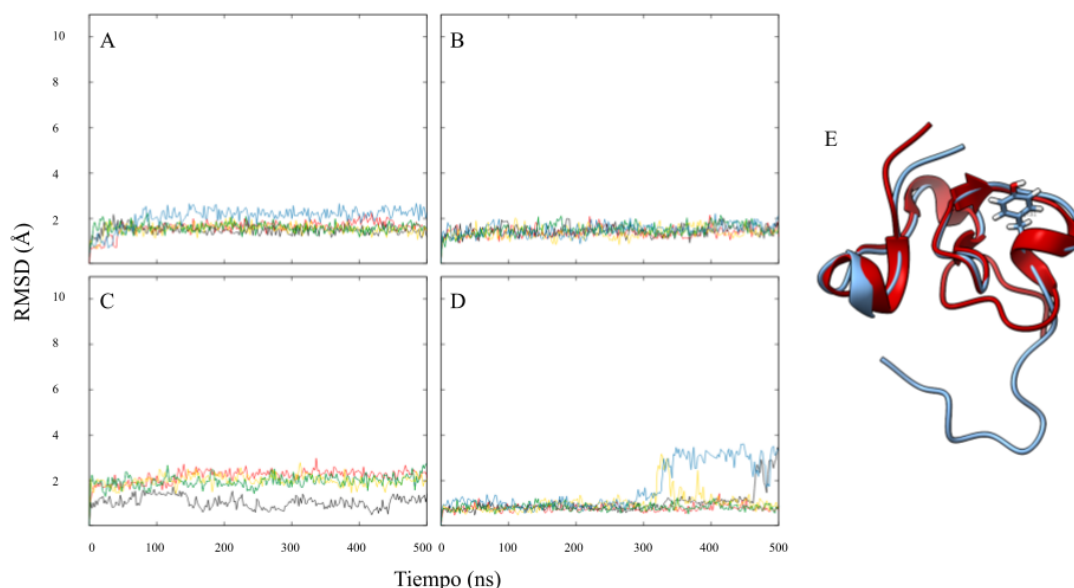
## Análisis de RMSD

La comparación de estructuras y RMSD, reveló que existe poca diferencia en los cambios conformacionales de cada dominios una vez que se simula su comportamiento a temperaturas superiores de su  $T_m$ . Por lo tanto, se optó por analizar los dominios aislados y la proteína a 20°, 67°, 80° ( $T_m$  de los dominios D y A, respectivamente) como temperaturas intermedias y 120° C.

Las simulaciones a 20° se utilizaron como referencia del comportamiento conformacional de los dominios. Como un primer paso para conocer dicho comportamiento, se calculó la RMSD. Cuando se calculó el RMSD de toda la trayectoria sobre todos los átomos pesados de los dominios, se obtuvieron valores de 10 Å, los cuales al inspeccionar los confórmeros de las trayectorias, no correspondían con los cambios conformacionales observados. Estos valores se debieron a la flexibilidad conformacional de los extremos. Estas regiones son generalmente las más móviles en proteínas [Ruvinsky et al., 2012]. Posteriormente, se hizo el cálculo de la RMSD respecto de los átomos pesados de los residuos [8-35], con la finalidad de observar el comportamiento de la proteína al remover los extremos terminales (**Figura 7**). A lo largo de todas las réplicas de los dominios A, B y C, se obtuvieron valores de RMSD  $< 2$  Å, lo que indica que las conformaciones muestreadas en las simulaciones son similares a la estructura cristalográfica. Únicamente el dominio D presenta algunas réplicas con valores de RMSD  $> 3$  Å. Al observar los confórmeros que presentan estos valores, observamos que el péptido presenta un cambio de conformación a partir de la Tyr<sup>31</sup> hasta el CT (**Figura 7.E**) En esta región se encuentra una hélice alfa, la cual se pierde debido al cambio conformacional.

## Dominios a su $T_m$ correspondiente

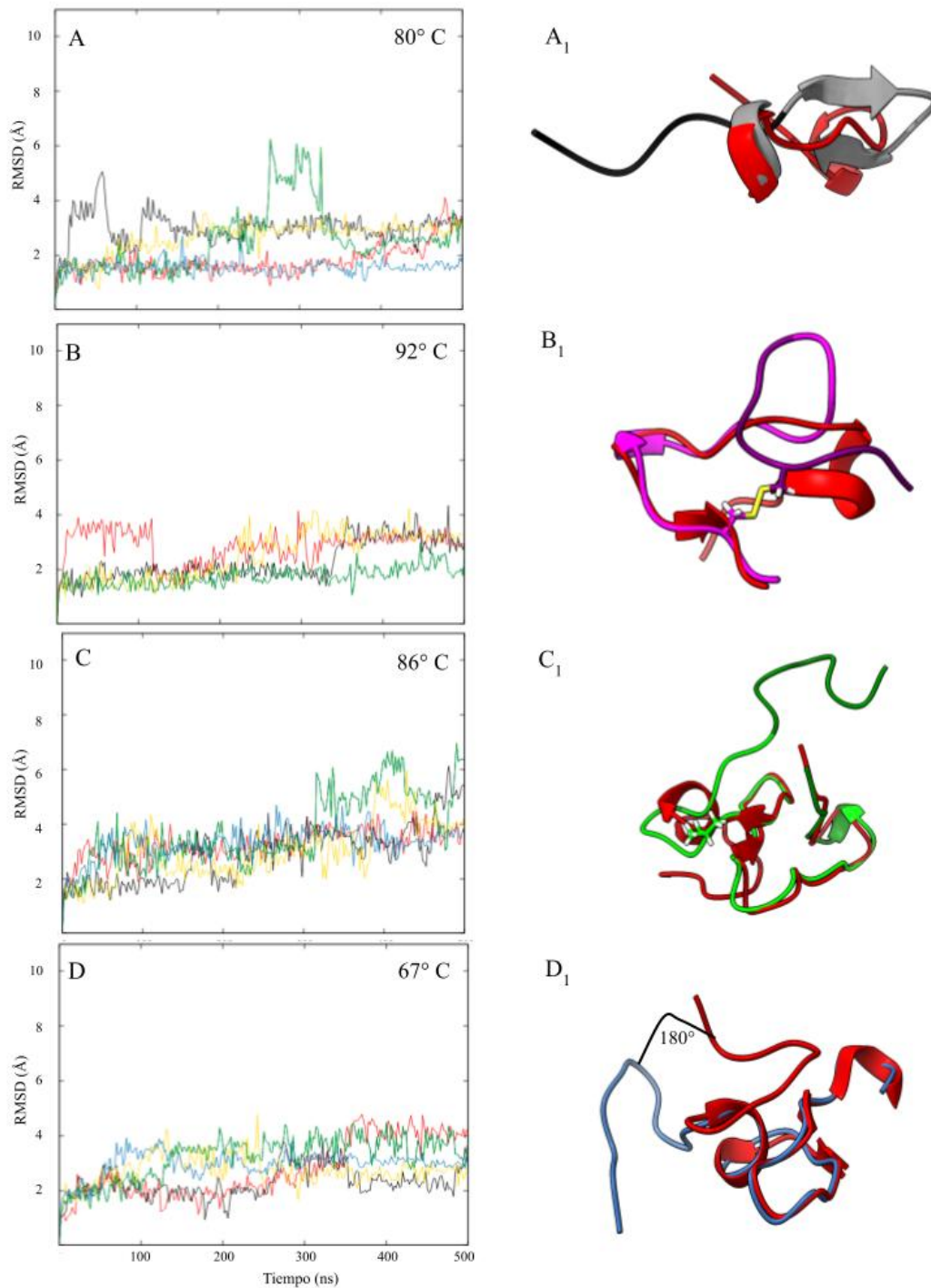
Para el dominio A se observaron dos réplicas que mantuvieron valores de RMSD  $< 2$  Å alrededor del 80 % de las trayectorias, mientras que en las otras tres réplicas se obtuvieron valores entre 3 y 6 Å (**Figura 8.A**). La inspección de estas estructuras con mayor RMSD reveló que el extremo del AT del núcleo formó dos láminas  $\beta$  nuevas en el rango de residuos 8-15, las cuales no se encuentran en la estructura de referencia. Esto explica el aumento de RMDS en el caso del dominio A (**Figura 8.A1**). Los dominios C y D tuvieron un comportamiento más homogéneo, ya que la mayoría de sus réplicas oscilaron entre los 2 y



**Figura 7. RMSD de los cuatro dominios aislados a 20° C.** La letra en la esquina izquierda superior de cada gráfica corresponde a las dinámicas de cada dominio. Se corrieron cinco réplicas de cada dominio, las cuales se representan con los distintos colores

4 Å. Las dos réplicas del dominio C que alcanzaron valores de  $\text{RMSD} > 6$  Å presentaron un desplegamiento a partir de la Leu<sup>16</sup>, el cual afectó la estructura secundaria de la región Glu<sup>1</sup>-Leu<sup>16</sup> ya que en la estructura cristalográfica se encuentra una hélice alfa que se pierde en las trayectorias (**Figura 8.C1**). En el dominio D, los residuos Glu<sup>0</sup>-Tyr<sup>16</sup> formaron un asa que gira 180° respecto a la estructura de referencia (estructura roja), manteniendo el plegamiento del asa y la hélice alfa. A pesar de los cambios conformacionales descritos, en la mayoría de las réplicas los dominios habitaron conformaciones cercanas a la estructura cristalográfica. Por otro lado, cuatro de las cinco réplicas del dominio B presentaron un RMSD promedio de 1.5 Å durante los primeros 250 ns, mientras que otra réplica presentó un promedio de 4 Å durante los primeros 100 ns. El resto de la trayectoria mantuvo un promedio de 3.5 Å. Los cambios conformacionales que provocan este aumento de RMSD, corresponden al asa flanqueada por el puente disulfuro Cys<sup>3</sup>-Cys<sup>12</sup> (**Figura 8.B1**), el cual se compacta y hace que se pierda el plegamiento nativo. Este cambio conformacional fue recurrente en todas las réplicas, sin embargo, se observó que en la mayoría de las réplicas presentaron una tendencia de adaptar una conformación de plegamiento similar al plegamiento nativo.

El análisis de las trayectorias a 120°C para los dominios A, C y D indicaron que el rango de residuos [20-35] es la única parte del dominio que presenta el plegamiento nativo,



**Figura 8.** Gráficas de RMSD a la  $T_m$  respectiva de cada dominio. En la esquina superior izquierda se indica al dominio que corresponde a la gráfica y en la esquina superior derecha se encuentra la temperatura correspondiente a cada gráfica.

lo que corresponde al 70 % del núcleo. La estructura secundaria de ambos extremos se pierde a esta temperatura. Por otro lado, el dominio B mantiene 78 % del plegamiento del núcleo, únicamente los nueve primeros residuos del núcleo pierden el plegamiento respecto a la estructura de referencia. Al comparar los valores de RMSD de las estructuras a 120° y a la  $T_m$ , se observó que los dominios pueden presentar conformaciones desplegadas a la  $T_m$ , lo cual es congruente con la definición de ese valor de temperatura. No obstante, el hecho de que a pesar de la alta temperatura y los saltos abruptos de RMSD, los dominios puedan regresar a un plegamiento cercano al nativo, indica que los cuatro dominios no se desnaturalizan del todo. Esto puede atribuirse en parte a los PdS bien caracterizados que tienen los cuatro dominios [Balčiūnaitė M and Dzikaras, 2021], [Hernández et al., 1995], [Tellez et al., 2011], los cuales representan un componente importante en la estabilidad de la proteína en pH ácido y altas temperaturas.

## Análisis de RMSF

Para identificar las zonas con mayor fluctuación dentro de los dominios, se calculó la RMSF, la cual indica el desplazamiento promedio de cada residuo a lo largo de la trayectoria. Mediante el RMSF es posible observar cuánto aporta cada residuo a los cambios conformacionales de los dominios y cuantificar la restricción de fluctuación que imponen los PdS al péptido [Fuglebakk et al., 2012]. Dado que todos los dominios presentan un plegamiento idéntico, debido a la similitud en la secuencia y misma posición de los PdS como se muestra en la **Figura 9**, se puede inferir conclusiones de los cuatro dominios a partir de la estructura general.

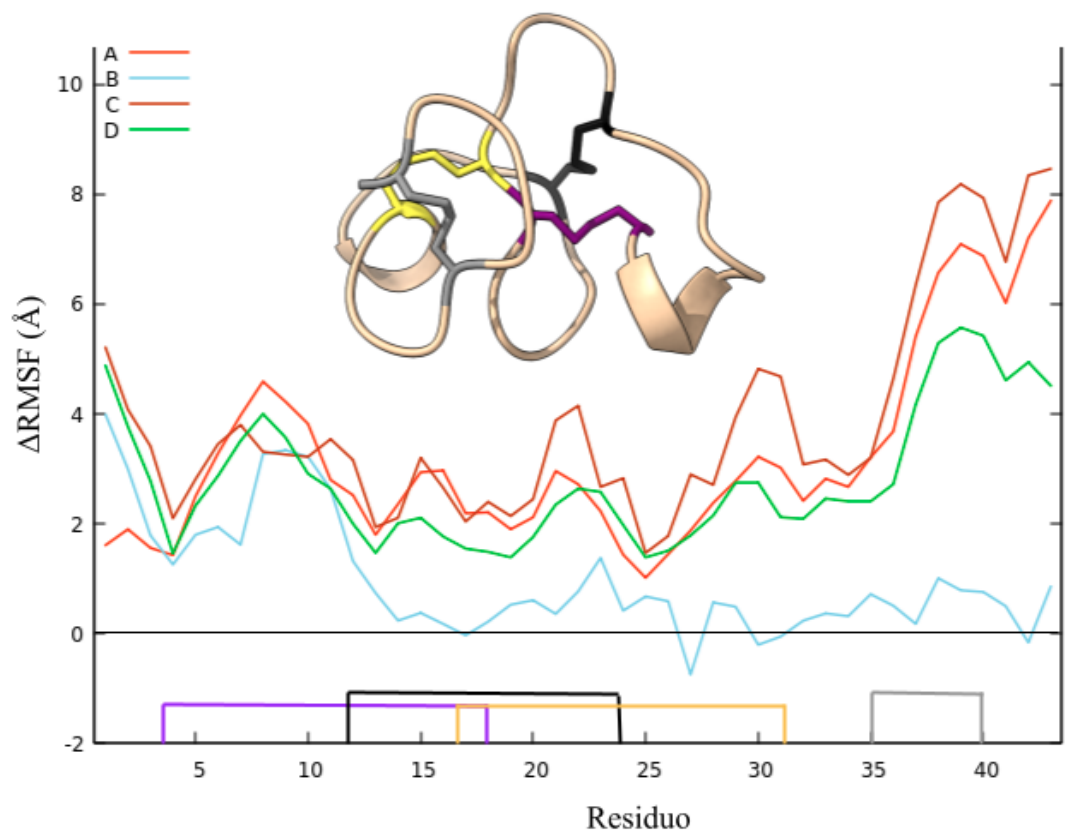
Para observar la influencia que presentan los PdS sobre la fluctuación de los residuos circundantes a estos, se calculó el RMSF promedio de las cinco réplicas, tanto a 20° como a 120° C y se calculó la diferencia de valores de RMSF ( $\Delta$ RMSF) a partir del valor promedio de cada residuo restando los valores pertenecientes a 120° ( $RMSF_{120^\circ}$ ) menos los de 20° ( $RMSF_{20^\circ}$ ), como se muestra en la siguiente ecuación. Por medio de este cálculo es posible ver el cambio en la fluctuación entre las temperaturas.

$$\Delta RMSF = RMSF_{120^\circ} - RMSF_{20^\circ} \quad (17)$$

Por medio de este cálculo fue posible observar de forma más clara cuál es el cambio en la

fluctuación entre  $C_\alpha$  a distintas temperaturas. En la **figura 10** se observa el  $\Delta$ RMSF de los cuatro dominios junto con los PdS marcados tanto en el eje  $X$  de la gráfica como en la estructura cristalográfica. Los dominios A, C y D muestran un comportamiento similar entre sí, ya que el AT es el que más fluctúa en los tres casos, teniendo un valor promedio de cada dominio para los últimos cinco residuos de  $7.01 \pm 0.67$ ,  $7.94 \pm 0.68$ ,  $5.01 \pm 0.47$ , respectivamente. El dominio D presenta valores más bajos de  $\Delta$ RMSF, dado que en sus trayectorias de  $20^\circ$  hubo una mayor fluctuación respecto a los demás dominios por lo que la diferencia con la fluctuación de  $120^\circ$  es menor, lo cual implica que su fluctuación a  $20^\circ$  es menor respecto a los demás dominios. De este modo el dominio C es el que mayor diferencia presenta entre las dos temperaturas, ya que a  $20^\circ$  C su fluctuación es baja mientras que a  $120^\circ$  C esta aumenta considerablemente, especialmente en el AT. Otra característica que tienen en común estos tres dominios es que a partir del residuo 36 todos fluctúan de la misma manera, cambiando únicamente la magnitud de la fluctuación. Los residuos que presentaron mayor fluctuación en los dominios A, C y D fueron: Ala<sup>39</sup>, Ala<sup>125</sup>, y Gly<sup>167</sup>, respectivamente. Por otro lado, el dominio B también muestra una diferencia pequeña entre los valores de RMSF de las dos temperaturas. Lo que ocurre en este caso, es que no hubo un aumento en la fluctuación de los residuos al estar a  $120^\circ$ , respecto a la fluctuación a  $20^\circ$ . Incluso llegó a ser menor en unos residuos como Cys<sup>24</sup>, Gly<sup>23</sup>, Tyr<sup>23</sup>, Cys<sup>74</sup>, Ala<sup>85</sup>. Otro comportamiento único del dominio B es que el terminal que más fluctúa es el CT, mientras que en los demás dominios resulta ser el AT.

Los PdS parecen encontrarse en una posición estratégica, como se muestra en la estructura cristalográfica de la **Figura 9**. Esta ubicación se concentra principalmente en medio de la secuencia de los dominios, lo cual hace que el péptido se pliegue de forma equivalente a un empaquetamiento y mantenga su plegamiento [Maag et al., 2021]. La zona donde se concentran los PdS no sólo es la parte más estable de los dominios, sino que también ahí se encuentra el sitio de unión al ligando. Con el análisis del RMSF es posible ver la importancia de los PdS en la estabilidad de los dominios y permite ver la contribución de cada residuo al movimiento de los dominios cuando se encuentra a un pH ácido. No obstante, no brinda una respuesta a la variación en los valores de la  $T_m$ , ya que dichos puentes se encuentran en la misma posición de cada dominio, por lo que debe de haber otro componente que aporte a las diferencias de las termoestabilidades.



**Figura 9.**  $\Delta\text{RMSF}$  entre  $20^\circ$  y  $120^\circ$  de los cuatro dominios. Diferencia del promedio de RMSF entre  $20^\circ$  y  $120^\circ$  de los cuatro dominios. En el eje  $x$  de la gráfica los PdS están marcados con un color que corresponde al puente marcado en la estructura.



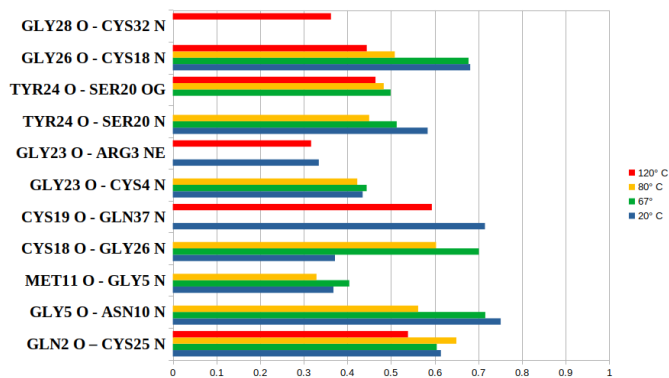
## Cálculo de P-H

De acuerdo con [Bowie, 2011], los P-H son un factor importante en la estabilización de proteínas, lo cual se corroboró por medio de un estudio en el que se mutaron residuos polares que formaban P-H por alanina o valina. Se encontró que dichos puentes contribuían favorablemente a la estabilidad de los péptidos dado que se observaba un decremento en la estabilidad cuando los residuos mutados no eran capaces de formar dichas interacciones [Pace et al., 2014]. Con base en esto, se identificaron los residuos que formaban P-H a 20° C y prevalecían a altas temperaturas. Este tipo de P-H se le llama puentes nativos. Posteriormente se cuantificó la fracción de tiempo que estaba presente dicha interacción a lo largo de las trayectorias. Estos cálculos se realizaron a 20°, 67°, 80° y 120° C con el fin de cuantificar la prevalencia de P-H nativos, los cuales podrían ayudar a explicar la estabilidad individual de cada dominio. Tanto en el dominio A aislado como en la proteína completa, los P-H que más perduran son aquellos que ocurren entre Gly<sup>25</sup> - Cys<sup>17</sup>, Cys<sup>18</sup> - Gln<sup>36</sup> y Gln<sup>1</sup> - Cys<sup>24</sup>, tanto a 20° como a 120° C (**Figura 10.a**). El hecho de que los P-H formados por residuos de cisteína sean los que más perduren, muy probablemente se debe a que los PdS obligan a residuos vecinos a mantener interacciones. Estos tienen un porcentaje de vida por encima del 60 % a 20° C, mientras que a 120° dichas interacciones disminuyen un 15 % respecto a la temperatura anterior. Esto al parecer se debe a las conformaciones expandidas que adopta el dominio en las que a partir de la Cys<sup>17</sup> hasta el CT se pierde el plegamiento y causa la pérdida de las interacciones entre Gly<sup>25</sup> - Cys<sup>1</sup>, Tyr<sup>24</sup> - Ser<sup>19</sup>, Gly<sup>4</sup> - Asn<sup>9</sup>, Gln<sup>1</sup> - Cys<sup>24</sup>. A 67° y Tm los P-H del dominio A tienen una fracción similar a la observada a 20°, ya que interacciones como Tyr<sup>23</sup> - Ser<sup>19</sup>, Gly<sup>22</sup> - Cys<sup>4</sup>, Ser<sup>19</sup> - Gly<sup>22</sup>, Cys<sup>17</sup> - Gly<sup>25</sup>, Met<sup>10</sup> - Gly<sup>4</sup> tienen fracciones de presencia que varían 0.46±0.11 en las tres temperaturas. A 120° se observa un cambio repentino, ya que se pierde la mayoría de las interacciones y la fracción de tiempo de presencia disminuye considerablemente. No obstante, prevalecen algunos P-H que también se encuentran formados a 20°. Cabe recalcar que a 120°, la interacción entre Cys<sup>17</sup> - Gln<sup>36</sup> y Gln<sup>10</sup> - Cys<sup>24</sup>, se encuentran en el dominio aislado más del 70 % de las trayectorias, mientras que en la proteína completa todos los P-H están por debajo del 40 %. A pesar de esto, es posible notar un patrón claro en los dos estados del dominio. Cuando se encuentra en la proteína, los P-H que se encontraban en los extremos a temperaturas inferiores se pierden y prevalecen principalmente los del núcleo, específicamente entre la Cys<sup>17</sup> - Gln<sup>36</sup> y Gln<sup>1</sup> - Cys<sup>24</sup>, Cys<sup>17</sup> y la Gly<sup>25</sup>, lo cual

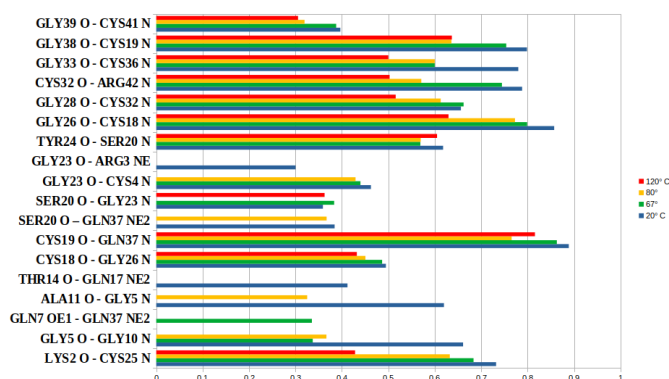
corresponde al sitio de unión del dominio. En el caso del dominio aislado, precisamente en dicha región es donde se pierden las interacciones a 120° y se mantienen en los extremos terminales.

Las interacciones entre Gly<sup>25</sup> – Cys<sup>17</sup>, Ser<sup>19</sup>-Tyr<sup>23</sup>, Gln<sup>1</sup> – Cys<sup>24</sup>, no son únicas del dominio A, ya que se encuentran también en B, C y D, siendo las interacciones que más perduran en todos los dominios. En el dominio B, están presentes más del 80 %, tanto aislado como en la proteína completa. Esta fracción es la más alta a lo largo de los cuatro dominios y en el caso del P-H entre Cys<sup>17</sup> – Gln<sup>36</sup> se mantiene ese tiempo en las cuatro temperaturas (**Figura 10.b**). A diferencia del dominio A, el B llega a presentar P-H únicos de cada temperatura, los cuales no se encuentran en temperaturas subsecuentes como es el caso de Gln<sup>6</sup> – Gln<sup>6</sup>, la cual vale la pena mencionar que a pesar de que estos dos residuos se encuentran cerca de los terminales puede formarse ese P-H, debido al empaquetamiento de este dominio. Otro ejemplo de estas interacciones es: Gly<sup>4</sup> – Ala<sup>7</sup> que se ubican en el AT formando una hélice alfa y por último, Ser<sup>19</sup> – Tyr<sup>23</sup> presente en el sitio de unión. Estas interacciones se presentan únicamente a 20°, 67° y 80°, respectivamente. Estas interacciones propias de cada temperatura sugieren contribuir favorablemente a la estabilidad del dominio en cada temperatura. A pesar de tener P-H espontáneos en las distintas temperaturas, el B aislado es el dominio con más P-H nativos que prevalecen a 120° con fracción de presencia > 20 %, ya que la mitad de los P-H que presenta a 120° tienen fracciones de duración  $\geq$  50 %. En el caso del dominio en la proteína no preserva los P-H de la misma manera, ya que la mayoría de sus interacciones están por debajo del 40 %, únicamente Tyr<sup>67</sup> – Ser<sup>63</sup> mantiene una fracción > 60 % y presente en las cuatro temperaturas. Esto se debe a que el dominio en este estado presenta movimientos más amplios en los extremos, provocados por los dominios A y C. Este movimiento es aún más notorio a alta temperatura. Uno de los P-H que se ve principalmente afectado por el cambio de la temperatura en la proteína es Cys<sup>61</sup> – Gln<sup>80</sup>, ya que a las temperaturas 20°, 67° y 80° está formado un 84±1 % y a 120° perdura 38 %.

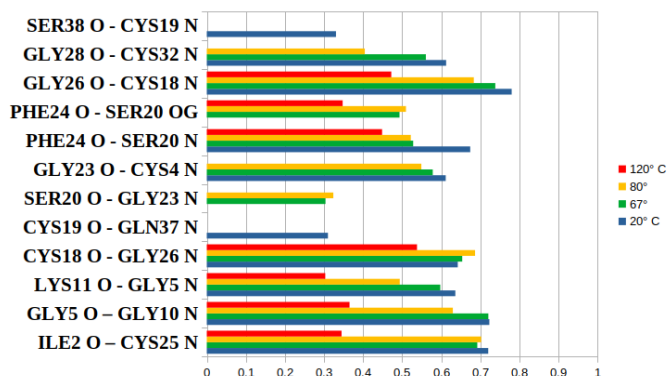
En el dominio C, la distribución de los P-H con mayor presencia es más homogénea a lo largo de la secuencia, por lo que no es tan definida la zona del sitio de unión como en el caso del dominio A. Sin embargo, al comparar los P-H de este dominio en la proteína, se observa una zona distintiva que comprende desde la Leu<sup>103</sup>-Cys<sup>118</sup> donde los P-H de las cuatro temperaturas están presentes entre el 40 y 60 %. En el AT también llega a presentar



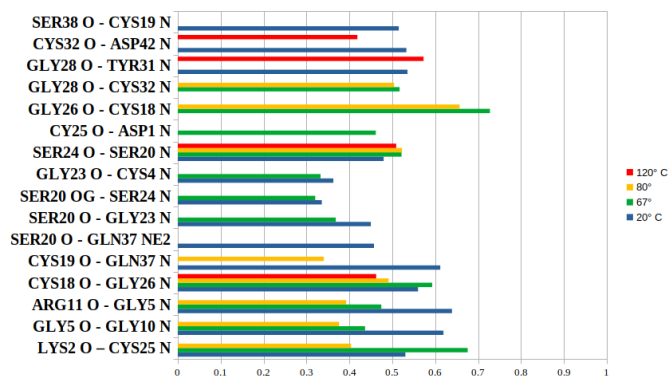
(a)



(b)



(c)



(d)

**Figura 10.** Figura 10. Fracciones de P-H a diferentes temperaturas de los dominios. El código de colores indica la temperatura a la que se presenta dicho P-H. En la izquierda de cada gráfica se muestran el número de residuos que forman la interacción.

interacciones con una alta presencia, mientras que en el CT únicamente las interacciones entre Ile<sup>87</sup>-Cys<sup>110</sup> y Gly<sup>90</sup>- Gly<sup>95</sup> es la que se mantiene a 120°, 23 % (**Figura 10.c**). Por último, el dominio D aislado es el dominio con menor cantidad de P-H tanto en cantidad como en porcentaje de presencia. A 67° y 80°, el sitio de unión mantuvo interacciones arriba del 20%. Vale la pena mencionar que cuando el dominio está a 80° es capaz de mantener interacciones con mayor presencia que cuando está a 20°, como es el caso de Gly<sup>25</sup> - Cys<sup>17</sup> y Arg<sup>1</sup> - Cys<sup>24</sup> entre otras, las cuales estuvieron presentes 70 y 67%, respectivamente. A 120°, sólo se observaron seis P-H, los cuales son Tyr<sup>16</sup> - Cys<sup>12</sup> y Arg<sup>10</sup> - Cys<sup>18</sup>, presentes 25 y 28% de las trayectorias, únicos de esta temperatura. Cuando el dominio D se encuentra en la proteína, los P-H tienen un comportamiento distinto, ya que presentan siete interacciones nativas a lo largo de las cuatro temperaturas, con una presencia del 40%. Estas son: Cys<sup>147</sup> - Gly<sup>155</sup>, Cys<sup>147</sup> - Gln<sup>166</sup>, Ser<sup>153</sup> - Ser<sup>149</sup>, Gly<sup>55</sup> - Cys<sup>147</sup>, Gly<sup>155</sup> - Cys<sup>161</sup> y Ser<sup>165</sup> - Cys<sup>147</sup> (**Figura 10.d**). Varias otras interacciones propias de cada temperatura perduran entre un 30 y 40

Con el fin de poder establecer una correlación entre la estabilidad propia de cada dominio aislado y los P-H que es capaz de formar y mantener, se tomaron en cuenta únicamente los P-H nativos que perduran a 120°, con un porcentaje de presencia en las trayectorias  $\geq 20\%$  y se hizo un análisis de correlación (Tabla 3) [Ragone, 2001]. Este análisis arrojó un coeficiente de determinación ( $R^2$ ) de 0.91, lo cual sugiere una alta correlación entre la presencia de los P-H y la estabilidad propia de cada dominio. Por lo tanto, este análisis sugiere que la estabilidad térmica está mediada hasta cierto punto por los P-H ( ).

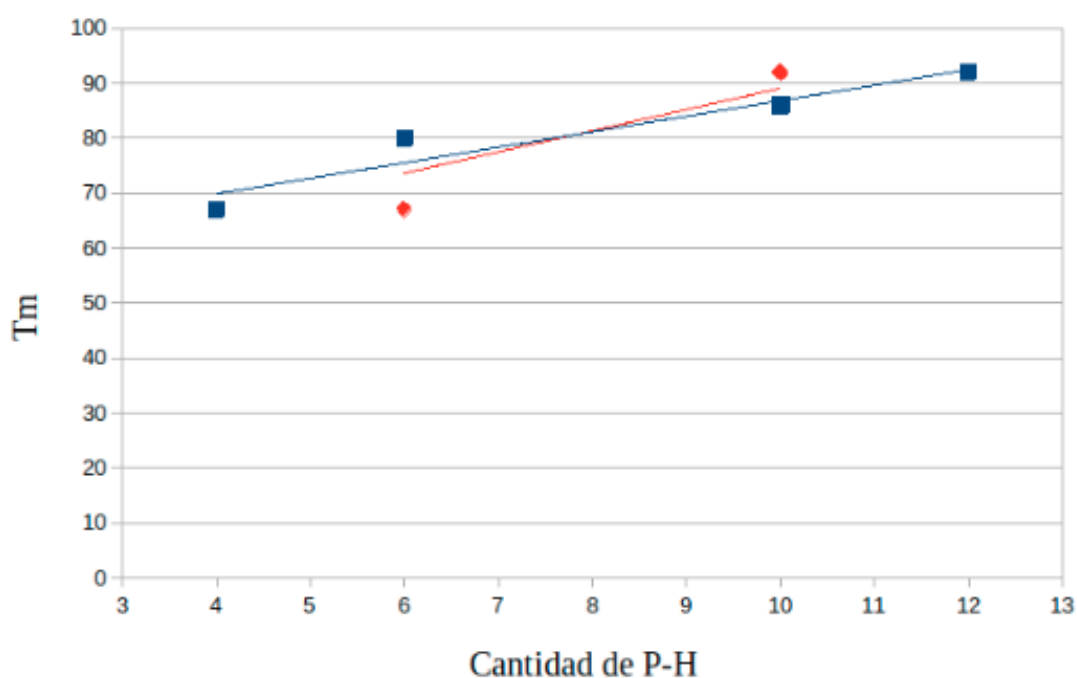
**Tabla 5.** Cantidad de P-H por dominio aislado y su  $T_m$  correspondiente

Dominio	P-H	$T_m$ (° C)
A	6	80
B	12	92
C	10	86
D	4	67

**Tabla 6.** Cantidad de P-H por dominio en la proteína y su  $T_m$  correspondiente

Dominio	P-H	$T_m$ (° C)
A	6	80
B	10	92
C	10	86
D	6	67

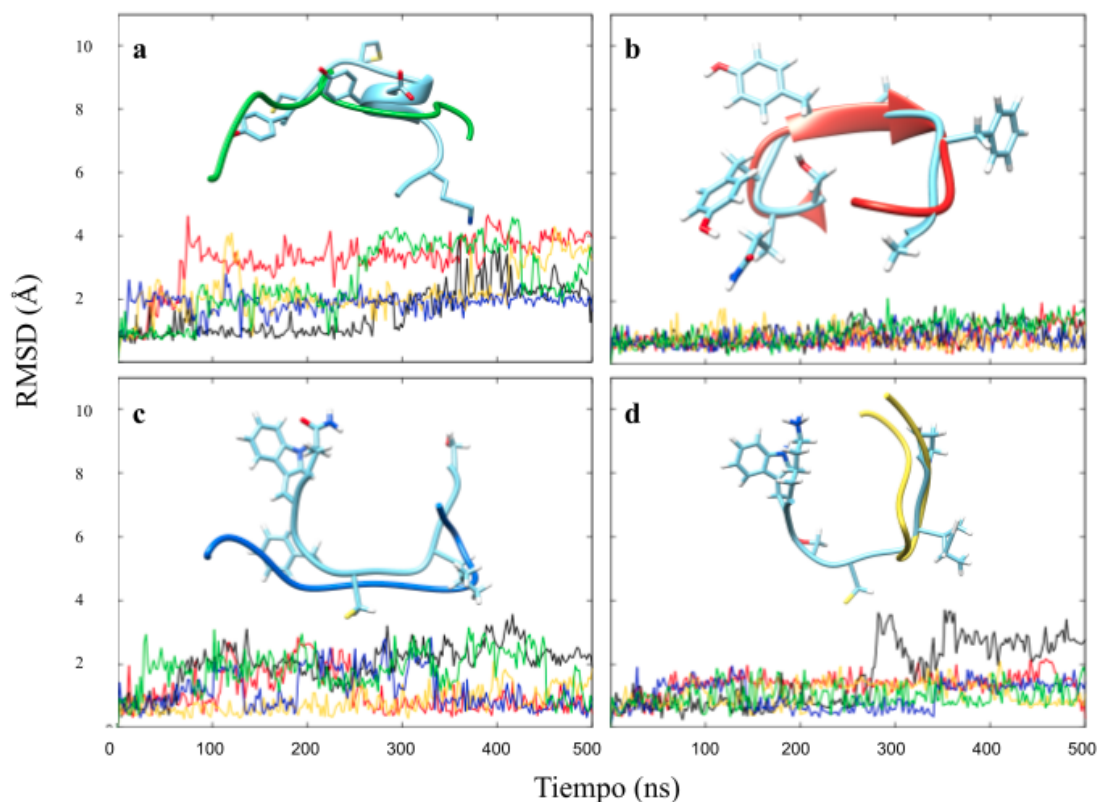
Por otro lado, se hizo este mismo análisis para los dominios en la proteína. En este caso los dominios A y D tuvieron ambos seis P-H con dichas características, mientras que tanto B y C presentaron 10 P-H. Esto arrojó un  $R^2$  de 0.7.



**Figura 11.** Gráfica de la regresión lineal entre la cantidad de P-H y  $T_m$  de los dominios aislados y en la proteína. El color de cada línea corresponde a las **Tablas 5 y 6**.

## Análisis del sitio de unión.

Otros P-H que se presentan en los cuatro dominios tanto aislados como en la proteína completa, son los P-H formados por la Ser19 del dominio A aislado y residuos aromáticos como Tyr, Trp y Phe, los cuales forman P-H persistentes y recurrentes a lo largo de todas las temperaturas en los cuatro dominios (**Figura 12**). Estos residuos forman parte del sitio de unión y tienen un papel importante en la unión con el ligando, ya que los residuos aromáticos mantienen interacciones hidrofóbicas, mientras que la Ser<sup>19</sup>, conservada en los cuatro dominios, forma el mismo P-H con el carbohidrato que está interactuando [Tellez et al., 2011]. Para analizar el comportamiento del sitio de unión, con el fin de observar si existe un cambio respecto a los dominios completos, se calculó la RMSD sobre los átomos pesados de los residuos Ser<sup>19</sup>-Cys<sup>31</sup> a 120°. Se calculó sólo a esta temperatura dado que ya se sabe que el núcleo de los dominios (residuos: 8-35) a 20° no sufre ninguna perturbación conformacional. Además de calcular el RMSD, también se obtuvo la estructura promedio de los conformeros correspondientes al rango de valores de RMSD más altos para analizar de manera cualitativa la diferencia del sitio de unión respecto a la estructura de referencia.



**Figura 12.** RMSD y estructura promedio del sitio de unión de los dominios a  $120^{\circ}$  C. Cada color en las gráficas representa las diferentes réplicas que se corrieron para cada temperatura. La letra en la esquina superior derecha, indica el dominio de las trayectorias. Cada estructura se obtuvo a partir del promedio de conformeros correspondientes a los valores de RMSD más altos, para observar la diferencia respecto a la estructura de referencia (azul claro).

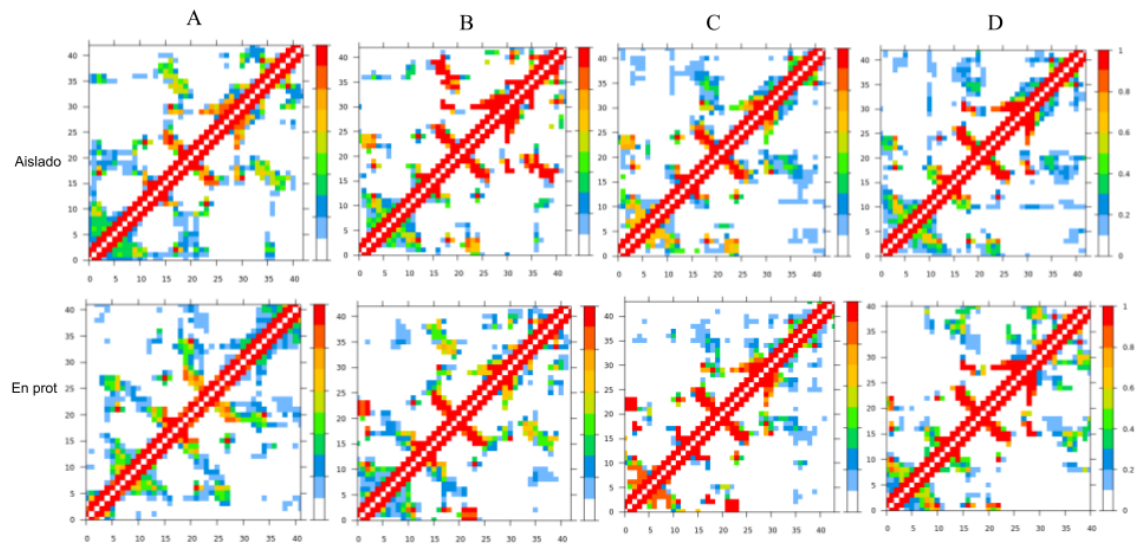
En la **Figura 12.a** se observa como el sitio del dominio A es el que más perturbaciones presenta en casi todas sus réplicas. Estos cambios en la estructura se deben a que el péptido se expande y pierde el plegamiento nativo (estructura verde). Lo mismo ocurre en el dominio C a una menor escala (**Figura 12.c**), ya que sus valores de RMSD están por debajo de los 4 Å. A pesar de que en los resultados anteriores, el dominio D siempre había confirmado ser el menos termoestable, su sitio de unión parece tener pocas perturbaciones conformacionales (**Figura 12.d**). Únicamente en una de las réplicas mostró cambiar el plegamiento a una compactación en el sitio y una rotación de  $90^{\circ}$ , respecto a la estructura cristalográfica (estructura azul). Por último, el sitio de unión del dominio B muestra valores de RMSD por debajo de los 2 Å, por lo que permanece prácticamente sin ninguna perturbación conformacional a pesar de estar a  $120^{\circ}$ .

## Cáculo de interacciones intermoleculares generales

Dado que los P-H no son el único tipo de interacciones moleculares que hay que tomar en cuenta, se hizo un análisis de los contactos ¿no polares? interdominios para identificar zonas en las que se concentra la mayor cantidad de interacciones al igual que observar el tiempo que prevalecen en las trayectorias y ver su variación respecto a la temperatura. Se cuantificó la fracción de presencia de cada interacción de los cuatro dominios, tanto aislados como en la proteína completa a las temperaturas de análisis y se graficaron en una matriz de contactos, proyectando la secuencia de los dominios en los dos ejes de la matriz (**Figura 13**). A través de un gradiente de color, se distingue la duración de cada interacción. Las matrices revelaron que las regiones que mantuvieron interacciones más recurrentes en todos los dominios son: los cinco primeros residuos del AT X<sup>1</sup>-Gly<sup>5</sup> con los residuos Ser<sup>19</sup>-Cys<sup>24</sup> del sitio de unión. También los residuos Asn<sup>14</sup>-Ser<sup>19</sup> del núcleo que interactúan con el CT Glu<sup>29</sup>-Ala<sup>39</sup>. Por último, los residuos aromáticos del sitio de unión 19-30 interactúan entre sí debido a su volumen y su ubicación de un espacio estrecho en la proteína.

No es de extrañar que las regiones equivalentes mantengan los mismos contactos en todos los dominios, tomando en cuenta que tienen un índice de similitud en secuencia del 46%. Se observó que el aumento de temperatura provoca un cambio en el comportamiento de los contactos, ya que incrementa la cantidad total de interacciones pero la mayoría cambia a una fracción de presencia  $\geq 20\%$ . Esto se debe a que el aumento de temperatura incrementa la energía cinética del sistema y por lo tanto, el movimiento de los átomos es más rápido [Zhang et al., 2021]. Al hacer esto, los dominios cambian más rápido de conformación, haciendo que se formen y rompan interacciones no específicas, especialmente en las regiones terminales y en la región X<sup>29</sup>-X<sup>39</sup>. Este rompimiento y formación de interacciones está dado por residuos puntuales, no en rangos de residuos continuos.

En la **Figura 13** se muestran las matrices de contactos para 120° C. En el dominio A aislado, las interacciones entre el rango de residuos Glu<sup>29</sup>-Ala<sup>39</sup>, Asn<sup>14</sup>-Ser<sup>19</sup> tienen una fracción de duración del 50%, mientras que en la proteína completa, la mayoría disminuye a fracciones  $\geq 20\%$ , además de cambiar el patrón de interacciones. Este dominio es el segundo con fracciones altas, ya que el B mantiene las mismas interacciones tanto aislado como en la proteína completa en el 100% de las trayectorias. Estas interacciones en los



**Figura 13.** Matrices de contactos interatómicos con fracción de presencia a  $120^{\circ}$  C. Las columnas corresponden a los dominios que se indica en la parte superior de la figura y los renglones corresponden a los dominios aislados y en la proteína como se indica en la parte izquierda.

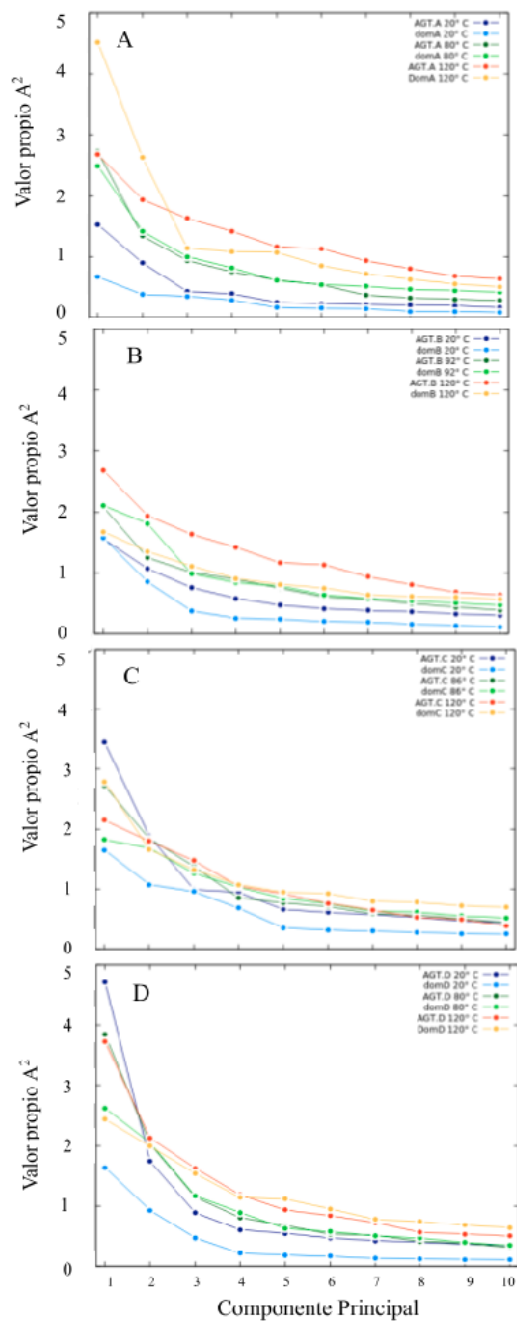
dominios C y D aislados están  $< 20\%$ , mientras que en la proteína el dominio C pierde por completo parte de estas interacciones, además de que el D aumenta la fracción de algunas de estas interacciones, dentro del rango mencionado anteriormente.

## Valores propios de dPCA en los dominios aislados y en la proteína

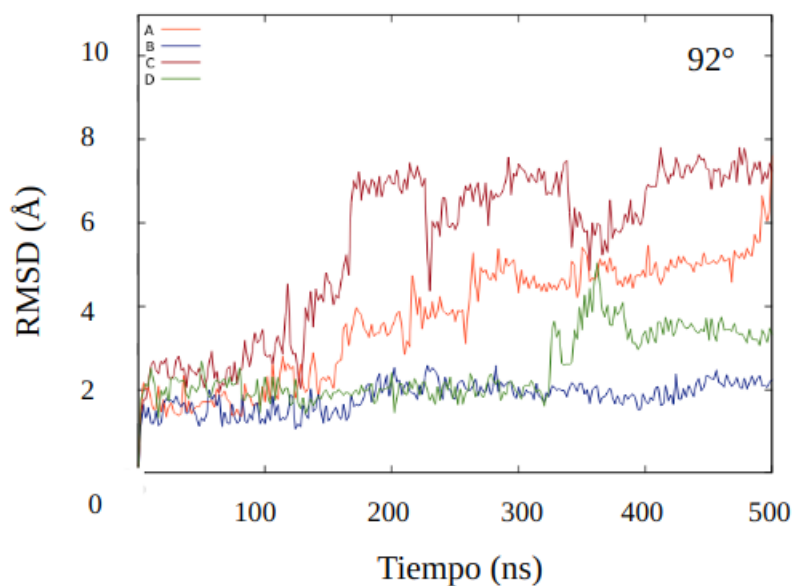
El último análisis comparativo entre los dominios aislados y en la proteína completa que se realizó, consistió en determinar si los dominios mostraban movimientos colectivos más amplios cuando se encuentran aislados. Para esto se hizo nuevamente un análisis de dPCA a  $20^{\circ}$ ,  $T_m$  respectiva y  $120^{\circ}$  C y se compararon los valores propios de cada dominio aislado y en la proteína [Papaleo et al., 2009].

Los valores propios de los PC's de las trayectorias del dominio A cuando se encuentra aislado se mantienen por debajo de los que pertenecen al dominio cuando está en la proteína. Esta relación se invierte a  $120^{\circ}$ , ya que el valor propio del PC 1 del dominio aislado es de 4.6, mientras que cuando está en la proteína completa es de 2.8, al igual que los de la  $T_m$ . El dominio B no presenta ninguna diferencia a  $20^{\circ}$  y  $T_m$  entre el





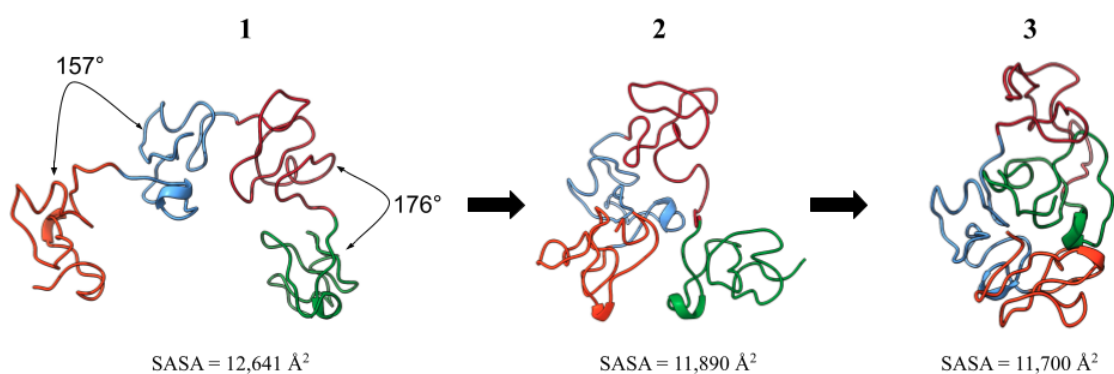
**Figura 14.** Valores propios de los PC de cada dominio a distintas temperaturas. La letra en la parte superior izquierda indica el dominio al que corresponde la gráfica.



**Figura 15.** RMSD de los dominios en la proteína a 92° C. Cada color representa cada unode los dominios, con el código que se indica en la parte superior izquierda.

dominio aislado o en la proteína, por lo que sus movimientos colectivos son de la misma magnitud. Aunque el valor propio del PC 1 del dominio aislado cuando está a 120°, es más bajo que cuando se encuentra a su  $T_m$ , indicando la posibilidad de que sea necesario realizar dinámicas de mayor duración para muestrear de manera satisfactoria el espacio conformacional a esta temperatura intermedia. Por otro lado, el dominio C presenta un comportamiento bastante similar al dominio D, en el que el estado con más varianza es cuando se encuentra a 20° en la proteína, incluso por encima de los valores propios de las trayectorias a 120°. No obstante, ambos dominios también presentan el valor propio más pequeño cuando están aislados. Por otro lado, el dominio C es el único que presenta mayor flexibilidad a 120° cuando se encuentra en la proteína completa.

A partir del dPCA fue posible determinar la diferencia y la magnitud de los cambios conformacionales de los dominios aislados, respecto a los dominios en la proteína. Para observar el comportamiento de los dominios en la proteína a lo largo de las trayectorias, se calculó el RMSD para el rango de residuos de la proteína correspondiente a cada dominio a distintas temperaturas (**Figura 15**). Se observó que a baja temperatura, los valores son congruentes con los resultados de RMSD de los dominios aislados, obtenidos previamente. Conforme aumentó la temperatura se comenzó a observar un comportamiento anómalo por parte del dominio C, en el que pareciera que es el dominio con mayor desnaturalización ya que presentó valores de RMSD más altos en la mayoría de las réplicas a partir de la



**Figura 16.** Estructuras promedio que describen el cambio conformacional (compactación) de la proteína a altas temperaturas. Este cambio conformacional, causa una disminución en el SASA y una alteración en el plegamiento del dominio AGT.C

temperatura de 80° y hasta 120°. Al analizar estas trayectorias en un visualizador molecular (**Figura 16.**), se observó que dichos valores de RMSD del dominio C se deben a una conformación recurrente en la mayoría de las réplicas, que consiste en una compactación de la proteína, en la que los dominios de los extremos (dominios A y D) se acercan, provocando un cambio en el plegamiento del dominio C en los segmentos Glu<sup>87</sup>-Cys<sup>142</sup> y Cys<sup>118</sup> - Thr<sup>129</sup>. Para que se lleve a cabo este acercamiento, el dominio C sufre un cambio conformacional, el cual altera el plegamiento y provoca que se compacte la proteína. Al ser tan recurrente la conformación compactada, sugeriría que es una conformación con mayor estabilidad. Para medir el nivel de compactación se calculó la superficie de área expuesta al solvente (SASA) de los tres estados conformacionales que describían este comportamiento. En la Figura 16 se observan las tres estructuras promedio con su respectivo valor de SASA. Se observa como los dominios AGT.A y AGT.D se pliegan hacia el interior de la proteína, lo que causa una compactación y por lo tanto una disminución de SASA. Este comportamiento se observó en las temperaturas de 86°, 92° y 120° C.

# Discusión

La AGT es la lectina que ha sido más caracterizada a nivel molecular. Su capacidad de unirse a ligandos con alto nivel de especificidad y con afinidad en el rango  $\mu\text{M}$ , han hecho que esta proteína haya sido utilizada en una gran variedad de aplicaciones en estudios de las ramas biomédica y biotecnológica. Sin embargo, a pesar de todos estos estudios, se desconocen los mecanismos de estabilidad de esta proteína, conocimiento que abriría la posibilidad de aumentar su eficiencia en tratamientos y posibilidades de aplicación. Por lo tanto, es necesario encontrar dichos mecanismos, con el fin de hacerla más eficiente como acarreador de fármacos y seguir explotando su potencial en la investigación biomédica y biotecnológica. En la presente tesis se realizó la caracterización *in silico* del desplegamiento térmico de la AGT mediante simulaciones de dinámica molecular de los cuatro dominios aislados y de la proteína completa, a distintas temperaturas. Mediante dPCA y modelos de estados de Markov se definió el estado nativo y desnaturalizado de la proteína y de los dominios aislados con base en la energía libre de cada conformero. Los cuatro PdS de cada dominio confirmaron ser un componente importante en la estabilidad térmica, aunque uno de los determinantes en la estabilidad propia de los dominios parecieron ser los P-H. Se observó que los dominios aislados presentan movimientos colectivos de menor amplitud que cuando se encuentran en la proteína. Por último, se observó que la proteína tiene conformaciones recurrentes que consisten en un acercamiento de los dominios de cada extremo, provocando una compactación.

La definición del estado plegado y desplegado que se determinó para los dominios aislados y en la proteína utilizando los modelos de estados de Markov [Zimmerman et al., 2017], mostró que el estado plegado se acerca a la estructura cristalográfica. Esto también se vio reflejado en la RMSD, ya que el estado plegado de los dominios aislados presentaba un

valor más bajo respecto a la estructura cristalográfica que el estado desnaturalizado. Estos valores fueron mayores cuando los dominios se encontraban en la proteína, ya que el desplegamiento no era únicamente en el AT, sino que el asa que conecta con el dominio siguiente parece afectar el plegamiento del CT. El macroestado que contiene a los conformeros plegados se encontró en la cuenca de atracción más poblada del PEL de cada dominio, indicando que los conformeros que pertenecen a ese macroestado presentan una menor energía respecto a los demás conformeros de las trayectorias.

En los dominios aislados, la diferencia en estructuras más notorias es en el CT para los dominios A, C y D, mientras que para B es el AT. El terminal contrario al desnaturalizado se mantenía prácticamente con el mismo plegamiento, a pesar de la alta temperatura a la que fue sometida. Esto indica un alto nivel de estabilidad, que como se mencionó antes, de la que son responsables los PdS. Una interacción de ese tipo puede contribuir de 5 a 10 kcal/mol a la estabilidad de una proteína en su estado nativo [Zavodszky et al., 2001]. Según lo reportado por [Pace et al., 1988] y [Zavodszky et al., 2001], el incremento en la estabilidad de la proteína se ve reflejado principalmente en la disminución de la entropía del estado desnaturalizado. La influencia de los PdS en el comportamiento de los dominios se analizó mediante el RMSF. Los PdS cumplen con la función de disminuir grados de libertad en la proteína, lo que sugiere que el comportamiento es congruente con lo mencionado por dichos investigadores. La restricción por parte de los PdS también se observó en el análisis que se hizo sobre los sitios de unión de los dominios, los cuales resultaron ser la parte más estable en todas las simulaciones. Esto abre la pregunta: ¿La restricción de los PdS tendría un papel importante en la especificidad por carbohidratos, característica de esta proteína? Ya que además de brindarle estabilidad, moldean la topología de los dominios y del sitio de unión. Se han realizado estudios en los que, al mutar los residuos de los sitios de unión de la AGT, se altera la red de interacción de los P-H, lo que conlleva un cambio en su especificidad por el N-acetil-ácido-neuroamínico (Neu5Aca(2-3)Gal) [Parasuraman et al., 2014]. No obstante, no se ha estudiado sin los PdS.

Con el análisis del RMSF es posible ver la importancia de los PdS en la estabilidad de los dominios y la contribución de cada residuo al movimiento de los dominios. Sin embargo, no brinda una respuesta a la variación en los valores de la  $T_m$ , ya que dichos puentes se encuentran en la misma posición de cada dominio, por lo que debe haber otro componente que aporte a las diferencias de las termoestabilidades.

Por otro lado, el análisis de regresión lineal de los P-H con la  $T_m$  de cada dominio y la

cuantificación de los contactos intramoleculares sugieren ser congruentes con lo que se tiene reportado sobre la aportación de los P-H en la estabilidad de proteínas. En el presente trabajo se puso especial atención a las interacciones intramoleculares como contactos hidrofóbicos y P-H. Se tiene reportado que la contribución entálpica de P-H en la superficie de una proteína es de  $-6$  kJ/mol, mientras que las interacciones al interior de la proteína aportan  $-25$  kJ/mol [Cramer et al., 2020]. En el análisis de los P-H de cada dominio a  $20^\circ$ , los P-H eran más abundantes y con una mayor fracción de vida, a diferencia de cuando se encuentran a  $120^\circ$ , donde las interacciones eran menos abundantes y menos persistentes, lo que implica que a  $20^\circ$  la aportación entálpica es mayor que a  $120^\circ$ .

Los valores propios de los componentes principales calculados para los dominios aislados tienden a ser más bajos que cuando se encuentran en la proteína. Esto indica que sus movimientos colectivos son más amplios que cuando se encuentran aislados y, por lo tanto, serían menos flexibles al estar aislados. Los resultados de [Bhaskara and Srinivasan, 2011] son un contra ejemplo adecuado para estos resultados, ya que en dicho estudio compararon el cambio de estabildades entre dominios aislados que normalmente pertenecen a proteínas con múltiples dominios vs proteínas sin alteración en su secuencia. A través de técnicas computacionales. Calcularon la energía libre de plegamiento ( $\Delta G_{pleg}$ ) para los dominios pertenecientes a proteínas multidominios estudiados en aislamiento y obtuvieron valores de  $7.70 \pm 0.96$  kcal/mol, mientras que formando la proteína completa presentaron valores de  $-6.57 \pm 2.37$  kcal/mol. Cabe recalcar que las proteínas seleccionadas para este estudio eran proteínas capaces de mantener interacciones de tipo hidrofóbicos entre los dominios, lo cual contribuye favorablemente a la estabilidad. El número de interacciones interdominio calculadas para la AGT fue de  $< 10$  y con fracciones de presencia que perduraban  $< 10\%$  de las trayectorias, por lo que no se tomaron en cuenta. [Bhaskara and Srinivasan, 2011] también reportan que la disminución en estabilidad de dominios pertenecientes a proteínas multidominio en estado aislado, puede deberse a una falta de energía de solvatación, es decir, que la cantidad de residuos hidrofóbicos expuestos al medio sea mayor que los residuos polares. Esto lo corroboraron al mutar residuos polares por hidrofóbicos en la parte de interacción interdominio, por lo que al interactuar con los demás dominios no quedan expuestos al medio. Si los dominios se aíslan, son privados de esas interacciones y los grupos hidrofóbicos quedan expuestos a medio, lo que conlleva a una solvatación pobre y  $\Delta G > 0$ . En el presente trabajo, los dominios aislados de la AGT tienen un área polar  $\approx 50\%$ ,

por lo que su energía de solvatación sería más favorable que los dominios del estudio de [Bhaskara and Srinivasan, 2011]. No obstante, cuando se encuentran en la proteína completa su área polar también equivaldría al 50 %, por lo que no se esperaría un cambio en la estabilidad. Otro factor que podría estar afectando la estabilidad de los dominios cuando se encuentran en la proteína completa son los movimientos que puedan causar los demás dominios, al estar conectados por la cadena principal. Estos resultados dejan bases para futuros estudios sobre el cambio de estabilidad y movimiento de los dominios en estado aislado de la AGT.

# Conclusiones

- Los dominios de la AGT presentan una alta termoestabilidad, mediada en gran parte por los PdS. No obstante, uno de los factores determinantes sobre la diferencia en las estabildades son los P-H e interacciones intermoleculares.
- El sitio de unión de los dominios de la AGT, resulta ser la parte más estable de toda la proteína.
- Los dominios de la AGT aislados, presentan movimientos colectivos de menor amplitud que cuando se encuentran en la proteína.



# Perspectivas

- Correlacionar los datos obtenidos con datos experimentales, obtenidos por calorimetría.
- Evaluar el comportamiento de los dominios aislados y la proteína sin los puentes disulfuro.
- Realizar experimentos de acoplamiento molecular para encontrar fármacos candidatos que puedan ser acarreados por la AGT y sus dominios aislados.

# Bibliografía

- [Adams, 2001] Adams, J. (2001). Bonding energy models. *Encyclopedia of Materials: Science and Technology*, pages 763–767.
- [Altis et al., 2007] Altis, A., Nguyen, P. H., Hegger, R., and Stock, G. (2007). Dihedral angle principal component analysis of molecular dynamics simulations. *The Journal of chemical physics*, 126(24):244111.
- [Aub et al., 1965] Aub, J. C., Sanford, B. H., and Cote, M. N. (1965). Studies on reactivity of tumor and normal cells to a wheat germ agglutinin. *Proceedings of the National Academy of Sciences*, 54(2):396–399.
- [Balčiūnaitė M and Dzikaras, 2021] Balčiūnaitė M, G. and Dzikaras, M. (2021). Wheat germ agglutinin—from toxicity to biomedical applications. *Applied Sciences*, 11(2):884.
- [Beck and Daggett, 2004] Beck, D. A. and Daggett, V. (2004). Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods*, 34(1):112–120.
- [Becker and Karplus, 2006] Becker, O. M. and Karplus, M. (2006). *Guide to biomolecular simulations*, volume 4. Springer Science & Business Media.
- [Bekker et al., 1993] Bekker, H., Berendsen, H., Dijkstra, E., Achterop, S., Vondrumen, R., VANDERSPOEL, D., Sijbers, A., Keegstra, H., and Renardus, M. (1993). Gromacs—a parallel computer for molecular-dynamics simulations. In *4th International Conference on Computational Physics (PC 92)*, pages 252–256. World Scientific Publishing.
- [Berman, 2000] Berman, H M. Westbrook, J. F. Z. G. G. T. B. H. W. I. S. P. B. (2000). The protein data bank. *Nucleic Acids Research*, 28:235–242.
- [Bhaskara and Srinivasan, 2011] Bhaskara, R. M. and Srinivasan, N. (2011). Stability of domain structures in multi-domain proteins. *Scientific reports*, 1(1):1–9.

- [Bhattacharyya and Brewer, 1988] Bhattacharyya, L. and Brewer, C. F. (1988). Lectin-carbohydrate interactions: Studies of the nature of hydrogen bonding between d-galactose and certain d-galactose-specific lectins, and between d-mannose and concanavalin a. *European journal of biochemistry*, 176(1):207–212.
- [Bowie, 2011] Bowie, J. U. (2011). Membrane protein folding: how important are hydrogen bonds? *Current opinion in structural biology*, 21(1):42–49.
- [Broadwell et al., 1988] Broadwell, R. D., Balin, B. J., and Saleman, M. (1988). Transcytotic pathway for blood-borne protein through the blood-brain barrier. *Proceedings of the National Academy of Sciences*, 85(2):632–636.
- [Cavada et al., 2020] Cavada, B. S., Pinto-Junior, V. R., Osterne, V. J. S., Oliveira, M. V., Lossio, C. F., Silva, M. T. L., Bari, A. U., Lima, L. D., Souza-Filho, C. H. D., and Nascimento, K. S. (2020). Comprehensive review on caelsalpinioideae lectins: From purification to biological activities. *International Journal of Biological Macromolecules*, 162:333–348.
- [Chettri et al., 2021] Chettri, D., Boro, M., Sarkar, L., and Verma, A. K. (2021). Lectins: Biological significance to biotechnological application. *Carbohydrate Research*, 506:108367.
- [Cramer et al., 2020] Cramer, J., Jiang, X., Schönemann, W., Silbermann, M., Zihlmann, P., Siegrist, S., Fiege, B., Jakob, R. P., Rabbani, S., Maier, T., et al. (2020). Enhancing the enthalpic contribution of hydrogen bonds by solvent shielding. *RSC chemical biology*, 1(4):281–287.
- [Dalla Pellegrina et al., 2005] Dalla Pellegrina, C., Rizzi, C., Mosconi, S., Zoccatelli, G., Peruffo, A., and Chignola, R. (2005). Plant lectins as carriers for oral drugs: Is wheat germ agglutinin a suitable candidate? *Toxicology and applied pharmacology*, 207(2):170–178.
- [Eastman et al., 2017] Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., et al. (2017). Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7):e1005659.

- [Feige et al., 2018] Feige, M. J., Braakman, I., and Hendershot, L. M. (2018). Disulfide bonds in protein folding and stability.
- [Fuglebakk et al., 2012] Fuglebakk, E., Echave, J., and Reuter, N. (2012). Measuring and comparing structural fluctuation patterns in large protein datasets. *Bioinformatics*, 28(19):2431–2440.
- [Gao et al., 2006] Gao, X., Tao, W., Lu, W., Zhang, Q., Zhang, Y., Jiang, X., and Fu, S. (2006). Lectin-conjugated peg–pla nanoparticles: preparation and brain delivery after intranasal administration. *Biomaterials*, 27(18):3482–3490.
- [Gao et al., 2007] Gao, X., Wu, B., Zhang, Q., Chen, J., Zhu, J., Zhang, W., Rong, Z., Chen, H., and Jiang, X. (2007). Brain delivery of vasoactive intestinal peptide enhanced with the nanoparticles conjugated with wheat germ agglutinin following intranasal administration. *Journal of Controlled Release*, 121(3):156–167.
- [Gómez-González et al., 2011] Gómez-González, B., Larios, H. M., and Escobar, A. (2011). Increased transvascular transport of wga-peroxidase after chronic perinatal stress in the hippocampal microvasculature of the rat. *International Journal of Developmental Neuroscience*, 29(8):839–846.
- [Hernández et al., 1995] Hernández, A. A., Rojo-Domínguez, A., Soriano-García, M., and Rodríguez-Romero, A. (1995). The thermal unfolding of hevein, a small disulfide-rich protein. *European journal of biochemistry*, 228(3):649–652.
- [Huang and MacKerell Jr, 2013] Huang, J. and MacKerell Jr, A. D. (2013). Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of computational chemistry*, 34(25):2135–2145.
- [Knapp et al., 2011] Knapp, B., Frantal, S., Cibena, M., Schreiner, W., and Bauer, P. (2011). Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible? *Journal of Computational Biology*, 18(8):997–1005.
- [Kumar et al., 2012] Kumar, K. K., Chandra, K. L. P., Sumanthi, J., Reddy, G. S., Shekar, P. C., Reddy, B., et al. (2012). Biological role of lectins: A review. *Journal of orofacial sciences*, 4(1):20.

- [Larsson et al., 2020] Larsson, P., Kneiszl, R. C., and Marklund, E. G. (2020). Mkvsites: A tool for creating gromacs virtual sites parameters to increase performance in all-atom molecular dynamics simulations. *Journal of Computational Chemistry*, 41(16):1564–1569.
- [Leyva et al., 2019] Leyva, E., Medrano-Cerano, J. L., Cano-Sánchez, P., López-González, I., Gómez-Velasco, H., del Río-Portilla, F., and García-Hernández, E. (2019). Bacterial expression, purification and biophysical characterization of wheat germ agglutinin and its four hevein-like domains. *Biopolymers*, 110(1):e23242.
- [Lis and Sharon, 1998] Lis, H. and Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chemical reviews*, 98(2):637–674.
- [Liu et al., 2011] Liu, S.-L., Zhang, Z.-L., Sun, E.-Z., Peng, J., Xie, M., Tian, Z.-Q., Lin, Y., and Pang, D.-W. (2011). Visualizing the endocytic and exocytic processes of wheat germ agglutinin by quantum dot-based single-particle tracking. *Biomaterials*, 32(30):7616–7624.
- [Maag et al., 2021] Maag, D., Putzu, M., Gómez-Flores, C. L., Gräter, F., Elstner, M., and Kubař, T. (2021). Electrostatic interactions contribute to the control of intramolecular thiol–disulfide isomerization in a protein. *Physical Chemistry Chemical Physics*, 23(46):26366–26375.
- [Miotto et al., 2019] Miotto, M., Olimpieri, P. P., Di Rienzo, L., Ambrosetti, F., Corsi, P., Lepore, R., Tartaglia, G. G., and Milanetti, E. (2019). Insights on protein thermal stability: a graph representation of molecular interactions. *Bioinformatics*, 35(15):2569–2577.
- [Murakami et al., 2014] Murakami, . Y., Hasegawa, Y., Nagano, K., and Yoshimura, F. (2014). Characterization of wheat germ agglutinin lectin-reactive glycosylated ompa-like proteins derived from porphyromonas gingivalis. *Infection and Immunity*, 82(11):4563–4571.
- [Muraki et al., 2002] Muraki, M., Ishimura, M., and Harata, K. (2002). Interactions of wheat-germ agglutinin with glnac $\beta$ 1, 6gal sequence. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1569(1-3):10–20.

- [Neutsch et al., 2014] Neutsch, L., Eggenreich, B., Herwig, E., Marchetti-Deschmann, M., Allmaier, G., Gabor, F., and Wirth, M. (2014). Biomimetic delivery strategies at the urothelium: targeted cytoinvasion in bladder cancer cells via lectin bioconjugates. *Pharmaceutical research*, 31(3):819–832.
- [Nizet et al., 2017] Nizet, V., Varki, A., and Aebi, M. (2017). Microbial lectins: hemagglutinins, adhesins, and toxins.
- [Pace et al., 2014] Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., et al. (2014). Contribution of hydrogen bonds to protein stability. *Protein Science*, 23(5):652–661.
- [Pace et al., 1988] Pace, C. N., Grimsley, G. R., Thomson, J. A., and Barnett, B. J. (1988). Conformational stability and activity of ribonuclease t1 with zero, one, and two intact disulfide bonds. *Journal of Biological Chemistry*, 263(24):11820–11825.
- [Panja et al., 2020] Panja, A. S., Maiti, S., and Bandyopadhyay, B. (2020). Protein stability governed by its structural plasticity is inferred by physicochemical factors and salt bridges. *Scientific reports*, 10(1):1–9.
- [Papaleo et al., 2009] Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., and De Gioia, L. (2009). Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case. *Journal of molecular graphics and modelling*, 27(8):889–899.
- [Parasuraman et al., 2014] Parasuraman, P., Murugan, V., Selvin, J. F., Gromiha, M. M., Fukui, K., and Veluraja, K. (2014). Insights into the binding specificity of wild type and mutated wheat germ agglutinin towards neu5ac $\alpha$  (2-3) gal: a study by in silico mutations and molecular dynamics simulations. *Journal of Molecular Recognition*, 27(8):482–492.
- [Parrinello and Rahman, 1981] Parrinello, M. and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190.
- [Pettersen et al., 2004] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612.

- [Price and Brooks III, 2004] Price, D. J. and Brooks III, C. L. (2004). A modified tip3p water potential for simulation with ewald summation. *The Journal of chemical physics*, 121(20):10096–10103.
- [Ragone, 2001] Ragone, R. (2001). Hydrogen-bonding classes in proteins and their contribution to the unfolding reaction. *Protein Science*, 10(10):2075–2082.
- [Rizzuti, 2022] Rizzuti, B. (2022). Molecular simulations of proteins: From simplified physical interactions to complex biological phenomena. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, page 140757.
- [Roe and Cheatham III, 2013] Roe, D. R. and Cheatham III, T. E. (2013). Ptraaj and cpptraj: software for processing and analysis of molecular dynamics trajectory data. *Journal of chemical theory and computation*, 9(7):3084–3095.
- [Ruvinsky et al., 2012] Ruvinsky, A. M., Kirys, T., Tuzikov, A. V., and Vakser, I. A. (2012). Structure fluctuations and conformational changes in protein binding. *Journal of bioinformatics and computational biology*, 10(02):1241002.
- [Ryva et al., 2019] Ryva, B., Zhang, K., Asthana, A., Wong, D., Vicioso, Y., and Parameswaran, R. (2019). Wheat germ agglutinin as a potential therapeutic agent for leukemia. *Frontiers in oncology*, 9:100.
- [Sharon and Halina, 2004] Sharon, N. and Halina, L. (2004). History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, 14(11):53–62.
- [Tellez et al., 2011] Tellez, M. d. C. P. et al. (2011). Plegamiento y homodimerización de la aglutinina de germén de trigo.
- [Teuschl et al., 2014] Teuschl, A. H., Neutsch, L., Monforte, X., Rünzler, D., Van Griensven, M., Gabor, F., and Redl, H. (2014). Enhanced cell adhesion on silk fibroin via lectin surface modification. *Acta biomaterialia*, 10(6):2506–2517.
- [Toofanny and Daggett, 2012] Toofanny, R. D. and Daggett, V. (2012). Understanding protein unfolding from molecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(3):405–423.
- [Uehara and Park, 2004] Uehara, T. and Park, J. T. (2004). The n-acetyl-d-glucosamine kinase of escherichia coli and its role in murein recycling. *Journal of bacteriology*, 186(21):7273–7279.

- [Wirth et al., 2002] Wirth, M., Kneuer, C., Lehr, C., and Gabor, F. (2002). Lectin-mediated drug delivery: discrimination between cytoadhesion and cytoinvasion and evidence for lysosomal accumulation of wheat germ agglutinin in the caco-2 model. *Journal of drug targeting*, 10(6):439–448.
- [Wu et al., 2006] Wu, J. H., Singh, T., Herp, A., and Wu, A. M. (2006). Carbohydrate recognition factors of the lectin domains present in the ricinus communis toxic protein (ricin). *Biochimie*, 88(2):201–217.
- [Zavodszky et al., 2001] Zavodszky, M., Chen, C.-W., Huang, J.-K., Zolkiewski, M., Wen, L., and Krishnamoorthi, R. (2001). Disulfide bond effects on protein stability: Designed variants of cucurbita maxima trypsin inhibitor-v. *Protein Science*, 10(1):149–160.
- [Zhang et al., 2021] Zhang, G., Yang, G., Li, S., Shen, Q., Wang, H., Li, Z., Zhou, Y., and Ye, W. (2021). Effects of hydration and temperature on the microstructure and transport properties of nafion polyelectrolyte membrane: A molecular dynamics simulation. *Membranes*, 11(9):695.
- [Zimmerman et al., 2017] Zimmerman, M. I., Hart, K. M., Sibbald, C. A., Frederick, T. E., Jimah, J. R., Knoverek, C. R., Tolia, N. H., and Bowman, G. R. (2017). Prediction of new stabilizing mutations based on mechanistic insights from markov state models. *ACS central science*, 3(12):1311–1321.