



**INSTITUTO POTOSINO DE INVESTIGACIÓN  
CIENTÍFICA Y TECNOLÓGICA, A.C.**

**POSGRADO EN CIENCIAS EN BIOLOGIA MOLECULAR**

**Reconocimiento molecular en interacciones proteína-  
ligando: estabilidad de CGI-58 y estructura  
cristalográfica de SdrP/ADN**

Tesis que presenta

**Miriam Livier Llamas García**

Para obtener el grado de

**Doctora en Ciencias en Biología Molecular**

Director de la Tesis:

**Dr. Samuel Lara González**

San Luis Potosí, S.L.P., Agosto de 2022



## Constancia de aprobación de la tesis

La tesis “**Reconocimiento molecular en interacciones proteína-ligando: estabilidad de CGI-58 y estructura cristalográfica de SdrP/ADN**” presentada para obtener el Grado de Doctora en Ciencias en Biología Molecular fue elaborada por la **M. en C. Miriam Livier Llamas García** y aprobada el **12 de agosto de 2022** por los suscritos, designados por el Colegio de Profesores de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

---

**Dr. Samuel Lara González**

Director

---

**Dra. Irene Beatríz Castaño Navarro**

Miembro del Comité Tutorial

---

**Dr. Luis Antonio Salazar Olivo**

Miembro del Comité Tutorial

---

**Dr. Roberto Sánchez Olea**

Miembro del Comité Tutorial



## **Créditos Institucionales**

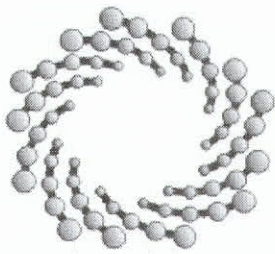
Esta tesis fue elaborada en el Laboratorio de Biología Estructural de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la dirección del Dr. Samuel Lara González.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología (395867) y del Instituto Potosino de Investigación Científica y Tecnológica, A. C.

Este trabajo fue financiado a través del proyecto de Ciencia Básica SEP-CONACYT No CB2011-168710 y por el proyecto CONACYT INFRA-2013-01 No 204373.

Agradecemos al Laboratorio Nacional de Biotecnología Agrícola, Médica y Ambiental (LANBAMA) del Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), por el apoyo en el servicio de secuenciación de ADN.

Agradecemos al Laboratorio Nacional de Estructura de Macromoléculas del Instituto de Química (LANEM-IQ) de la Universidad Nacional Autónoma de México (UNAM), por el apoyo en la colecta, reducción y análisis de los datos de difracción de rayos-X.



**IPICYT**

# Instituto Potosino de Investigación Científica y Tecnológica, A.C.

## Acta de Examen de Grado

La Secretaria Académica del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 131 del Libro Segundo de Actas de Exámenes de Grado del Programa de Doctorado en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 29 días del mes de agosto del año 2022, se reunió a las 12:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

<b>Dr. Luis Antonio Salazar Olivo</b>	<b>Presidente</b>	<b>IPICYT</b>
<b>Dra. Irene Beatriz Castaño Navarro</b>	<b>Secretaria</b>	<b>IPICYT</b>
<b>Dr. Roberto Sánchez Olea</b>	<b>Sinodal externo</b>	<b>UASLP</b>
<b>Dr. Samuel Lara González</b>	<b>Sinodal</b>	<b>IPICYT</b>

a fin de efectuar el examen, que para obtener el Grado de:

**DOCTORA EN CIENCIAS EN BIOLOGÍA MOLECULAR**

sustentó la C.

**Miriam Livier Llamas García**

sobre la Tesis intitulada:

*Reconocimiento molecular en interacciones proteína-ligando: estabilidad de CGI-58 y estructura cristalográfica de SdrP/ADN*

que se desarrolló bajo la dirección de


**Dr. Samuel Lara González**

El Jurado, después de deliberar, determinó

**APROBARLA**

Dándose por terminado el acto a las 14:27 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe la Secretaria Académica del Instituto.

A petición de la interesada y para los fines que a la misma convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 29 días del mes de agosto de 2022.

  
**Dra. Lina Raquel Riego Ruiz**  
Secretaria Académica

  
**Mtra. Ivonne Lizette Cuevas Vélez**  
Jefa del Departamento del Posgrado



## **Dedicatorias**

Esta tesis está dedicada a aquellos seres que me han hecho crecer y seguir evolucionando en la vida.

## **Agradecimientos**

A mi asesor de tesis, Dr. Samuel Lara González, por recibirme en su laboratorio y enseñarme de su experiencia, pero sobre todo por la confianza de poder aportar en sus proyectos.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) y al Instituto Potosino de Investigación Científica y Tecnológica (IPICYT) por la oportunidad de estudiar el doctorado.

A mi comité tutorial, Dra. Irene Castaño, Dr. Luis Salazar y Dr. Roberto Sánchez, por todas sus aportaciones al trabajo. Al Dr. Sergio Casas y Dr. Rubén López-Revilla, por todo el apoyo brindado para lograr concluir el posgrado.

A la Dra. Gabriela Montero, Dra. Claudia Benítez y Dr. Edgar Paéz por sus valiosas contribuciones en el capítulo 01 de la tesis.

Al Dr. Mauricio Esguerra Neira (Uppsala University) por su apoyo con el software x3ADN.

A Esaú Bojórquez Velázquez por todo su apoyo emocional y académico en los años que tomo hacer esta tesis.

A mi familia y amigos que estuvieron presente durante este viaje.

Gracias a todos los alumnos que estuvieron en breve estancia en el laboratorio y que participaron en la investigación desarrollada.

Al personal de biblioteca, que me apoyo en conseguir los artículos que forman parte de la bibliografía.

# Contenido

Constancia de aprobación de la tesis	ii
Créditos institucionales	iii
Acta de examen	iv
Dedicatorias	v
Agradecimientos	vi
Abreviaturas	viii
Resumen	x
Abstract	xi
Introducción	01
<b>Capítulo 1.</b> Mejoramiento de la estabilidad de la proteína CGI-58 de humano inducida por fosfomimetización mediante la mutación S237E	04
Objetivos	05
Introducción	06
Materiales y métodos	08
Resultados y discusión	17
Conclusiones	35
<b>Capítulo 2.</b> Estructura cristalográfica de SdrP de <i>Thermus thermophilus</i> en complejo con ADN: importancia del residuo Y180 en el reconocimiento del ADN	38
Objetivos	39
Introducción	40
Materiales y métodos	42
Resultados y discusión	47
Conclusiones y perspectivas	70
Referencias	71
Anexos	82

## Abreviaturas

<b>2-ME</b>	2-Mercaptoetanol
<b>ADN</b>	Ácido desoxirribonucleico
<b>ATGL</b>	Lipasa de triglicéridos en el adipocito
<b>cAMP</b>	Monofosfato de adenosina cíclico
<b>CAP</b>	Proteína Activadora por Catabolitos
<b>CBD</b>	Dominio de unión a cAMP
<b>CD</b>	Dicroísmo Circular
<b>CDS</b>	Síndrome de Chanarín-Dorfman
<b>CGI-58</b>	Identificación comparativa de genes 58
<b>Clp</b>	Proteína similar a CRP
<b>cNMP</b>	Monofosfato de nucleótidos cíclicos
<b>CRP/FNR</b>	Proteína receptora de cAMP/Proteína reguladora de la reducción de fumarato y nitrato
<b>CV</b>	Volumen de Columna
<b>DBD</b>	Dominio de Unión a ADN
<b>DLS</b>	Dispersión Dinámica de Luz
<b>DTT</b>	Ditiotreitol
<b>EDTA</b>	Ácido etilendiaminotetracético
<b>EMSA</b>	Ensayo de cambio de movilidad electroforética
<b>HSL</b>	Lipasa Sensible a Hormona
<b>HTH</b>	Hélice Giro Hélice
<b>IPTG</b>	Isopropil $\beta$ -D-tiogalactósido
<b>iTOL</b>	Árbol de la vida interactivo
<b>LB</b>	Luria Bertani
<b>LD</b>	Gota lipídica
<b>LdrP</b>	Proteína reguladora dependiente de LitR
<b>LitR</b>	Regulador de la transcripción inducida por la luz
<b>mARN</b>	Ácido Ribonucleico Mensajero
<b>MGL</b>	Lipasa de monoacilglicerol



<b>ML</b>	Máxima Verosimilitud
<b>MSA</b>	Alineamiento Múltiple de Secuencia
<b>ORF</b>	Marco de Lectura Abierto
<b>PCR</b>	Reacción en cadena de la polimerasa
<b>PDB</b>	Banco de Datos de Proteínas
<b>PK</b>	Proteína Quinasa
<b>PKA</b>	Proteína Quinasa A
<b>PLIN1</b>	Perilipina 1
<b>PP</b>	Proteína fosfatasa
<b>PTM</b>	Modificación post-traducciona
<b>R<sub>h</sub></b>	Radio hidrodinámico
<b>RMSD</b>	Raíz de la desviación cuadrática media
<b>RNAP</b>	ARN polimerasa
<b>SdrP</b>	Proteína reguladora dependiente de fase estacionaria
<b>SDS-PAGE</b>	Electroforesis en gel de poliacrilamida con dodecilsulfato de sodio
<b>SEC</b>	Cromatografía de exclusión por tamaño
<b>TAG</b>	Triacilglicérido
<b>Vfr</b>	Regulador del factor de virulencia
<b>WT</b>	Tipo Silvestre

## Resumen

### Reconocimiento molecular en interacciones proteína-ligando: estabilidad de CGI-58 y estructura cristalográfica de SdrP/ADN

El reconocimiento molecular es el proceso de interacción de macromoléculas biológicas entre sí o con varias moléculas pequeñas para formar un complejo específico, y constituye la base de todos los procesos en los organismos vivos, como en el catabolismo de triglicéridos y la expresión de genes. En el caso de la lipólisis, estudiamos a la proteína CGI-58, cuya función de activación de la lipasa ATGL es regulada por su interacción con la proteína de periferia PLIN1 y su unión a la gota lipídica. Una vez que la lipólisis es estimulada por catecolaminas, la fosforilación mediada por PKA permite la disociación del complejo CGI-58/PLIN1 y el reclutamiento de las lipasas ATGL, HSL y MGL para iniciar la liberación de los ácidos grasos almacenados en los triglicéridos. En este trabajo, evaluamos la estabilidad y solubilidad de la proteína CGI-58 WT y las mutantes 3WA y 3WA/S237E. Nuestros resultados mostraron que los triptófanos tienen un efecto en el estado de agregación de la proteína, ya que las mutantes 3WA y 3WA/S237E se encuentran mayormente como monómero en comparación con la proteína WT que tiende a agregarse. También determinamos que la mutante en el sitio de fosforilación favorece un incremento en la estabilidad térmica y el contenido de estructura secundaria en comparación con la 3WA, lo que sugiere que la fosforilación tiene un rol en la integridad de la proteína. En el caso de la transcripción, el factor SdrP que pertenece a la familia CRP/FNR, regula principalmente la expresión de genes del metabolismo energético y de reparación durante la fase estacionaria. El presente estudio muestra la estructura cristalográfica de SdrP en complejo con la secuencia consenso de ADN, determinamos que el “doblez” del ADN es diferente al reportado previamente para el factor de transcripción CAP, esto como resultado de los aminoácidos que realizan el contacto directo con el ADN (R160, E161, K165 y Y180) y las interacciones indirectas de la proteína con el ADN, que difieren respecto a CAP. Finalmente, determinamos la importancia de la participación de la Y180 en la interacción directa con el ADN, un residuo externo al motivo hélice-giro-hélice que no se había descrito en la familia CRP/FNR.

PALABRAS CLAVE.

CGI-58, estabilidad, fosfomimetización, SdrP, CRP/FNR, ADN

## **Abstract**

### **Molecular recognition in protein-ligand interactions: CGI-58 stability and SdrP/DNA X-ray crystal structure**

Molecular recognition is the process of interaction of biological macromolecules with each other or with several small molecules to form a specific complex and is the basis of all processes in living organisms, such as triglyceride catabolism and gene expression. In lipolysis, we studied the CGI-58 protein, whose function as a co-activating protein of ATGL is regulated by its interaction with periphery protein PLIN1 and its binding to the lipid droplet. Once lipolysis is stimulated by catecholamines, PKA-mediated phosphorylation triggers the dissociation of the CGI-58/PLIN1 complex, allowing the recruitment of ATGL, HSL, and MGL lipases to initiate the release of fatty acid stored in triglycerides. In this work, we evaluated the stability and solubility of the CGI-58 WT protein and the 3WA and 3WA/S237E mutants. Our results showed that tryptophan residues have an effect on the aggregation state of the protein, as the 3WA and 3WA/S237E mutants were mostly found as monomers compared to the WT protein, which tends to aggregate. We also established that the phosphorylation site mutant increases the thermal stability and secondary structure content of the 3WA/S237E compared to 3WA, suggesting that phosphorylation has a role in protein integrity. In transcription, the stationary phase-dependent regulatory protein (SdrP), which belongs to the CRP/FNR family, activates transcription at promoters related to energy metabolism and repair during the stationary phase. The present study shows the X-ray crystal structure of SdrP in complex with its consensus DNA binding sequence. We showed that DNA bending differs from what has been previously reported for CAP due to the differences observed in amino acids that participate in direct (R160, E161, K165, and Y180) and indirect recognition of DNA. Finally, we determined the participation of Y180 in the direct interaction with DNA, a residue located outside of the helix-turn-helix DNA binding motif that has not been described before in the CRP/FNR family.

#### **KEY WORDS.**

CGI-58, stability, phosphomimetic, SdrP, CRP/FNR, DNA

## Introducción

Las proteínas son una clase importante de macromoléculas biológicas, cuya diversidad funcional está determinada por su secuencia y la proporción de aminoácidos que las constituyen, por lo que se pueden desempeñar en procesos tan variados como transporte, soporte, comunicación, actividad enzimática, regulación, etc. Además de las propiedades químicas intrínsecas de una proteína, definidas por su estructura primaria, lo cual define su papel en un sistema biológico, muchas de ellas también desempeñan su función mediante procesos de reconocimiento molecular, es decir, mecanismos mediante los cuales las macromoléculas biológicas interactúan entre sí o con moléculas pequeñas, con una alta especificidad y afinidad para formar un complejo específico de forma no covalente, lo que constituye la base de todos los procesos bioquímicos en los organismos vivos. El reconocimiento molecular es un elemento de un mecanismo más complejo y funcionalmente importante que involucra los elementos esenciales de la vida como la autorreplicación, el metabolismo y el procesamiento de información por lo que una comprensión detallada de las interacciones proteína-ligando es fundamental para comprender la biología a nivel molecular[1].

Otro mecanismo mediante el cual se modula la actividad de las proteínas es a través de modificaciones postraduccionales (PTM por sus siglas en inglés) que consisten en la alteración de las cadenas polipeptídicas de las proteínas después de su síntesis, ya sea por escisión proteolítica o por la adición covalente de un grupo funcional en la cadena lateral de los aminoácidos, cuyo fin es modular la estructura, estabilidad, función, localización e interacciones de estas macromoléculas[2]. En la actualidad, se han reportado cientos de ejemplos de PTMs que generan o modulan la función de la proteína blanco. Entre las PTMs más estudiadas se encuentran la fosforilación, metilación, acetilación, glicosilación y ubiquitinación[3,4].

La fosforilación es una de las PTMs mejor caracterizadas, la cual juega un papel esencial en casi todas las vías de señalización y procesos biológicos, desde

eucariotas hasta procariotas. Este proceso dinámico reversible está modulado con precisión por quinasas (PK) y fosfatasas (PP), que participan en la incorporación o eliminación de un grupo fosfato en residuos específicos de sustratos proteínicos, por ejemplo en las cascadas de señalización mediadas por quinasas[3,5]. La fosforilación eucariótica ocurre principalmente en residuos específicos de serina (S), treonina (T) y tirosina (Y), mientras que la fosforilación de proteínas procariotas puede ocurrir en tipos adicionales de aminoácidos, como arginina (R), ácido aspártico (D), histidina (H) y cisteína (C)[5].

En esta tesis se abordó el estudio del reconocimiento molecular en interacciones proteína-ligando en la lipólisis y en la transcripción. En el primer capítulo se estudió a la proteína CGI-58, elemento clave en la regulación de la lipólisis del tejido adiposo y cuya función es regulada por la fosforilación desencadenada con la señalización de catecolaminas. En condiciones basales CGI-58 interactúa con la proteína PLIN1, lo que impide que reclute a la lipasa ATGL y se inicie la lipólisis; cuando existe una estimulación por catecolaminas, tanto CGI-58 como PLIN1 son fosforiladas, lo que da lugar al reclutamiento de la lipasa ATGL para iniciar con la hidrólisis de los ácidos grasos almacenados en los triglicéridos de las gotas lipídicas[6]. Las interacciones de CGI-58 tanto con PLIN1 como ATGL son puntos de regulación de la lipólisis, por lo que, en este trabajo, estudiamos los efectos de mutaciones puntuales en tres triptófanos del N-terminal de CGI-58 (W19, W23, W27), que están involucrados en la unión a la gota de lípido y en la activación de ATGL, los cuales sustituimos por residuos de alanina, generando la triple mutante 3WA. También realizamos la mutante en el sitio de fosforilación S237, donde sustituimos la serina por ácido glutámico para estudiar el efecto de la fosfomimetización en la proteína generando la cuádruple mutante 3WA/S237E. Nuestros resultados experimentales muestran que los triptófanos tienen un efecto en el estado agregación de la proteína y también determinamos que la mutante en el sitio de fosforilación tiende a incrementar la estabilidad térmica y contenido de estructura secundaria en comparación con la 3WA, lo que sugiere que la fosforilación tiene un rol en la integridad de la proteína.

En el segundo capítulo se estudió a la proteína SdrP, un factor de transcripción de la bacteria termófila *Thermus thermophilus*, que pertenece a la familia CRP/FNR y cuyos miembros regulan la expresión de genes blanco en forma de homodímero mediante el reconocimiento directo de regiones específicas del ADN. Este reconocimiento se da a través de su dominio de unión al ADN en el C-terminal y está controlado por el dominio sensor N-terminal en respuesta a varias señales ambientales/celulares[7]. En este trabajo determinamos la estructura tridimensional de SdrP en complejo con la secuencia consenso del ADN y determinamos los residuos de aminoácidos que realizan el reconocimiento directo al ADN, así como la importancia del papel de la Y180 en la unión al ADN, un residuo adyacente a la hélice de reconocimiento que no se ha reportado en otros factores de transcripción.

# **CAPÍTULO 1**

**Mejoramiento de la estabilidad de la proteína CGI-58 de humano inducida por fosfomimetización mediante la mutación S237E**

### **Objetivo general**

Evaluar el impacto de los triptófanos del N-terminal y el sitio de fosforilación S237 en las propiedades estructurales y fisicoquímicas de la proteína CGI-58 humana mediante mutaciones puntuales.

### **Objetivos específicos**

- Predecir las características bioquímicas de la proteína mediante enfoques *in silico*
- Realizar las construcciones que codifiquen para las proteínas recombinantes hCGI-58 WT, 3WA y 3WA-S237E.
- Estandarizar los protocolos de expresión y purificación de las proteínas recombinantes.
- Caracterizar el impacto en la estabilidad térmica, solubilidad y propiedades estructurales de las sustituciones por técnicas biofísicas.
- Determinar la funcionalidad del sitio de fosforilación S237.



## 1.1 Introducción

La proteína CGI-58 (por sus siglas en inglés de *Comparative Gene Identification-58*) fue identificada en un estudio donde se comparó el proteoma de *Caenorhabditis elegans* con una base de datos de Secuencias Expresadas con Etiquetas (EST por su siglas en inglés) de humano, lo que reveló que se expresa ampliamente en diversos tejidos, incluyendo la piel, linfocitos, hígado, músculo esquelético y cerebro[8]. Esta proteína pertenece a la familia de  $\alpha/\beta$  hidrolasas, la cual se caracteriza por contener una triada catalítica formada por un residuo nucleofílico, un ácido y una Histidina. Sin embargo, CGI-58 difiere de otros miembros de la familia porque contiene un residuo de Asparagina en lugar del residuo nucleofílico Serina, por lo que CGI-58 no tiene actividad hidrolítica. En humanos, la triada está formada por N153, H327 y D301 o E177[9]. [10]. La relevancia de CGI-58 en el metabolismo de lípidos se dio a conocer en 2001 cuando se encontró que mutaciones en el gen que codifica para la proteína (deleciones, inserciones y mutaciones puntuales en los residuos 7, 33, 130 y 260), causan una enfermedad de almacenamiento de lípidos neutros llamada Síndrome de Chanarin-Dorfman (CDS). Desde entonces, se han descrito más de 40 mutaciones a lo largo de la secuencia de la proteína relacionadas con el CDS [11]. Los pacientes con CDS se caracterizan por la presencia de ictiosis y la acumulación intracelular de gotas lipídicas (LD) en varios tejidos[9].

Actualmente, es ampliamente conocido que la proteína CGI-58 está involucrada en la regulación de diferentes rutas metabólicas y vías de señalización tejido específico a través de su interacción con otras proteínas[11]. La lipólisis en el tejido adiposo blanco es el proceso metabólico en el cual se ha reportado que CGI-58 tiene una actividad esencial para activar a la lipasa de triglicéridos del tejido adiposo (ATGL), enzima que inicia la hidrólisis de los triglicéridos (TAG). En ratones, CGI-58 incrementa la actividad de ATGL hasta 20 veces mediante la interacción directa proteína-proteína; sin embargo, la región de interacción es desconocida. Recientemente, se ha reportado que la región del C-terminal de CGI-58, específicamente los residuos R299 y G328, son esenciales para la activación de ATGL[12–14]. Además de incrementar su actividad, CGI-58 extiende la

regioselectividad de ATGL por el ácido graso en la posición sn-1 del TAG. En condiciones basales, ATGL hidroliza el ácido graso en la posición sn-2, generando el diacilglicerol sn-1-3, y en presencia de CGI-58 también genera el sn-2,3, lo cual permite aumentar la tasa de lipólisis e incrementa la liberación de ácidos grasos libres para la producción de energía[15].

La función de activación de ATGL por CGI-58 es regulada por su unión a la proteína de periferia Perilipina 1 (PLIN1) que restringe el acceso de las lipasas citosólicas a los TAG almacenados dentro de la LD[16], y su liberación es controlada por fosforilación. En condiciones basales, CGI-58 interacciona con el C-terminal de PLIN1 y permanece anclada a la LD del adipocito a través de los residuos de triptófano en su N-terminal (W21, W25 y W29, en la proteína de ratón). La región del N-terminal es una extensión del dominio estructural de hidrolasa de la proteína, la cual carece de estructura y ha sido propuesta como un ancla a la LD independiente[13,17]. Iniciada la estimulación por catecolaminas, la quinasa A dependiente de cAMP (PKA), fosforila seis residuos de serina en PLIN1 y la serina 239 del motivo KYSS en CGI-58 de ratón (residuo 237 en la proteína de humano), resultando en la liberación de CGI-58 de PLIN1 y permitiendo que CGI-58 reclute a ATGL a la LD[16,18,19]. Un estudio reciente reportó dos compuestos sintéticos que se unen a CGI-58 e interrumpen su interacción con PLIN1 sin pasar por la vía de señalización de PKA, además se encontró que el residuo Y330 está involucrado con la unión de la proteína al oleoil-CoA, logrando identificar la región de unión al ligando [20].

A pesar de la relevancia de los residuos de triptófano del extremo N-terminal y la fosforilación del residuo de serina en la regulación de la lipólisis, el papel funcional de estos residuos en CGI-58 no se conoce a detalle. Por ello, en este trabajo analizamos el impacto de la sustitución de los triptófanos del N-terminal por residuos de Alanina mediante la generación de una triple mutante (3WA), así como la introducción de la mutación fosfomimética S237E en la construcción 3WA (3WA/S237E) para evaluar su impacto en las propiedades estructurales y

fisicoquímicas de la proteína CGI-58 humana. Los resultados bioquímicos y biofísicos sugieren que los residuos de Triptófano en el extremo N-terminal funcionan como un sitio de nucleación de agregación. Para la mutante cuádruple 3WA/S237E, los resultados mostraron una mayor estabilidad térmica y solubilidad con un plegamiento adecuado y un sitio de unión funcional para oleoil-CoA. La contribución positiva de la mutación fosfomimética S237E podría explicarse por una posible interacción intramolecular entre el ácido glutámico y un residuo del dominio  $\alpha/\beta$  hidrolasa.

## **1.2 Materiales y Métodos**

### **1.2.1 Análisis bioinformático de CGI-58**

Las secuencias de la proteína de CGI-58 de humano (Q8WTS1.1) y ratón (Q9DBL9.1) fueron descargadas del sitio web NCBI. Se realizó un alineamiento de secuencias con el servidor T-Coffe[21]. La predicción de la estructura secundaria se realizó con los servidores web Jpred4 (<https://www.compbio.dundee.ac.uk/jpred/>)[22] y PredictProtein (<https://predictprotein.org/home>)[23]. El análisis de la propensión a agregación de las secuencias de proteínas se realizó con AGGRESCAN[24] usando los parámetros predeterminados.

### **1.2.2 Clonación y mutagénesis**

La secuencia de ADN codificante para la proteína de humano (NP\_001342115) fue optimizada para su expresión en bacterias y adquirida a GeneOracle en el plásmido pGOV4. El gen fue amplificado por PCR usando los oligos Fw 5'-GTAATTCCATATGGCAGCGGAAGAGGAAGAGG y Rv 5'-CGCGGATCCAAGCTT TCAATCGACAG. La PCR se realizó con la Taq polimerasa *Platinum* (Invitrogen). El producto de PCR fue purificado y digerido con las enzimas de restricción *NdeI* y *HindIII*, al igual que el vector de expresión pET28a modificado, el cuál tiene sustituido el sitio de corte de Trombina por el de la proteasa de rinovirus 3C. La triple mutante W19A/W23A/W27A (3WA) y la cuádruple mutante 3WA/S237E fueron

generadas por PCR usando los oligos de la Tabla 1. Todas las construcciones fueron verificadas por secuenciación por el método de Sanger.

### **1.2.3 Expresión y purificación de proteínas recombinantes**

Los plásmidos que expresan la secuencia completa de CGI-58 WT o las mutantes fueron transformados en las células competentes *Escherichia coli* BL21-CodonPLUS (DE3)-RIPL siguiendo las instrucciones del fabricante y cultivadas en placas de agar Luria-Bertani (LB) con kanamicina (50  $\mu\text{g}/\text{mL}$ ), estreptomycin (75  $\mu\text{g}/\text{mL}$ ) y cloranfenicol (34  $\mu\text{g}/\text{mL}$ ). A partir de las células transformadas se creció un preinóculo de 25 mL en medio LB líquido con los antibióticos y se incubó a 37 °C por 12 h a 200 rpm. El preinóculo se utilizó para inocular medio LB fresco con los antibióticos en una dilución 1:20 y se incubó a 37 °C a 200 rpm hasta una  $\text{DO}_{600}$  de 0.6-0.8, una vez alcanzada la densidad óptica, la expresión de las proteínas recombinantes fue inducida por la adición de isopropil  $\beta$ -D-1-tiogalactopiranosido (IPTG) a una concentración de 0.5 mM. El cultivo se incubó por 24 h a 18°C a 200 rpm. Las células fueron colectadas por centrifugación a 12,000 rpm por 15 min a 4 °C y el pellet se almacenó a -20 °C hasta su uso o inmediatamente sonicado (30 ciclos 15 ON/45 OFF a 50% amplitud) a 4 °C en amortiguador A [50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% (v/v) glicerol, 1% (w/v) octil glucósido, 6 mM ditioneitol (DTT), 15 mM 2-ME y 10 mM imidazol]. Se separó la fracción soluble del lisado celular por centrifugación a 12,000 rpm/15 min/4°C y se cargó a una columna de Ni-NTA (Qiagen) equilibrada con el amortiguador A. La columna se lavó con 3 volúmenes de columna (CV) de amortiguador B [50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% (v/v) glicerol, 0.1% (w/v) octil glucósido, 6 mM DTT, 15 mM 2-ME y 10 mM imidazol]. La elución de las proteínas recombinantes unidas a la columna se realizó en un solo paso con la adición de 3CV de amortiguador B suplementado con 500 mM de imidazol. Las fracciones que contenían la proteína recombinante fueron tratadas con 5 unidades/mL de la proteasa HRV 3C a 4°C por 12 h para remover la etiqueta de histidinas. La remoción de la etiqueta de histidinas fue monitoreada por SDS-PAGE y Western-Blot usando el anticuerpo primario dirigido a la etiqueta de Histidinas (6His-tag).

**Tabla 1. Oligonucleótidos usados para las construcciones WT, S237E, 3WA y 3WA/S237E**

Nombre de los oligos	Secuencia
WT_Forward	5' GTAATTCCATATGGCAGCGGAAGAGGAAGAGG
WT_Reverse	5' CGCGGATCCAAGCTTCAATCGACAG
3WA_Forward	5' GGAATTCCATATGGCAGCGGAAGAGGAAGAGGTTGACTC T GCTGACACCGGTGAGAGAAGTGGAGCGCTTACAGGGGCA CTTCCAACGCGGTGTCCCACTTC
S237E_Forward	5' CGATTTTAAACGTAAGTACGAATCTATGTTTCGAGGATGATA
S237E_Reverse	5' TATCATCCTCGAACATAGATTCTGACTTACGTTTAAAATCG

En rojo se señalan los sitios de restricción para la clonación y en verde los codones que generan las mutaciones puntuales.

La proteína libre de la etiqueta 6xHis se cambió al amortiguador C [50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10% (v/v) glicerol, 0.1% (w/v) octil glucósido, 6 mM DTT y 15 mM 2-ME] usando una columna PD-10 y se cargó a la columna HiTrap-Q HP (GE Healthcare). La proteína recombinante se eluyó con un gradiente lineal de 100 a 500 mM de NaCl en amortiguador C (15 CV). Las fracciones con la proteína eluída se colectaron y concentraron usando un dispositivo de ultrafiltración (Sartorius Vivaspin Turbo 15, 10 kDa) y posteriormente se purificó por cromatografía de exclusión molecular utilizando una columna Superdex-200 10/300 GL (GE Healthcare) en el amortiguador optimizado [50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% (v/v) glicerol, 0.1% (w/v) octil glucósido, 6 mM DTT y 15 mM 2-ME]. La concentración óptima de glicerol, octil glucósido, DTT y B-ME usado en el amortiguador optimizado fue determinada con la proteína 3WA por el ensayo de estabilidad térmica (Tabla 2), el amortiguador optimizado fue utilizado para la purificación de las proteínas WT y 3WA/S237E. La columna Superdex 200 10/300 GL fue calibrada con un marcador de proteínas que incluyen: tiroglobulina (670 kDa),  $\gamma$ -globulina (158 kDa), ovoalbúmina (44 kDa), mioglobina (17 kDa) y Vitamina B12 (1.3 kDa). El rendimiento de proteína obtenido al final de la purificación fue de 0.36 mg/L, 1.6 mg/L y 4.8 mg/L para la proteína WT, 3WA y 3WA/S237E respectivamente. Se monitoreó la expresión de proteínas y los procedimientos de purificación por SDS-PAGE. La concentración de proteína se determinó por el método de Bradford usando albúmina de suero bovino como proteína estándar.

#### **1.2.4 Mediciones de solubilidad de proteínas**

Los límites de solubilidad relativa de las proteínas 3WA y 3WA/S237E se midieron por ultrafiltración. Para ello, se prepararon muestras de proteína en el amortiguador optimizado y se concentraron hasta que la concentración de proteína en la solución retenida se mantuvo constante; la solución retenida se repuso según fue necesario. Se usó un concentrador Sartorius Vivaspin Turbo 15 de corte de 10 kDa a 3000 g a 4 °C. Las muestras se mezclaron regularmente mediante pipeteo para evitar el bloqueo de la membrana. La concentración de proteína se midió tomando alícuotas

**Tabla 2. Temperatura de desnaturalización media aparente ( $T_m$ ) determinada por el ensayo de estabilidad térmica para la mutante CGI-58 3WA.**

Condición		$T_m$ , °C	a	$E^2$
Amortiguador no optimizado		47.22	2.93	0.7605
Tris pH 8.0 <sup>+</sup>	50 mM	50.79	2.52	$4.60 \times 10^{-4}$
	100 mM	51.12	2.50	$2.30 \times 10^{-4}$
	150 mM	50.52	2.36	$1.24 \times 10^{-4}$
NaCl <sup>+</sup>	100 mM	50.45	2.06	$6.15 \times 10^{-4}$
	200 mM	50.23	2.04	$3.92 \times 10^{-4}$
	300 mM	49.59	2.50	$2.59 \times 10^{-4}$
	400 mM	48.82	2.44	$3.63 \times 10^{-4}$
	500 mM	48.24	2.35	$5.21 \times 10^{-4}$
2-ME	5 mM	50.04	2.13	$8.80 \times 10^{-4}$
	10 mM	50.57	2.06	$1.24 \times 10^{-4}$
	15 mM*	49.92	1.92	$6.51 \times 10^{-4}$
	20 mM	50.11	2.01	$3.10 \times 10^{-4}$
DTT	1 mM	50.55	2.23	$4.14 \times 10^{-4}$
	2 mM	50.38	1.72	$1.36 \times 10^{-3}$
	4 mM	50.08	1.70	$3.03 \times 10^{-4}$
	6 mM*	49.89	1.54	$1.57 \times 10^{-4}$
OG	0.05%	50.55	2.16	$2.53 \times 10^{-4}$
	0.1%*	50.29	2.12	$3.98 \times 10^{-4}$
	0.25%	49.07	2.35	$2.45 \times 10^{-3}$
	0.5%	48.47	4.18	$4.47 \times 10^{-3}$
Glicerol	5%	51.41	2.00	$6.15 \times 10^{-5}$
	10%*	51.33	1.90	$8.15 \times 10^{-4}$
	15%	51.08	2.39	$4.86 \times 10^{-4}$

“a” pendiente de la curva

“ $E^2$ ” error cuadrático

\* condiciones seleccionadas para amortiguador optimizado

+ Como no se observó un cambio significativo en la pendiente de la curva, se puede utilizar cualquier concentración indistintamente.

de 10  $\mu\text{L}$  en varios intervalos. Cuando se alcanzó la concentración máxima en la solución retenida, las muestras se transfirieron a un tubo de centrifuga de 1.5 mL y luego se incubaron durante la noche a 4 °C, enseguida fueron centrifugadas durante 15 minutos a 15,000 rpm a 4 °C para descartar cualquier proteína precipitada antes de medir la concentración de proteína.

### **1.2.5 *Dispersión dinámica de luz (DLS)***

Los ensayos DLS se llevaron a cabo utilizando un sistema APS2000 (Malvern Instruments). Las mediciones de los radios hidrodinámicos ( $R_h$ ) se tomaron a una concentración de proteína de 0.4 mg/mL en el amortiguador optimizado [50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10% (v/v) glicerol, 0.1% (w/v) octil glucósido, 6 mM DTT y 15 mM 2-ME]. El equipo se programó para realizar 15 lecturas de 40 s cada una a 25 °C; las muestras se incubaron durante 5 minutos a 25 °C para equilibrarlas antes de la medición. El peso molecular se estimó a partir del  $R_h$  usando el software Malvern Zetasizer versión 7.13.

### **1.2.6 *Ensayo de estabilidad térmica***

La determinación de la estabilidad térmica de las proteínas se llevó a cabo en placas de 96 pocillos utilizando el sistema 7500 Fast Real-Time PCR (Applied Biosystems). Las muestras que contenían la condición a evaluar, 5X SYPRO ORANGE (Invitrogen) y 0.05 mg/mL de proteína, se prepararon en un volumen final de 40  $\mu\text{L}$ . Los experimentos preliminares de desnaturalización térmica se realizaron en un rango de temperatura de 25-95 °C con incrementos de 1 °C y 30 s de incubación en cada paso. Una vez identificado el rango de temperatura para la curva de transición térmica, el intervalo de temperatura se delimitó como se indica en la figura correspondiente; todos los ensayos se realizaron por triplicado. Se siguió el despliegue de la proteína mediante la medición de la fluorescencia, se utilizó una longitud de onda de excitación de 455-485 nm y se registró la emisión entre 567-596 nm. La intensidad de fluorescencia de las muestras de proteína se corrigió restando la intensidad de fluorescencia de las muestras de control sin proteína. La curva de desnaturalización se obtuvo graficando las intensidades de fluorescencia



corregidas en función de la temperatura. Los datos se ajustaron a la ecuación de Boltzmann para la determinación de la temperatura de desnaturalización media aparente ( $T_m$ ) según lo recomendado por Po-Hsien et al [25].

### 1.2.7 Ensayos de extinción de fluorescencia

Los datos de fluorescencia intrínseca se adquirieron en un espectrofluorómetro LS 55 (Perkin Elmer). Los espectros de emisión de fluorescencia se midieron en el intervalo de 300-400 nm y la excitación de las muestras se realizó a 280 nm. Las muestras se prepararon a una concentración de 0.5  $\mu$ M de proteína en el amortiguador optimizado [Tris-HCl 50 mM, pH 8,0, NaCl 300 mM, glicerol al 10 % (v/v), octil glucósido al 0,1 % (p/v), DTT 6 mM y 15 mM 2-ME]. Los datos se obtuvieron a 25 °C con o sin concentraciones crecientes de oleoil-CoA hasta que la extinción alcanzó la meseta. Se midieron los espectros de fluorescencia del amortiguador y se restaron de los espectros de las muestras de proteína. La presencia de cualquier efecto de filtro interno se evaluó registrando la absorción del ligando en la longitud de onda de excitación y emisión a las mismas concentraciones utilizadas para los experimentos de fluorescencia. Cada espectro de fluorescencia del oleoil-CoA se corrigió utilizando la ecuación 1, donde  $F_{cor}$  y  $F_{obs}$  son las intensidades de fluorescencia corregidas y observadas, respectivamente.  $A_{ex}$  y  $A_{em}$  representan las diferencias en los valores de absorbancia de la muestra tras la adición del ligando en la excitación (280 nm) y longitudes de onda de emisión (300-400 nm), respectivamente[26].

$$F_{cor} = F_{obs} \times 10^{(A_{ex}+A_{em})/2} \quad (1)$$

Para estimar la constante de disociación aparente ( $K_d$ ), los cambios en la intensidad de la fluorescencia en la longitud de onda de emisión máxima se trazaron en función de la concentración del ligando y se analizaron mediante ajuste no lineal a la ecuación cuadrática 2 utilizando GraphPad PRISM 8.

$$Fraction\ bound = \frac{[P]_t + [L]_t + K_d - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t}}{2[P]_t} \quad (2)$$

Donde  $[P]_t$  es la concentración de proteína total, y  $[L]_t$  es el total de la concentración de ligando. Esta aproximación puede ser aplicada cuando  $[P]_t \ll K_d$  o cuando  $[P]_t$  es similar a o en un modesto exceso sobre la  $K_d$  (~exceso de 10 veces)[27]. Todos los valores reportados son el promedio de tres experimentos independientes.

### 1.2.8 Análisis de Dicroísmo Circular (CD)

Los espectros se adquirieron en un espectropolarímetro CD J-815 (Jasco Inc., Easton, MD) equipado con un dispositivo térmico Peltier (PFD-425S) para el control de la temperatura. Debido a la alta absorbancia del amortiguador optimizado a longitudes de onda bajas en la región UV, las muestras se prepararon a una concentración de proteína de 3.7  $\mu\text{M}$  en el siguiente amortiguador [20 mM fosfato de potasio pH 8,0, 20 mM NaCl, 5 % (v/v) glicerol, 0.1% (p/v) de octil glucósido]. Los datos de CD de UV lejano se registraron de 192.5 a 250 nm usando una cubeta de cuarzo de 0.1 cm de paso óptico a 25, 65 y 80 °C. Los espectros de CD de UV cercano se registraron de 250 a 320 nm utilizando una cubeta de 1.0 cm de paso óptico a 25 y 80 °C. La elipticidad se reportó como elipticidad media por residuo  $[\theta]$ [28]. El contenido de la estructura secundaria se estimó a partir de los datos de CD utilizando el servidor web BestSel (<http://bestsel.elte.hu/>)[29].

Los experimentos de desnaturalización térmica se realizaron aumentando la temperatura de las muestras de 25 a 80 °C a una velocidad de calentamiento constante de 2 °C·min<sup>-1</sup> mientras se registraban los cambios en la elipticidad a 222 nm. La concentración de proteína y las condiciones del amortiguador fueron las mismas que las descritas anteriormente. Los datos experimentales de CD fueron analizados asumiendo un modelo de transición de dos estados entre el nativo (N) y el monómero no plegado (D) sin un estado intermediario[30].



La constante de equilibrio de la reacción se definió como:

$$K_D = \frac{[D]}{[N]} \quad (4)$$

La fracción de la proteína no plegada, acorde a la ecuación 5,

$$f_D = \frac{[D]}{P_t} \quad (5)$$

donde  $P_t$  es la concentración total de proteína en unidades de monómero.

La determinación del cambio de entalpía ( $\Delta H_D$ ) asociado con la desnaturalización térmica se realizó usando la siguiente relación termodinámica clásica[31]:

$$\Delta H_D(T) = -R[\partial \ln K_D / \partial (1/T)] = RT^2(\partial \ln K_D / \partial T) \quad (6)$$

$$\Delta S_D(T_m) = \Delta H_D(T_m) / T_m \quad (7)$$

donde  $T_m$  es el punto medio de desnaturalización térmica. Los diagramas de van't Hoff ( $\ln K_D$  vs  $1/T$ ) de desnaturalización térmica fueron aproximadamente lineales a lo largo de la región  $T_m$ , lo que permite una estimación de la entalpía y la entropía del despliegue en  $T_m$ .

### **1.2.9 Modelado estructural**

El modelo tridimensional (3D) de la secuencia completa de CGI-58 de humano se generó utilizando el servidor web en línea I-TASSER[32]. Se realizó una minimización energética del modelo 3D generado por I-TASSER mediante un protocolo de 100 pasos en UCSF Chimera[33]. La estructura minimizada de la proteína CGI-58 WT se usó como plantilla para generar la mutante S237E. Para ello, se sustituyó la cadena lateral de S237 por ácido glutámico con el software COOT[34]. El modelo se sometió a otra ronda de minimización de energía para mejorar los rotámeros de las cadenas laterales, los enlaces de hidrógeno y eliminar los choques entre átomos. Los potenciales electrostáticos se calcularon utilizando el servidor web PDB2PQR y el Adaptive Poisson-Boltzmann Solver APBS[35]; el cálculo del solvente implícito se realizó con el campo de fuerza PARSE mientras que los estados de protonación se asignaron con PROPKA. La visualización y representación molecular se realizaron con UCSF CHIMERA[33].

## 1.3 Resultados y Discusión

### 1.3.1 CGI-58 contiene un sitio (*hot-spot*) de agregación en la región N-terminal

La proteína CGI-58 de humano consiste en una secuencia de 349 residuos de aminoácidos con una masa molecular teórica de 39.1 kDa. En la Figura 1A se muestran los motivos conservados de la proteína, como el motivo de unión a LD en el extremo N-terminal y el dominio estructural perteneciente a la familia de  $\alpha/\beta$  hidrolasas. En el alineamiento de la secuencia, se observa que la proteína de humano comparte un 94% de identidad con el ortólogo en ratón. El dominio estructural formado por los residuos 52 a 349 (mostrado en azul en la Figura 1A) consiste en el plegamiento  $\alpha/\beta$  hidrolasa con 8 laminas  $\beta$  en el núcleo rodeado de 2 y 4 hélices  $\alpha$  en cada lado[10]. La región de los residuos 178 a 276 (color cian en la Figura 1A) se considera una inserción al núcleo  $\alpha/\beta$  hidrolasa, la cual muestra una propensión para formar estructuras tipo hélices  $\alpha$  ( $\alpha_6$ - $\alpha_{11}$ ), y ha sido propuesta como una tapa cuya función es proteger el sitio catalítico del acceso directo a la superficie[10]. El motivo de unión a la LD es una extensión del dominio estructural, el cual está localizado en el extremo N-terminal y comprende los residuos 10 al 43. Este motivo se caracteriza por la presencia de residuos de triptófano en la posición 19, 23 y 27, los cuales actúan como un ancla para unirse a la gota lipídica y son necesarios para la activación de ATGL; posterior a estos tres residuos se encuentra una región flexible (residuos 30 al 38) que permite la orientación sin restricciones del dominio  $\alpha/\beta$  hidrolasa[13,17].

Se analizó la secuencia de aminoácidos de la proteína CGI-58 de humano en el servidor AGGREGSCAN y se identificaron trece posibles *hot-spots* de agregación (Figura 1B). Estos sitios han sido descritos como secuencias cortas y específicas de aminoácidos que funcionan como puntos de nucleación en los procesos de agregación[24]. Doce de los posibles trece *hot-spots* identificados en la secuencia se encuentran distribuidos a lo largo del dominio estructural de CGI-58 entre los residuos 52 a 349 y, por lo tanto, están ocultos en el núcleo de la proteína, participando en la red de contacto que estabilizan a la misma. Este tipo de



muestra en el recuadro rojo; el alineamiento se realizó con la herramienta T-Coffee. Debajo del alineamiento, se muestra la predicción de estructura secundaria de la proteína de humano con el servidor web Jpred4 y PredicProtein; las hélices alfa se representan por rectángulos amarillos, mientras que las láminas beta con flechas azules. Por último, se muestran los dominios estructurales de CGI-58: en naranja se muestra la región de unión a la gota lipídica (residuos 8-41), el núcleo de  $\alpha/\beta$  hidrolasa se representa en azul (residuos 52-349) y el sitio activo de la tapa (CAP) en color cian (residuos 178-276). **(B)** Se muestra el análisis de predicción de agregación, el gráfico del área de *hot-spot* (HSA) para la secuencia de aminoácidos CGI-58. El *hot-spot* marcado con \* se pierde en el mutante 3WA. **(C)** Se muestran los valores normalizados de propensión a la agregación (Na4vSS) estimados para las secuencias de aminoácidos WT y 3WA. Se espera una baja propensión a la agregación con valores negativos más bajos, mientras que los valores positivos sugieren una mayor propensión a la agregación. El análisis de datos se realizó utilizando el servidor web AGGRESCAN.

distribución se ha propuesto como un mecanismo protector contra la agregación en proteínas globulares[24]. De forma interesante, el primer *hot-spot* de la secuencia, que comprende los residuos 19 a 25, cumple con las características de un verdadero *hot-spot*, ya que está expuesto al solvente y está ubicado en una región no estructurada; por lo tanto, podría funcionar como un sitio de nucleación de agregación. La composición de aminoácidos de esta región muestra residuos con una alta propensión a formar agregados, como los residuos de Leucina en las posiciones 20 y 24 y los tres residuos de Triptófano en las posiciones 19, 23 y 27[36]. Considerando esta información al igual que el papel (o la función) de los tres residuos de Triptófano en la unión a la gota lipídica[13,17], decidimos generar la triple mutante W19A/W23A/W27A (3WA) para analizar el papel de esta región en la estabilidad de CGI-58. Cabe mencionar que el análisis de secuencia del triple mutante 3WA en el servidor AGGRESCAN muestra la eliminación del “*hot-spot*” en los residuos 19 a 25. Además, el valor promedio de propensión a la agregación (Na4vSS) calculado por AGGRESCAN fue -1.0 para CGI- 58 proteína WT y -1.9 para la mutante 3WA (Figura 1C), lo que sugiere una mayor solubilidad para la proteína mutante[37].

### **1.3.2 La triple mutante 3WA y la sustitución fosfomimética de S237 promueve el estado monomérico y mejora la solubilidad de CGI-58**

La proteína CGI-58 WT y la mutante 3WA se purificaron por cromatografía de afinidad a Níquel, intercambio iónico y exclusión molecular al igual que la cuádruple mutante 3WA/S237E, en la cual se sustituyó la serina por ácido glutámico para mimetizar la fosforilación por PKA. Decidimos realizar la mutación S237E ya que en un trabajo previo se mostró que la sustitución fosfomimética de la S239 de CGI-58 de ratón por ácido glutámico, mejora la solubilidad de la proteína[19,38]. Los ensayos iniciales de purificación de la proteína CGI-58 WT resultaron en rendimientos muy bajos de proteína, los cuales no se pudieron mejorar al evaluar diferentes condiciones tales como cepas, tiempo y temperatura de expresión. Para evitar esta situación, decidimos purificar la proteína 3WA y analizar la estabilidad térmica de la proteína en presencia de varios aditivos mediante un ensayo de

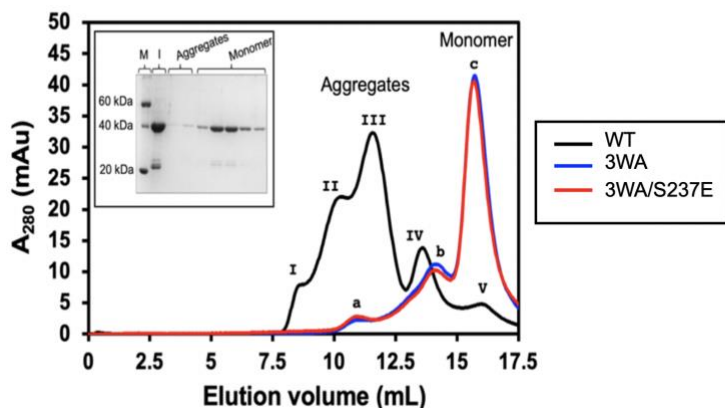
desplazamiento térmico basado en fluorescencia[39]. En la Tabla 2, se muestran las temperaturas de desnaturalización ( $T_m$ ) con los diferentes aditivos. En el amortiguador no optimizado la  $T_m$  de 3WA fue de 47.22 °C con una pendiente ( $a$ ) de 2.93, lo que nos indica una sigmoidea no pronunciada (Tabla 2), por lo tanto, la transición del estado nativo al desnaturalizado no es de manera abrupta y posiblemente estén presentes diferentes especies de plegamiento u oligomerización. El objetivo de analizar diferentes aditivos es encontrar las condiciones que tengan un efecto en la sigmoidea, ya sea en la disminución del valor de la pendiente, es decir sea más pronunciada o un desplazamiento de la sigmoidea al aumentar de la temperatura de desnaturalización. Este efecto se observó en las condiciones de glicerol al 10%, 2-ME a 15 mM, DTT a 6 mM y octil-glucósido al 0.1%. Estas condiciones se utilizaron en la composición del amortiguador optimizado, el cual se utilizó para purificar las proteínas WT, 3WA y 3WA/S237E.

Los datos experimentales de la purificación por cromatografía de exclusión molecular (*size-exclusion chromatography*, SEC) se muestran en la Figura 2A; las tres proteínas fueron purificadas en las mismas condiciones. El perfil de elución de la proteína WT (Figura 2A, línea color negro) resultó en una población heterogénea que muestra que la proteína eluye principalmente como agregados de diferente peso molecular (107.4, 314.8, y 644.7 kDa) con una pequeña fracción del monómero (31.0 kDa). En contraste, la proteína 3WA mostró un perfil menos heterogéneo con dos picos principales, de los cuales el pico del monómero mostró absorbancia ~5 veces mayor que el pico de los agregados (Figura 2A, línea color azul).

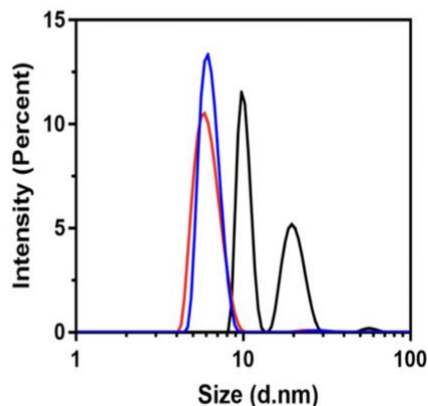
De manera similar, la mutante 3WA/S237E mostró el mismo perfil observado para 3WA con una absorbancia casi idéntica, ya que la proteína total cargada en la columna fue la misma que la de 3WA y CGI-58 WT. En el recuadro de la Figura 2A se muestra el perfil electroforético de las fracciones purificadas de 3WA/S237E en condiciones desnaturalizantes. Como se puede observar, la proteína 3WA/S237E



## A. Perfil de SEC



## B. DLS



**Figura 2. Cromatogramas de Exclusión Molecular y análisis por Dispersión Dinámica de Luz de CGI-58 WT, 3WA y 3WA/S237E.**

**(A)** Perfil de elución SEC de las proteínas CGI-58 WT, 3WA y 3WA/S237E separadas en una columna Superdex 200; la purificación se realizó en condiciones idénticas. Se indican los picos de elución para el monómero y los agregados. I: ND, II: 644.7, III: 314.8, IV: 107.4, V: 31.1, a: 441.1, b: 82.5 y c: 37.6 kDa. Línea negra: WT; línea azul: 3WA; línea roja: 3WA/S237E. El recuadro muestra el análisis SDS-PAGE de las fracciones purificadas de 3WA/S237E (línea roja); el gel se visualizó mediante tinción con azul de Coomassie. M: marcador de peso molecular; I: muestra de proteína cargada en la columna; Agregados: pico a y b; Monómero: fracciones de elución de 15 a 17.5 mL. **(B)** Las fracciones de monómero que se muestran en el panel (A) se recolectaron y analizaron mediante dispersión dinámica de luz. La proteína WT (línea negra) mostró dos picos, con un tamaño de partícula de  $10.2 \pm 0.4$  y  $19.5 \pm 2.5$  nm de diámetro. El tamaño de partícula estimado de 3WA (línea azul) y 3WA/S237E (línea roja) fue de  $6.6 \pm 0.8$  y  $6.4 \pm 1.0$  nm de diámetro, respectivamente, que se aproxima al peso molecular del monómero ( $55.7 \pm 15.8$  kDa y  $55.8 \pm 14$  kDa, respectivamente); ambas proteínas muestran un único pico monodisperso. Se muestran experimentos representativos en (A) y (B).

eluye principalmente como monómero; se observaron resultados similares para la proteína 3WA.

Después de la purificación por SEC, las fracciones de monómero de CGI-58 WT, 3WA y 3WA/S237E se analizaron mediante dispersión dinámica de luz (DLS) para obtener información sobre sus propiedades hidrodinámicas (Figura 2B). Interesantemente, la fracción del monómero de la proteína CGI-58 WT mostró una alta tendencia a la agregación, ya que el análisis DLS mostró dos picos con un diámetro de tamaño de partícula de  $10.2 \pm 0.4$  y  $19.5 \pm 2.5$  nm, lo que corresponde a un peso molecular estimado de  $156.5 \pm 9.5$  y  $697.2 \pm 102.9$  kDa, respectivamente. Por el contrario, la mutante 3WA mostró un único pico monodisperso con un tamaño de partícula de  $6.6 \pm 0.8$  nm de diámetro ( $55.7 \pm 15.8$  kDa) que se aproxima al peso molecular del monómero. De manera similar, la mutante 3WA/S237E también mostró un único pico monodisperso con el peso molecular del monómero (diámetro de  $6.4 \pm 1.0$  nm y  $55.8 \pm 14$  kDa). Estos resultados sugieren que los residuos de triptófano 19, 23 y 27 pueden estar involucrados en el proceso de agregación de la proteína CGI-58 WT, ya que la fuerte tendencia a la agregación observada para la proteína después de la purificación por SEC no se observó en la mutante 3WA ni en la 3WA/S237E. Además, los resultados experimentales concuerdan con el análisis AGGRESCAN, según el cual los residuos 19 a 25 se predicen como un punto crítico que confiere una mayor propensión a agregación a la proteína WT. La región N-terminal es esencial para la localización de CGI-58 en LD, la cual se propuso como un motivo funcional independiente ya que cuando la proteína fluorescente amarilla se fusionó con los residuos 19 a 35, la proteína de fusión se encontró localizada en las LD. Además, el análisis estructural de un péptido que contenía los residuos 10 a 43 mediante experimentos de Resonancia Magnética Nuclear y CD confirmó su conformación no estructurada[17]. Por lo tanto, el proceso de agregación en la proteína WT probablemente se inicia por interacciones intermoleculares entre la región N-terminal de monómeros adyacentes. Para la sustitución fosfomimética en la posición 237, el análisis SEC o DLS no nos permitió

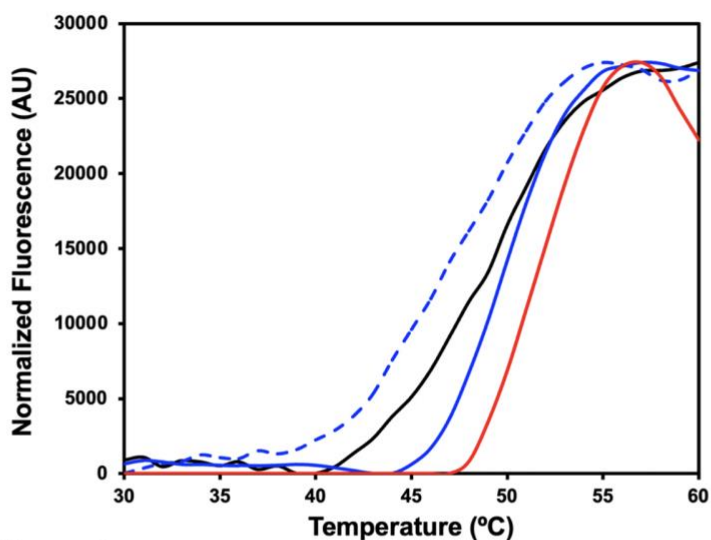
observar ninguna diferencia entre 3WA y 3WA/S237E, lo que sugiere una conformación similar para ambas proteínas.

Posteriormente evaluamos si la mutación fosfomimética de la serina 237 en la construcción 3WA/S237E tuvo algún impacto en la solubilidad de la proteína, como se mencionó anteriormente. Para ello, muestras de proteína de 3WA y 3WA/S237E en idénticas condiciones de amortiguador se concentraron por ultrafiltración hasta alcanzar los límites de solubilidad[40,41]. La solubilidad relativa alcanzada por la mutante 3WA/S237E (7.7 mg/mL) fue el doble del valor observado de 3WA (3.9 mg/mL); este valor también es el doble de la concentración reportada anteriormente para la mutante de ratón S239E[38]. Estos resultados sugieren que la mutación fosfomimética induce un cambio estructural que aumenta la solubilidad, muy probablemente una interacción electrostática con un residuo vecino ya que la Serina 237 se localiza en la tapa que protege al sitio activo putativo.

### ***1.3.3 La cuádruple mutante 3WA/S237E muestra una mayor estabilidad térmica en comparación con el mutante 3WA y está correctamente plegada***

Para evaluar el efecto de las mutaciones en la estabilidad de CGI-58, se analizó la desnaturalización térmica de las proteínas CGI-58 WT, 3WA y 3WA/S237E mediante un ensayo de desplazamiento térmico basado en fluorescencia (Figura 3A). Se determinó el punto medio aparente de desnaturalización térmica ( $T_m$ ) para las proteínas WT, 3WA y 3WA/S237E en el amortiguador optimizado (Tris-HCl 15 mM, pH 8.0, NaCl 85 mM, glicerol al 2.85 %, octil glucósido al 0.03 %, DTT 0.6 mM y 2-ME 1.5 mM), los valores observados fueron  $49.2 \pm 0.5$  °C,  $50.1 \pm 0.4$  °C y  $51.7 \pm 0.1$  °C, respectivamente (Figura 3B). La proteína WT muestra una curva de transición sigmoidea entre 40 y 55 °C que se ve extendida y menos pronunciada en comparación con la mutante 3WA. Las curvas WT y 3WA difieren en la temperatura inicial de la transición, pero se superponen en la región superior de 50 a 55 °C, indicando diferencias en los estados plegados y desplegados de las proteínas, lo que puede explicarse por la presencia de agregados en la proteína WT, como se describió anteriormente.

## A. Desnaturalización Térmica



## B. T<sub>m</sub> estimada

Proteína	T <sub>m</sub>	Condición
3WA	47.2 ± 0.5 °C	Amortiguador no optimizado
WT	49.2 ± 0.5 °C	Amortiguador optimizado
3WA	50.1 ± 0.4 °C	
3WA/S237E	51.7 ± 0.1 °C	

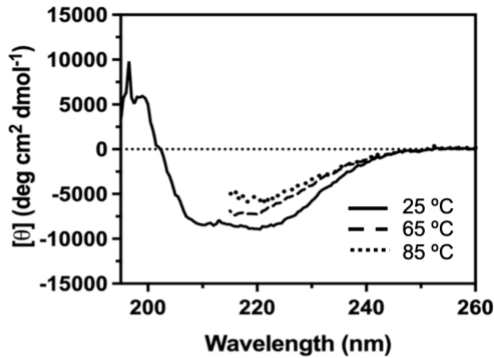
### Figura 3. Ensayo de estabilidad térmica de CGI-58 WT, 3WA, y 3WA/S237E.

Curvas de desnaturalización térmica de CGI-58 WT (línea negra), 3WA (línea azul) y 3WA/S237E (línea roja) en el amortiguador optimizado. También se muestra la curva de desnaturalización de 3WA (línea discontinua azul) en el amortiguador no optimizado. La T<sub>m</sub> estimada se calculó ajustando los datos a la ecuación de Boltzmann. Los datos mostrados corresponden a la media de tres mediciones independientes.

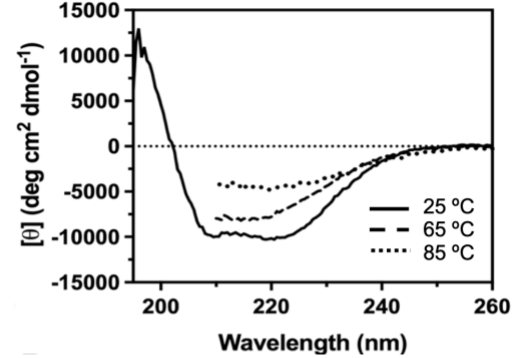
En contraste con el perfil observado para la proteína WT, la curva de desnaturalización térmica de las proteínas 3WA y 3WA/S237E es una sigmoide pronunciada que muestra una transición brusca al pasar del estado plegado al desplegado, lo que sugiere un plegamiento homogéneo y compacto de las proteínas[42]. Es interesante que la mutante cuádruple 3WA/S237E mostró una mayor  $T_m$ , lo que implica que la mutación fosfomimética induce un cambio estructural que mejora la estabilidad térmica. El cambio en la termoestabilidad de las proteínas de forma positiva, nula o negativa asociado a una modificación postraduccional como la fosforilación ha sido previamente reportado[43]. Por ejemplo, un efecto positivo de  $\sim 2$  °C fue reportado para la mutante fosfomimética (Cp183-EEE) del dominio C-terminal rico en arginina de la proteína central del virus de la hepatitis B23 [44], y un efecto positivo similar de  $\sim 2$  °C fue reportado para la mutante fosfomimética S17D en el dominio N-terminal de la proteína MDM2[45]. Cabe mencionar que el valor de  $T_m$  estimado para la proteína 3WA en el amortiguador no optimizado (Tris-HCl 50 mM, pH 8.0, NaCl 300 mM, glicerol al 10%, octil glucósido al 0.1 %, DTT 6 mM y 2-ME 15 mM) fue de 47.2 °C (Fig. 3B), muy similar a los valores determinados previamente para las proteínas WT y S237E por fluorimetría diferencial de barrido que fueron 47.4 °C y 46.2 °C, respectivamente[46]. Por lo tanto, el uso del amortiguador optimizado nos permitió observar una ganancia en la estabilidad térmica de  $\sim 3.0$  °C para el 3WA en comparación con las condiciones no optimizadas, así como detectar una diferencia de  $\sim 1.0$  °C entre la proteína WT y la mutante 3WA, y una mejora de  $\sim 1,5$  °C para la mutante 3WA/S237E respecto a la mutante 3WA.

Para obtener más información sobre el contenido de la estructura secundaria, la estabilidad térmica y la estructura terciaria de 3WA y 3WA/S237E, realizamos experimentos de espectroscopia de CD[28]. Debido a su alta tendencia a la agregación y poca solubilidad, CGI-58 WT no se sometió a este análisis. Las Figuras 4A y B muestran los espectros de CD en la región UV lejana de 3WA y 3WA/S237E a 25, 65 y 80 °C. A 25 °C, ambas proteínas presentan los espectros esperados para una proteína  $\alpha/\beta$  con dos picos negativos centrados en aproximadamente 210 nm

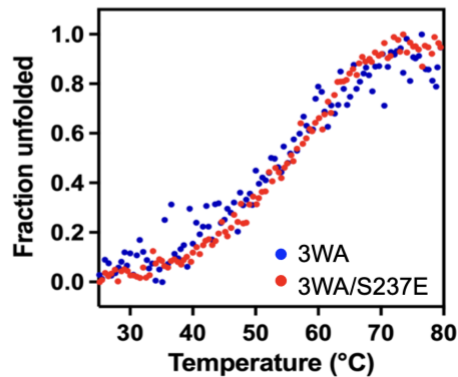
### A. CD UV lejano 3WA



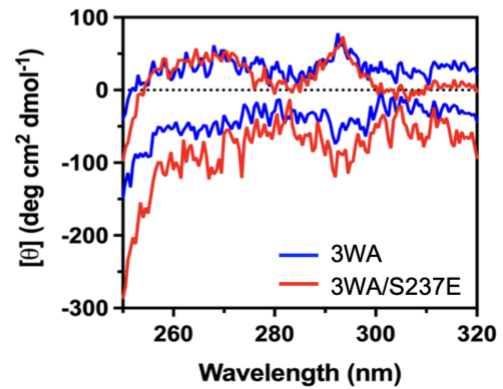
### B. CD UV lejano 3WA/S237E



### C. Desnaturalización térmica



### D. CD UV cercano



**Figura 4. Análisis de 3WA and 3WA/S237E por espectroscopía CD.**

Los espectros de CD de UV-lejano de **(A)** 3WA y **(B)** 3WA/S237E fueron determinados a 25 (línea negra), 65 (línea discontinua), y 80 °C (línea punteada) a una concentración de proteína de 3.7  $\mu$ M. **(C)** Los perfiles de desnaturalización térmica de 3WA (línea azul) y 3WA/S237E (línea roja) fueron monitoreados por espectroscopía de CD a 222 nm. **(D)** Espectro de CD de UV-cercano de 3WA (línea azul) y 3WA/S237E (línea roja) fueron obtenidos de 250 a 320 nm a 25 (líneas superiores) y 80 °C (líneas inferiores).

(-8466.0 y -9970.6 deg·cm<sup>2</sup>·dmol<sup>-1</sup>, 3WA y 3WA/S237E, respectivamente) y 222 nm (-8610.8 y -10051.9 deg·cm<sup>2</sup>·dmol<sup>-1</sup>, 3WA y 3WA/S237E, respectivamente).

Está claro que, aunque los espectros de CD de UV lejano para ambas proteínas son comparables, los valores de elipticidades registrados para 3WA fueron significativamente más bajos que los de 3WA/S237E. Esta diferencia podría atribuirse principalmente a la ligera agregación observada para 3WA cuando se incubaba en el amortiguador utilizado para las mediciones de CD. Un análisis de espectro con el servidor web BeStSel [29] dio como resultado un contenido estimado de ~24% de hélices  $\alpha$ , 29% de hebras  $\beta$ , 10% de giros  $\beta$  y 37% de *coil* para 3WA/S237E, que es cercano a la predicción de la estructura secundaria que se muestra en la Figura 1A. No se realizó un análisis de estructura secundaria de 3WA debido a su ligera tendencia a agregarse. A pesar de eso, el reconocimiento de plegamiento para ambas proteínas a partir de datos de CD con el servidor BeStSel predice la presencia del plegamiento tipo sándwich  $\alpha\beta\alpha$  observado en la familia de  $\alpha/\beta$  hidrolasas[47]. A temperaturas más elevadas (65 and 80 °C), ambas proteínas muestran una disminución en la elipticidad, lo que indica que las proteínas experimentan un despliegue inducido por la temperatura. Para analizar más a fondo el despliegue observado inducido por la temperatura, se realizaron experimentos de desnaturalización térmica para ambas proteínas siguiendo el cambio de elipticidad a 222 nm en función de la temperatura (Figura 4C). En ambos casos, las transiciones aparecen como curvas sigmoideas simples. Los datos se ajustaron a un modelo de dos estados de desplegamiento de proteínas, lo que nos permitió estimar el punto medio de la desnaturalización térmica y el cambio de entalpía aparente ( $\Delta H_D$ ). La estabilidad térmica de 3WA/S237E ( $T_m$  de  $57.5 \pm 1.0$  °C) fue 3.1 °C mayor que la de 3WA ( $T_m$  de  $54.4 \pm 1.1$  °C). De manera similar, la  $\Delta H_D$  de desplegamiento de 3WA/S237E ( $158.8 \pm 2.8$  kJ mol<sup>-1</sup>) fue mayor que la de 3WA ( $150.9 \pm 2.0$  kJ mol<sup>-1</sup>). La mayor estabilidad térmica observada para 3WA/S237E es consistente con los resultados de desplazamiento térmico monitoreado por fluorescencia descritos anteriormente y sugiere que la sustitución fosfomimética induce un cambio estructural, muy probablemente interacciones intramoleculares

favorables, que estabilizan la estructura de 3WA/S237E[48]. Además, el cambio de entalpía aparente asociado a los ensayos de desplegamiento indica una reacción más cooperativa para 3WA/S237E, lo que podría atribuirse a un incremento modesto, no despreciable, en el número o la robustez de los contactos terciarios.

Para examinar la estructura terciaria de ambas proteínas, se analizaron los espectros de CD en la región UV cercana (250 a 320 nm). Los espectros de CD de UV cercano proporcionan evidencia de estructura terciaria ya que cuando se ubican en ambientes quirales, las cadenas laterales de aminoácidos aromáticos tienen un perfil de CD característico que proporciona una huella digital de una proteína en particular[49].

Aunque se utilizó una concentración de proteína relativamente baja (0.15 mg/mL) para los experimentos de CD, dado que se observó una baja estabilidad de la proteína en el amortiguador de trabajo (Figura 4D), se observó el pico esperado cercano a 290 nm para los residuos de Triptófano junto con una amplia región positiva entre 260 y 280 nm, que podría corresponder a la señal de residuos de tirosina y fenilalanina. Ambos espectros se superponen de 250 a 300 nm, lo que sugiere que la estructura terciaria de 3WA/S237E es similar a la de la mutante 3WA. En ambos casos, el aumento de la temperatura provocó la pérdida de señal a 290 nm y de 260-280 nm, lo que significa que el entorno alrededor de los residuos aromáticos cambió como consecuencia de la desnaturalización térmica.

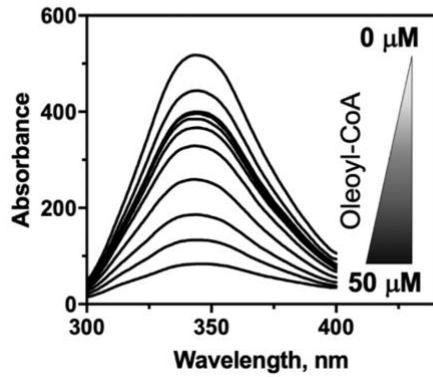
En conjunto, los datos de CD de UV lejano indican que 3WA y 3WA/S237E mantienen la estructura secundaria característica de la familia de  $\alpha/\beta$  hidrolasas. Los experimentos de desnaturalización térmica confirmaron la conformación más estable de 3WA/S237E, y el  $\Delta H_D$  aparente sugiere que 3WA/S237E está estabilizado por la mejora de las interacciones intramoleculares debido a la presencia del ácido glutámico fosfomimético en la posición 237. Además, se puede inferir una estructura terciaria similar entre ambas proteínas a partir de los resultados de CD de UV cercano.



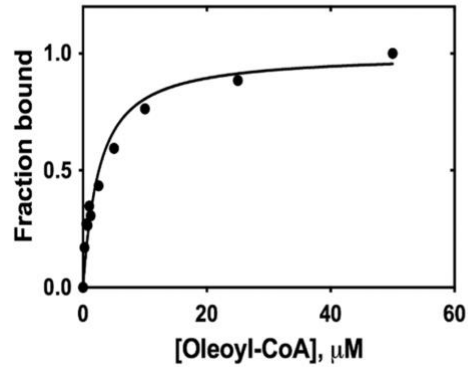
#### **1.3.4 La cuádruple mutante 3WA/S237E tiene un sitio de unión funcional para oleoil-CoA**

Estudios previos han encontrado que el oleoil-CoA es un ligando endógeno de CGI-58 de humano que comparte un sitio de unión común con compuestos sintéticos[20]. También se ha reportado la unión de oleoil-CoA a CGI-58 de ratón[50]. Considerando esta información, nos interesamos en evaluar si la proteína 3WA/S237E tiene un sitio de unión funcional para oleoil-CoA. Para determinar la afinidad de unión hacia oleoil-CoA, realizamos experimentos de espectroscopia de fluorescencia de triptófano. Este enfoque se ha utilizado con frecuencia para estudiar la unión de ligandos a proteínas [51], y el análisis de los datos de extinción de la fluorescencia se ha revisado recientemente[26]. La Figura 5A muestra un ejemplo representativo de los espectros de emisión de fluorescencia de 3WA/S237E en presencia de concentraciones crecientes de oleoil-CoA. Los datos de fluorescencia observados para oleoil-CoA se corrigieron por el efecto del filtro interno, ya que las muestras del ligando mostraron una absorbancia significativa a la longitud de onda de excitación de 280 nm, con un valor de 0.2 a la concentración más alta de 50  $\mu$ M. La adición de concentraciones crecientes de oleoil-CoA a 3WA/S237E resultó en una reducción progresiva de la intensidad de la fluorescencia, probablemente debido a cambios estructurales inducidos alrededor de los residuos aromáticos, lo que sugiere su interacción con la proteína. En la concentración más alta de oleoil-CoA, observamos un 65% de extinción de la fluorescencia. Los cambios en la intensidad de la fluorescencia se representaron frente a la concentración del ligando y se analizaron para determinar la constante de disociación aparente ( $K_d$ ) por el ligando. La proteína 3WA/S237E mostró una  $K_d$  aparente por el oleoil-CoA de  $2.3 \pm 0.3 \mu$ M (Figura 5B). Este resultado sugiere que el mutante 3WA/S237E tiene un sitio de unión al ligando funcional y probablemente intacto, ya que nuestros resultados son similares a los valores reportados para la proteína CGI-58 de ratón, que muestra una  $K_d$  aparente de 1.1  $\mu$ M para oleoil-CoA[50]. En comparación con el ortólogo de ratón, la proteína 3WA/S237E mostró una afinidad dos veces menor por el oleoil-CoA.

### A. Fluorescencia intrínseca



### B. Estimación de $K_d$



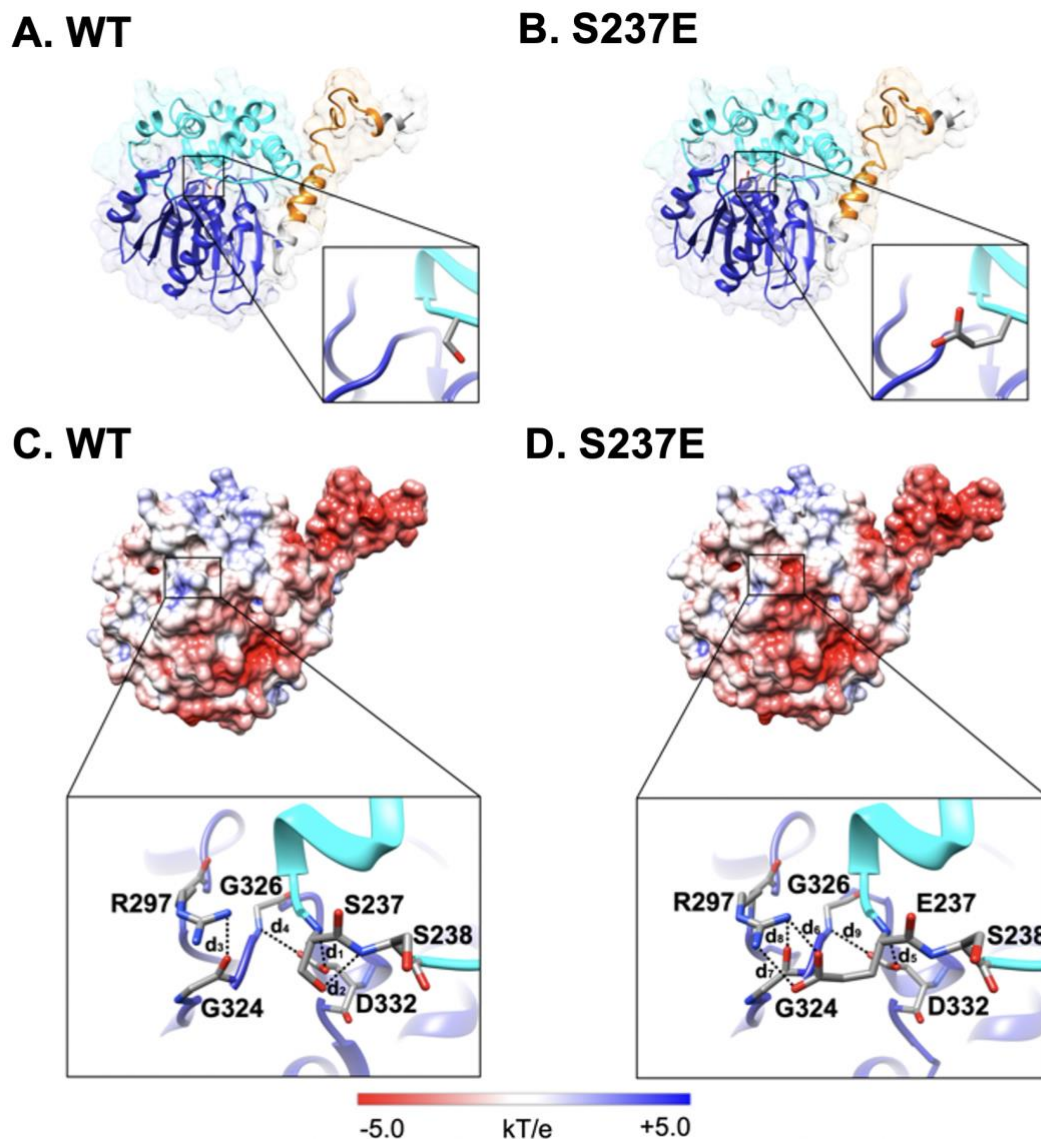
**Figura 5. Ensayos de afinidad de 3WA/S237E a oleoil-CoA monitoreados por fluorescencia intrínseca.**

**(A)** Espectros de fluorescencia de 3WA/S237E con concentraciones crecientes de oleoil-CoA. Los datos de fluorescencia se corrigieron por el efecto del filtro interno. Los ensayos se realizaron con una concentración de proteína de  $0.5 \mu\text{M}$  a  $25^\circ\text{C}$ ; se muestra un experimento representativo. **(B)** Los cambios de intensidad de fluorescencia en el punto de emisión máxima observada en (A) se representaron en función de la concentración de oleoil-CoA y se analizaron para determinar la  $K_d$  aparente. Se muestra la media de tres mediciones independientes.

### **1.3.5 Características del papel estructural del residuo S237**

Para obtener más información y ofrecer una explicación sobre cómo la mutación fosfomimética del residuo S237 aumenta la estabilidad y la solubilidad de 3WA/S237E, generamos modelos 3D por homología para analizar posibles contactos intramoleculares, ya que aún no se ha determinado la estructura cristalográfica de CGI-58. La Figura 6 muestra los modelos 3D generados para CGI-58 con la topología de la familia de  $\alpha/\beta$  hidrolasas que se caracterizan por tener hélices- $\alpha$  y hojas- $\beta$  conectadas por asas. El núcleo de  $\alpha/\beta$  hidrolasa, la tapa del sitio activo y la secuencia de unión de LD se representan como se describe en la Figura 1A. La Figura 6A muestra el segmento de los residuos 236 a 242 donde se encuentra la S237, aproximadamente en el medio de la tapa del sitio activo. El modelo muestra que la cadena lateral de S237 forma un enlace de hidrógeno con el N de la cadena principal de S238 ( $O_{\gamma}$ -N 2.87 Å), mientras que el N de la cadena principal de S237 forma un enlace de hidrógeno con la cadena lateral de D332 (N -  $O_{\delta 1}$  2.77 Å) (Recuadro de la Figura 6C). Los residuos vecinos de S237 incluyen a R297, que forma un enlace de hidrógeno con el carbonilo O de G324 ( $N_{\eta 1}$ -O 2.94 Å). La Tabla 3 muestra las distancias de las interacciones intramoleculares de la S237 y los residuos vecinos. Las interacciones descritas anteriormente difieren de las reportadas para la proteína de ratón. En el modelo computacional para CGI-58 de ratón, el R299 (R297 en CGI-58 humano) interactúa a través de un puente salino con la cadena lateral de D334 (D332 en humanos)[14], lo que indica algunas diferencias con el modelo humano. Sin embargo, en ambos casos, la arginina interactúa con los residuos del núcleo de  $\alpha/\beta$  hidrolasa.

El reemplazo de S237 por ácido glutámico (E237) provoca cambios en las interacciones intramoleculares, como se muestra en la Figura 6D y la Tabla 3. La cadena lateral de E237 se dirige hacia el asa donde se encuentra R297. Este cambio de orientación hace que las cargas negativas de  $O_{\epsilon 1}$  y  $O_{\epsilon 2}$  de E237 formen interacciones con las cargas positivas de  $N_{\eta 1}$  y  $N_{\eta 2}$  de R297, respectivamente.



**Figura 6. Análisis electrostático y de contacto molecular de la serina 237 (CGI-58 WT) y su sustitución fosfomimética por ácido glutámico (S237E).**

Representación de listones de **(A)** WT y **(B)** S237E, el motivo de unión a la gota lipídica se muestra en naranja (residuos 8-41), el núcleo  $\alpha/\beta$  hidrolasa en azul (residuos 52-349) y la tapa del sitio catalítico en cian (residuos 178-276). El recuadro de (A) muestra la serina 237 (CGI-58 WT), mientras que el recuadro de (B) muestra la fosfomimetización por el ácido glutámico (S237E). El potencial electrostático de superficie de la proteína WT y S237E se muestran en **(C)** y **(D)**, respectivamente. Las regiones con carga positiva son mostradas en azul, y las regiones con carga

negativa se muestran en rojo. **(C)** Una vista de primer plano de S237 y residuos adyacentes (R297, G324, G326 y D332). **(D)** Una vista de primer plano del ácido glutámico fosfomimético (S237E) y los residuos adyacentes antes mencionados. Las líneas discontinuas indican interacciones intramoleculares entre pares de átomos; las distancias correspondientes se enumeran en la Tabla 3.

**Table 3.** Distancias moleculares entre el residuo de serina o ácido glutámico en la posición 237 y residuos cercanos; las distancias medidas para pares de átomos específicos  $d(a_1, a_2)$  se muestran en Å.

Proteína	$d(a_1, a_2)^*$	$a_1$	$a_2$	Distancia, Å
WT	1	S237 (N)	Asp332 (O $\delta_1$ )	2.77
	2	S237 (O $\gamma$ )	Ser238 (N)	2.87
	3	Arg297 (N $\eta_1$ )	Gly324 (O)	2.94
	4	Gly326 (N)	Asp332 (O $\delta_2$ )	2.82
S237E	5	E237 (N)	Asp332 (O $\delta_1$ )	2.80
	6	E237 (O $\epsilon_1$ )	Arg297(N $\eta_1$ )	2.71
	7	E237 (O $\epsilon_2$ )	Arg297(N $\eta_2$ )	2.77
	8	Arg297 (N $\eta_1$ )	Gly324 (O)	2.93
	9	Gly326 (N)	Asp332 (O $\delta_2$ )	2.83

\*mostrados en la figura 6

Según las distancias entre los átomos ( $O_{\epsilon 1}-N_{\eta 1}$ , 2.71 Å; y  $O_{\epsilon 2}-N_{\eta 2}$ , 2.77 Å), estos forman un puente salino, lo que sugiere la estabilización de la tapa del sitio activo y, en particular, de la R297, que ha sido reportado como esencial en la activación de ATGL[14]. Otros residuos que se identificaron a corta distancia (corte de 5 Å) que podrían participar en una interacción intramolecular con E237 incluyen a la A325 y G324 (cuyo N de la cadena principal se encuentra a 4.3 y 4.4 Å, respectivamente) y la cadena lateral de Q333 (a 4.8 Å). Interesantemente, R297, A325, G324 y Q333 se encuentran en el núcleo de  $\alpha/\beta$  hidrolasa de la proteína.

El análisis anterior sugiere que la cadena lateral de S237 en la proteína WT desempeña un papel local al interactuar a través de un enlace de hidrógeno con el N de la cadena principal de S238; ambos residuos están ubicados en la tapa del sitio activo. De acuerdo con nuestra interpretación de los posibles contactos de E237, la sustitución fosfomimética puede generar un efecto estructural global ya que este residuo podría interactuar con un residuo del núcleo de  $\alpha/\beta$  hidrolasa. Esto sugiere que la sustitución fosfomimética funciona como un ancla que fija la tapa al núcleo de  $\alpha/\beta$  hidrolasa, lo que favorece una conformación más estable de CGI-58.

#### **1.4 Conclusiones**

Nuestros resultados muestran que los residuos de triptófano del extremo N-terminal (W19, W23, and W27) conforman un sitio de nucleación que promueve la agregación de la proteína CGI-58. Esto sugiere que, en un contexto celular, la interacción con PLIN1 en la gota lipídica o con ATGL estabiliza a CGI-58 evitando su agregación. Los resultados de solubilidad y estabilidad térmica demuestran que la mutación de fosfomimetización en el sitio S237 mejora la solubilidad y estabilidad de CGI-58. Nuestros resultados en los ensayos de unión a ligando, donde observamos valores de afinidad similares a los reportados para el ortólogo en ratón, junto con los resultados de espectroscopia de CD en regiones UV-lejanos y -cercanos, sugieren que la mutante cuádruple (3WA/S237E) es funcional y mantiene su conformación nativa. Además, el análisis de los modelos bioinformáticos nos sugiere la existencia de una interacción intramolecular entre la cadena lateral del

ácido glutámico (237), localizado en la tapa que protege el posible sitio catalítico, con un residuo del dominio  $\alpha/\beta$  hidrolasa resultando en una conformación más estable y compacta de CGI-58.

## CAPITULO 2

**Estructura cristalográfica de SdrP de *Thermus thermophilus* en complejo con ADN: importancia del residuo Y180 en el reconocimiento del ADN**



### **Objetivo general**

Determinar el mecanismo de reconocimiento en la interacción de SdrP-ADN consenso.

### **Objetivos específicos**

- Realizar un estudio filogenético para identificar aminoácidos conservados en SdrP.
- Generar la construcción que codifique para la proteína recombinante SdrP WT y mutantes.
- Determinar la estructura cristalográfica del complejo SdrP-ADN y analizar las interacciones directas e indirectas.
- Analizar el impacto de la sustitución del residuo Y180 por Ala, Gly y Glu de SdrP en la unión al ADN.

## 2.1 Introducción

Los factores de transcripción tienen un rol fundamental en la expresión génica de las bacterias, ya que le permiten adaptarse a diversos cambios en el ambiente a través de la regulación positiva o negativa de la transcripción de genes en específico [52]. Una de las principales familias de factores de transcripción es la familia CRP/FNR, cuyos miembros se caracterizan por tener un dominio de unión a nucleótidos cíclico (CBD) en el extremo N-terminal y un dominio de unión a ADN (DBD) tipo hélice-giro-hélice (helix-turn-helix, HTH) en el extremo C-terminal [53].

La bacteria *Thermus thermophilus* HB8 (*T. thermophilus*), es un organismo termófilo que no forma esporas y que fue aislado de aguas termales de Japón. Este organismo puede crecer a temperaturas de 47-85 °C y tiene un rango de temperatura óptima de 65-72 °C [54]. En su genoma se encontraron cuatro ORFs (por sus siglas en inglés *Open Reading Frames*) que codifican para proteínas de la familia CRP/FNR, TTHA1437, TTHA1567, TTHA1359 and TTHB099 (No. de acceso de NCBI: YP\_144703, YP\_144833, YP\_144625 and YP\_145338 respectivamente), las cuales tienen una similitud de secuencia de aminoácidos global entre 29-39% [54].

TTHA1437, también conocida como ttCRP, fue el primer factor de transcripción en estudiarse y se determinó que es el ortólogo de la proteína CAP de *Escherichia coli*, el miembro de la familia CRP/FNR más estudiado, ambas proteínas actúan como activadores transcripcionales de manera dependiente de cAMP [55]. A la fecha, se han depositado las estructuras de diversos miembros de la familia CRP/FNR, de las que podemos destacar las estructuras apo-CAP (PDB 3HIF), CAP-cAMP (1G6N) y CAP-cAMP-ADN (1ZRC), entre otras, las cuales han revelado los cambios conformacionales necesarios para que la proteína se una al ADN [56–58]. El mecanismo de control alostérico de CAP es claro y simple y ha sido empleado para describir el mecanismo de regulación de los factores de transcripción: el cAMP se une al sitio CBD del homodímero de CAP y hace contacto directo con los aminoácidos T127 y S128, lo cual induce una transición de asa a hélice (coil-to-

helix) del segmento V126 a F136 y que extiende la hélice C generando la rotación de 60° de la hélice de reconocimiento. Esto provoca el desplazamiento del DBD del dímero de CAP aproximadamente 7° y coloca la hélice de reconocimiento de ADN de los DBDs del dímero en la orientación y posición correcta para interactuar con el surco mayor del ADN [59].

Otro de los factores de *T. thermophilus* HB8 que se ha estudiado es TTHB099. Esta proteína es homóloga de la proteína TT\_P0055 de *T. thermophilus* HB27, con solamente una sustitución de aminoácido [60]. El gen de TT\_P0055 se encuentra localizado en el operon litR del megaplasmido pTT27 y su expresión es regulada por el regulador transcripcional inducible por luz (LitR). En condiciones de luz, la expresión de TT\_P0055 incrementa y regula positivamente la transcripción de genes involucrados en la biosíntesis de carotenoides sin necesidad de alguna molécula efectora o luz [61]. En 2012, Agari et al [60], resolvieron por difracción de rayos X la estructura de la proteína TTHB099 (PDB 3B02), la cual reveló que su estructura es similar a la forma de CAP unida a ADN y reforzó la observación de que TT\_P0055 no requiere de una molécula efectora para realizar su función de activador de la transcripción. Recientemente, Feng et al [62], resolvieron la estructura de un complejo de activación de la transcripción (PDB 5I2D) que incluía la proteína TTHB099, la holoenzima RNAP  $\sigma^A$  de *T. thermophilus*, un promotor dependiente de TTBH099 y un primer ribotetranucleotido. La estructura indica que TTBH099 estimula la isomerización a través de interacciones simples, adhesivas, estabilizadoras proteína-proteína con la holoenzima RNAP y no requiere algún mecanismo alostérico para llevar a cabo su función.

Por último, se encuentra la proteína TTHA1359, también conocida como SdrP (*stationary phase-dependent regulatory protein*). La expresión del mRNA de SdrP se ve incrementada durante la fase estacionaria del microorganismo a 70 °C y aunque el gen que codifica para SdrP no es esencial para *T. thermophilus*, si muestra un defecto en su crecimiento [54]. Los genes que regula este factor de transcripción, están involucrados principalmente en el metabolismo energético y de

nutrientes, control redox y en el metabolismo de ácidos nucleicos durante la fase estacionaria, sin embargo, en el 2010, Agari et al., determinaron que la expresión de SdrP se induce durante la fase logarítmica si la bacteria se encuentra en condiciones de estrés oxidativo [63]. Lo interesante es, que la proteína no tiene residuos de cisteínas o algún cofactor que detecte el estrés oxidativo y en el 2008, Agari et al [54] determinaron que, al igual que TTHB099, SdrP no requiere una molécula efectora para activar la transcripción de sus genes blanco. La estructura de SdrP en ausencia de ADN (apo-SdrP) fue resuelta por difracción de rayos X (PDB 2ZCW) y reveló que tiene una alta identidad estructural con CAP en complejo con el ADN (RMSD de 1.197 Å) [54], sin embargo, aún se desconoce cuál es el mecanismo por el cual se regula su actividad. Con respecto a TTHA1567, no se ha descrito información a nivel estructural para este factor de transcripción.

En este estudio, nosotros determinamos dos estructuras de SdrP en complejo con el ADN por difracción de rayos X, las cuales revelaron las interacciones directas con el ADN por los aminoácidos localizados en las hélices de reconocimiento. De forma interesante, encontramos que el aminoácido Y180, localizado fuera del DBD, participa en el reconocimiento directo de ADN mediante un puente de hidrógeno con el grupo fosfato en la posición 2A. Este mecanismo no se ha reportado en las proteínas de la familia CRP/FNR.

## **2.2 Materiales y métodos**

### **2.2.1 Análisis bioinformático**

La secuencia de aminoácidos de la proteína SdrP (Q5SIL0) fue descargada de UniProt. Para determinar el número de secuencias que pertenecen a la familia CRP/FNR usamos la base de datos InterPro. La construcción del árbol filogenético, se realizó con el árbol guía obtenido en la base de datos Phylome en la opción de búsqueda por secuencia usando la secuencia de SdrP como semilla. De este árbol se identificaron secuencias representativas de la familia CRP con dominio hélice-giro-hélice y el dominio de unión a cNMP. Adicionalmente, se seleccionaron ortólogos estructurales del Protein Data Bank y ortólogos reportados en la

literatura[64]. En total obtuvimos 50 secuencias y se realizó un alineamiento múltiple de secuencia (MSA) con MAFFT versión 7 (<https://mafft.cbrc.jp/alignment/software/>). Posteriormente, se construyó un árbol Máxima Verosimilitud (Maximum Likelihood, ML) con el servidor CIPRES science gateway usando como parámetros la matriz de sustitución de proteínas LG[65]. Finalmente, el árbol se editó utilizando la aplicación web Interactive Tree of Life (ITOL).

La identificación de sitios conservados se realizó mediante el servidor para la Identificación de Regiones Funcionales en Proteínas The ConSurf Server [66]. Se analizó por secuencia de aminoácidos y estructura y utilizó el MSA obtenido en MAFFT con SdrP como secuencia de consulta. Las imágenes se descargaron y visualizaron en Chimera siguiendo las instrucciones del servidor.

### 2.2.2 Clonación de SdrP y construcción de mutantes puntuales

El gen que codifica para la proteína SdrP de *T. thermophilus* fue optimizado para su expresión en *E. coli* y se adquirió clonado en el vector pGOV4 a la empresa GeneOracle Inc. El cADN se subclonó entre los sitios de corte *NdeI* y *HindIII* del vector de expresión pET28b modificado, el cual tiene sustituido el sitio de corte de trombina por el de la proteasa HRV 3C, que permite remover la etiqueta de histidinas en el extremo N-terminal. Esta construcción la denominamos pET28-PPS/SdrP\_WT y a partir de esta realizamos las variantes SdrP\_Y180E, SdrP\_Y180G y SdrP\_Y180A por mutagénesis sitio-dirigida. Para sustituir la Tyr180, se utilizaron oligos externos Fw: 5' GGAAGCAGCCCAGTAGTAGGTTGAGG y Rv: 5' CCGAGATAGGGTTGAGTGTTGTTCC y oligos internos con el codón del aminoácido de interés: Y180E, Fw 5' CCGCTCCGGG**GAG**GGGAAGATCC y Rv 5' GGATCTTCCC**CT**CCCCGGAGCGG, Y180G, Fw 5' CCGCTCCGGG**GGC**GGGAA GATCC y Rv 5' GGATCTTCCC**CG**CCCCGGAGCGG y Y180P, Fw 5' CCGCTCCG **GGCC**GGGAAGATCC y Rv 5' GGATCTTCCC**GG**CCCCGGAGCGG. Los productos de PCR se purificaron y clonaron en el vector pET28-PPS entre los sitios

*NdeI* y *HindIII*. Todas las construcciones se analizaron por secuenciación usando los oligos T7.

### **2.2.3 Expresión y purificación de proteínas recombinantes**

Las construcciones de la proteína SdrP WT y las variantes se transformaron por choque térmico en células competentes de *E. coli* BL21(DE3) *star* y se sembraron en placas de agar LB con kanamicina a una concentración de 50 µg/mL. A partir de las células transformadas se creció un preinoculo de 25 mL en medio LB líquido con el antibiótico y se incubó a 37 °C por 12 h a 200 rpm. El preinóculo se utilizó para inocular medio LB fresco con el antibiótico en una dilución 1:20 y se incubó a 37 °C a 200 rpm hasta una DO<sub>600</sub> de 0.6-0.8; una vez alcanzada la densidad óptica, la expresión de las proteínas recombinantes fue inducida por la adición de IPTG a una concentración de 1.0 mM. El cultivo se incubó por 18 h a 16 °C a 200 rpm. Las células fueron colectadas por centrifugación a 12,000 rpm por 10 min a 4 °C y el pellet se almacenó a -20 °C hasta su uso. Las células fueron resuspendidas en 25 mL de amortiguador de lisis D (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazol) y lisadas por sonicación con 30 ciclos 15 ON/45 OFF a 50% amplitud a 4 °C. Se separó la fracción soluble del lisado celular por centrifugación a 12,000 rpm/10 min/4°C y se cargó a una columna de Ni-NTA-agarose (Qiagen) equilibrada con el amortiguador D. La columna se lavó con 3 volúmenes de columna (CV) del amortiguador de lisis y se realizó la elución con un gradiente escalonado de imidazol (10, 50, 300 y 500 mM). A las fracciones que contenían la proteína recombinante se les adicionó 5 unidades/mL de la proteasa HRV 3C y se incubó a 4°C por 12 h para remover la etiqueta de histidinas. La remoción de la etiqueta de histidinas fue monitoreada por SDS-PAGE y Western-Blot usando el anticuerpo primario dirigido a la etiqueta de Histidinas (6His-tag).

La proteína libre de la etiqueta 6xHis se cambió al amortiguador E [20 mM Tris-HCl, pH 8.0 y 2 mM EDTA] usando una columna PD-10 y se cargó a la columna HiTrap-Q HP (GE Healthcare). La proteína recombinante fue eluída con un gradiente lineal de 0 a 500 mM de NaCl en amortiguador E (15 CV). Las fracciones con la proteína

eluída fueron colectadas y concentradas usando un dispositivo de ultrafiltración (Sartorius Vivaspin Turbo 15, 10 kDa) y posteriormente se purifico por cromatografía de exclusión molecular utilizando una columna Superdex-75 10/300 GL (GE Healthcare) pre-equilibrada en amortiguador F (20 mM Tris pH 8.0, 2 mM EDTA y 150 mM NaCl). Las fracciones con la proteína se colectaron y concentraron con un concentrador Vivaspin 20 (corte de peso molecular de 10 kDa; Sartorius AG). Al finalizar cada paso de purificación, se determinó la concentración de proteína usando Bio-Rad Protein Assay y se monitoreó la pureza de la proteína mediante geles SDS-PAGE al 15%.

#### **2.2.4 Preparación del ADN para ensayos de cristalización**

Para los experimentos de cristalización se utilizaron dos secuencias de ADN dúplex de diferente tamaño (38 y 30 pb), las cuáles se ensamblaron a partir de dos segmentos dúplex de ADN con extremos cohesivos que se sobrelapan en los extremos 5' reconstituyendo la secuencia consenso reportada por Ebright et al.[67] para la proteína CAP. Se adquirieron los siguientes pares de oligonucleótidos purificados por HPLC de fase reversa a IDT (Integrated ADN Technologies, Inc.). Para iSdrP38: 5'-ATTTTCGAAAATTGTGAT-3' (17-mer) y 5'-CTAGATCACAATTTTCGAAAT-3' (21-mer) y para iSdrP30: 5'-CGAAAATTGTGAT-3' (13-mer) y 5'-CTAGAT CACAATTTTCG-3' (17-mer). El ADN dúplex se preparó en una relación de los oligos 1:1 y se calentó a 95°C por 5 min, estos se enfriaron hasta 25°C en gradiente escalonado descendiente de 5 °C incubando 2 minutos en cada temperatura usando un termociclador (Bio-Rad).

#### **2.2.5 Cristalización**

Las condiciones iniciales de cristalización del complejo SdrP\_WT-ADN se realizaron por la técnica de difusión de vapor en gota sentada usando los kits Crystall Screen 1 y 2 (Hampton research, Aliso Viejo, CA, USA) y Morpheus (Molecular Dimensions, Suffolk, England) a 14°C. Se montaron gotas de 2.6 µL en una relación 1:1 de complejo (0.22 mM SdrP, 0.33 mM ADN dúplex, 20 mM Tris pH 8.0, 50 mM MgCl<sub>2</sub>, 150 mM NaCl, 2 mM EDTA, 1 mM DTT y 0.02% azida de sodio) y precipitante, con

150  $\mu$ L en reservorio. Después de 2-3 días de incubación, se obtuvieron cristales en forma de aguja de un tamaño aproximado de 150-200  $\mu$ m de longitud en las condiciones CS1-33, CS2-06 y CS2-15 para el complejo con el ADN iSdrP38 y en la condición CS1-23 para el complejo con el ADN iSdrP30. Los cristales se incubaron 5 segundos en una solución crio-protectora, la cual se consiguió adicionando 20% de glicerol a la solución precipitante y se montaron en crio-loops (Mitegen) para posteriormente ser congelados a 100 K (Flash-frozen).

### **2.2.6 Colección de datos de la difracción de rayos X y determinación de la estructura**

Los datos de difracción de rayos-X fueron colectados usando un difractómetro in-house Rigaku/MSC Micromax-007 HF (Instituto de Química, UNAM) con un detector DECTRIS-PILATUS 3R/200K-A, bajo condiciones criogénicas a 100 K. Los datos de difracción para determinar la estructura fueron de un solo cristal con rotación de 1° por imagen. El programa HKL3000 fue usado para integrar y escalar las intensidades de difracción.

La estructura del complejo SdrP-ADN fue resuelta por remplazo molecular, utilizando el programa Phaser. La estructura de SdrP (PDB: 2ZCW) y CAP-ADN (PDB: 1ZRC) fueron utilizadas como modelos de búsqueda para el remplazo molecular (search models). El refinamiento y la construcción del modelo se realizó de forma iterativa utilizando los programas phenix.refine del paquete PHENIX versión dev-2776-000 y el programa COOT. La validación se realizó con el programa MOLPROBITY.

### **2.2.7 Parámetros helicoidales del ADN**

Los parámetros del ADN-helicoidal fueron analizados con el software 3ADN[68]. Los ángulos Twist, Roll y overall bend del ADN fueron determinados siguiendo las instrucciones de Lu & Olsen[69].



### **2.2.8 Análisis de la unión proteína-ADN**

Para determinar la capacidad de unión de SdrP por el ADN consenso y promotores nativos de *T. thermophilus* se realizó el ensayo de cambio de movilidad electroforética (EMSA)[70]. Se usó ADN marcado con radioactividad como trazador a concentración constante y variando la concentración de las proteínas en diluciones seriadas. La siguiente secuencia de ADN dúplex fue utilizada: iSdrP30: 5'-Cy5-CGAAAATTGTGATCTAGATCACAATTTTCG, adquirida a la empresa IDT ADN Technologies. Las muestras para EMSA fueron preparadas en reacciones de 20  $\mu$ L que contenían 1 nM de ADN consenso, Tris-HCl 10 Mm pH 7.5, 50 mM NaCl, 0.05 mg/ml BSA, 1.5 Mm DTT y la concentración de proteína WT, Y180A, Y180E y Y180G correspondiente. Las reacciones se incubaron durante 20 min a 4 °C y se detuvieron con la adición del amortiguador que contenía 40% sacarosa, 20 mM EDTA, y 0.05 % de azul de bromofenol. Las muestras, se corrieron en geles al 6% y se expusieron durante 3 días para su digitalización.

## **2.3 Resultados y Discusión**

### **2.3.1 Características estructurales de ortólogos de SdrP**

SdrP es un factor de transcripción de la familia CRP/FNR de la bacteria termófila *T. thermophilus* HB8 y su expresión se incrementa en la fase estacionaria durante su incubación a 70 °C[54]. Los miembros de la familia CRP/FNR se caracterizan por ser proteínas de 200 a 250 aa que contienen dos dominios importantes: el motivo Hélice-Giro-Hélice (HTH, por las siglas en inglés de Helix-Turn-Helix) en el C-terminal que consiste en dos hélices- $\alpha$  unidas por un giro y que se une al surco mayor del ADN; el segundo es el dominio de unión a nucleótido que se encuentra en el extremo N-terminal, por sí mismo, este dominio definido por la similitud con los dominios de unión a cAMP, no confiere especificidad a la familia de reguladores [53]. En 1994, cuando se propuso la clasificación de los reguladores CRP/FNR, solamente existían 22 proteínas reportadas y CAP/CRP era la proteína modelo para describir a los otros miembros, no obstante, gracias a las tecnologías de análisis de genomas, el número de proteínas pertenecientes a esta familia ha aumentado considerablemente[53]. De acuerdo con la base de datos InterPro para la

clasificación de familias de proteínas, actualmente existen 93,409 secuencias de la familia CRP/FNR en bacterias, de las cuales los filos más representativos son el *Proteobacteria* con 52,033 secuencias, *Bacteroidetes* 9,152, *Firmicutes* 12,543, *Actino bacteria* 10,187 y *Deinococcus-Thermus* 526. Dentro del filo *Deinococcus-Thermus* se encuentra *Thermus thermophilus* HB8, que además de SdrP tiene otras tres proteínas pertenecientes a la familia CRP/FNR, ttCRP, LdrP y TTHA1567 con las cuales tiene una similitud de 32%, 36% y 33.5%, respectivamente, lo que sugiere que son evolutivamente diferentes. Para determinar la relación evolutiva de SdrP con otros factores de la familia CRP/FNR realizamos un árbol filogenético de Máxima Verosimilitud. En la Figura 7A podemos observar que SdrP se agrupa en el clado con otros factores de transcripción del filo *Deinococcus-Thermus* y mantiene una relación cercana con LdrP de *T. thermophilus*. No obstante, ttCRP se agrupa en el segundo clado que tiene mayor relación con *Proteobacterias*, *Firmicutes* y *Actino bacterias*. Interesantemente, en este segundo clado se agrupan los factores de transcripción de organismos que han sido ampliamente estudiados y con los que se cuenta estructura 3D resuelta por cristalografía de rayos X (color verde en Fig. 7A). Ejemplos de ellos son CAP/CRP de *E. coli*, el factor de transcripción más estudiado de la familia CRP/FNR, Clp de *Xanthomonas campestris*, CRP de *Mycobacterium tuberculosis*, Vfr de *Pseudomonas aeruginosa*, entre otras. Se ha reportado que la unidad funcional biológica de estos factores de transcripción está presente como dímero y se une al ADN a través de su hélice de reconocimiento. Sin embargo, el mecanismo de regulación es diferente en algunos casos, como CAP que se activa por la unión del cAMP, y Clp que tiene una regulación negativa al unirse a c-di-GMP [59]. Para determinar las diferencias de SdrP con distintos miembros de la familia CRP/FNR, realizamos un alineamiento múltiple de secuencias (MSA) donde identificamos los residuos de aminoácidos conservados (Fig. 7B) y la estructura secundaria, en la que se observa que el dominio N-terminal de SdrP consiste en una hélice  $\alpha$  ( $\alpha_1$ , residuos 62-65) y ocho hojas  $\beta$  ( $\beta_1$ , residuos 8-10;  $\beta_2$  residuos 15-17;  $\beta_3$ , residuos 29-33;  $\beta_4$ , residuos 36-41,  $\beta_5$ , residuos 47-53;  $\beta_6$ , residuos 58 y 59;  $\beta_7$ , residuos 72-75; y  $\beta_8$ , residuos 79-83) que adoptan un plegamiento de una doble hoja de  $\beta$ -hélices con una topología “jelly roll”. El domino

C-terminal consiste en cuatro hélices  $\alpha$  ( $\alpha_3$ , residuos 119-130;  $\alpha_4$ , residuos 149-156;  $\alpha_5$ , residuos 160-172; y  $\alpha_6$ , residuos 188-195) y cuatro hojas  $\beta$  ( $\beta_9$ , residuos 136-139;  $\beta_{10}$ , residuos 142-146;  $\beta_{11}$ , residuos 176-179; y  $\beta_{12}$ , residuos 182-185) que adoptan un plegamiento hélice-giro-hélice. Estos dos dominios son conectados por una gran hélice  $\alpha$  ( $\alpha_2$ , residuos 90-116), que constituye la región de interacción para la formación del homodímero. En la tabla 4 se reportan los porcentajes de identidad del alineamiento de secuencia y los valores de la raíz de la desviación cuadrática media (RMSD, por las siglas en inglés root-mean-square deviation) del alineamiento estructural, el cual es la medida cuantitativa más utilizada de la similitud entre dos coordenadas atómicas superpuestas; el RMSD es 0 para estructuras idénticas y su valor aumenta a medida que las dos estructuras se vuelven más diferentes. Los valores de RMSD se consideran indicadores fiables de variabilidad cuando se aplican a proteínas muy similares [71]. Como ya se había mencionado anteriormente, SdrP tiene un bajo porcentaje de identidad en secuencia con los otros dos factores de transcripción presentes en *T. thermophilus*, sin embargo, el alineamiento estructural resultó en un valor de RMSD de 1.38 Å con LdrP y de 10.03 Å con ttCRP. SdrP presenta una mayor homología estructural con LdrP, factor de transcripción que no requiere de molécula efectora para unirse al ADN. Por el contrario, el valor de RMSD con ttCRP que requiere de cAMP para modular su actividad es cercano a 10, lo que nos indica una menor homología estructural, no obstante, es interesante que SdrP presenta homología estructural cercana con otros factores de transcripción que requieren de cAMP como el factor de *E. coli* (1.15 Å), *S. typhimurium* (1.64 Å), *P. aeruginosa* (1.57 Å) y *M. tuberculosis* (1.14 Å) o que tienen otra molécula efectora como los casos de c-di-GMP en Clp de *X. campestris* (2.0 Å) y el 2-oxoglutarado de NtcA de *S. elongatus* (1.80 Å). En 2008, Agari y col. reportaron que SdrP no requiere de molécula efectora para unirse al ADN y que al superponer la estructura de apo-SdrP con la de CAP/CRP de *E. coli*, los residuos G72, E73, L74, R83, S84, T128 y S129 de CAP/CRP de *E. coli*, los cuales se ubican en el sitio de unión de cAMP, corresponden con los residuos G60, E61, E62, R69, Y71, A107 y Y108 en SdrP (Fig. 7B) y concluyeron que las cadenas laterales de los residuos E61, E62, Y71 y Y108 penetran en un espacio correspondiente al bolsillo

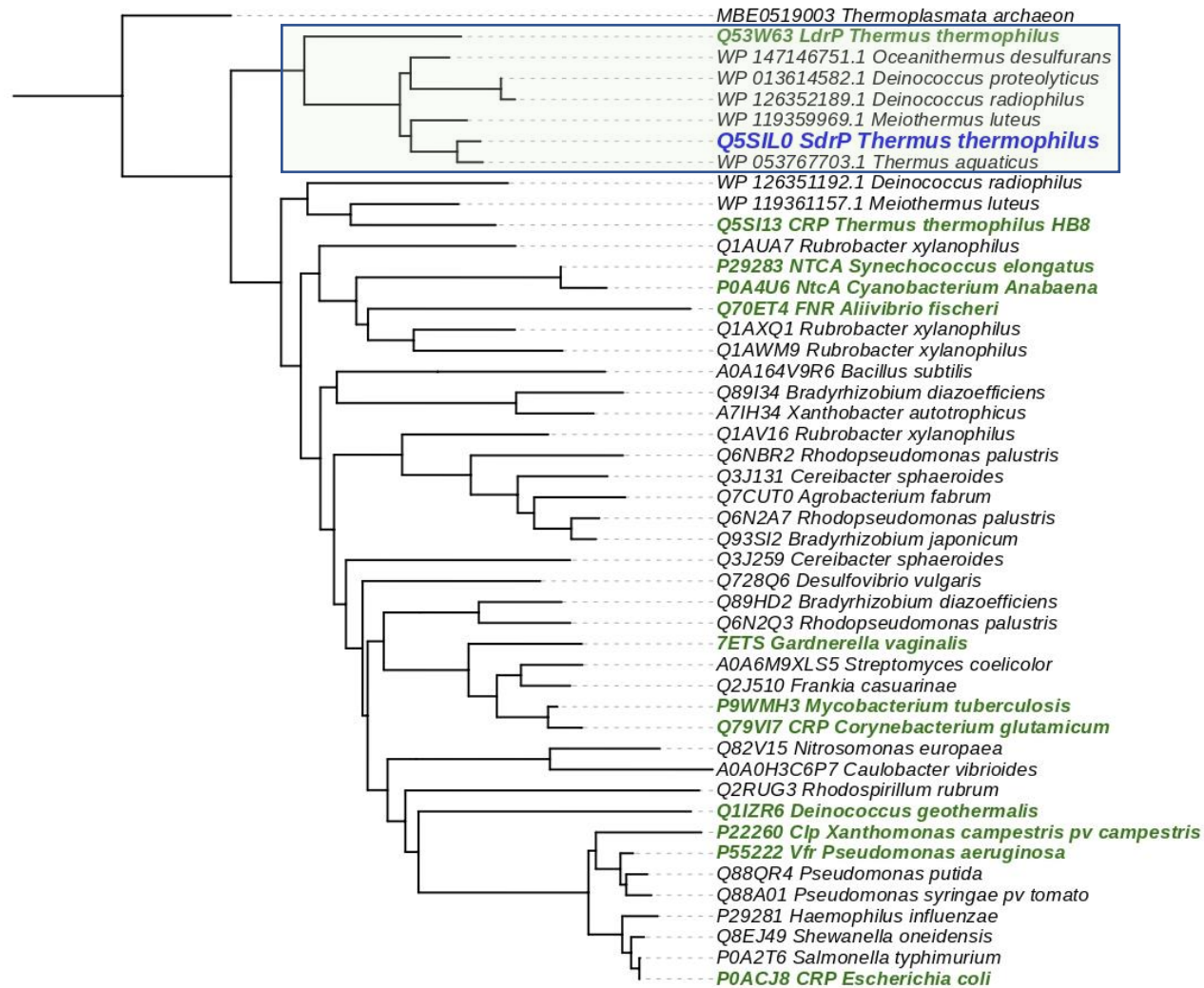
de unión de cAMP de CAP/CRP de *E. coli*[54]. En cuanto a los residuos de aminoácidos en el dominio de unión al ADN del C-terminal, podemos observar que la hélice  $\alpha_5$  que corresponde a la hélice de reconocimiento del ADN, se conservan los aminoácidos con carga positiva y negativa (R160 y E161), los cuales realizan una interacción directa con las bases nitrogenadas del ADN.

Estos datos respaldan que, los ortólogos estudiados presentan una estructura similar pero la identidad de secuencia y mecanismo de regulación, son diferentes. Se ha establecido que los ortólogos de esta familia presentan estructuras altamente conservadas en distintas especies de bacterias y mantienen su capacidad para reconocer secuencias de ADN independientemente de su tipo de regulación, sin embargo, la filogenómica bacteriana no puede explicarse usando los marcos teóricos que sustentan la genética de poblaciones clásica, lo que posiblemente se deba a la transferencia horizontal de genes entre estos organismos [64]. Con esto en mente, nos enfocamos en analizar las características fisicoquímicas y estructurales de los factores de transcripción de la familia CRP/FNR cuya estructura ha sido reportada.

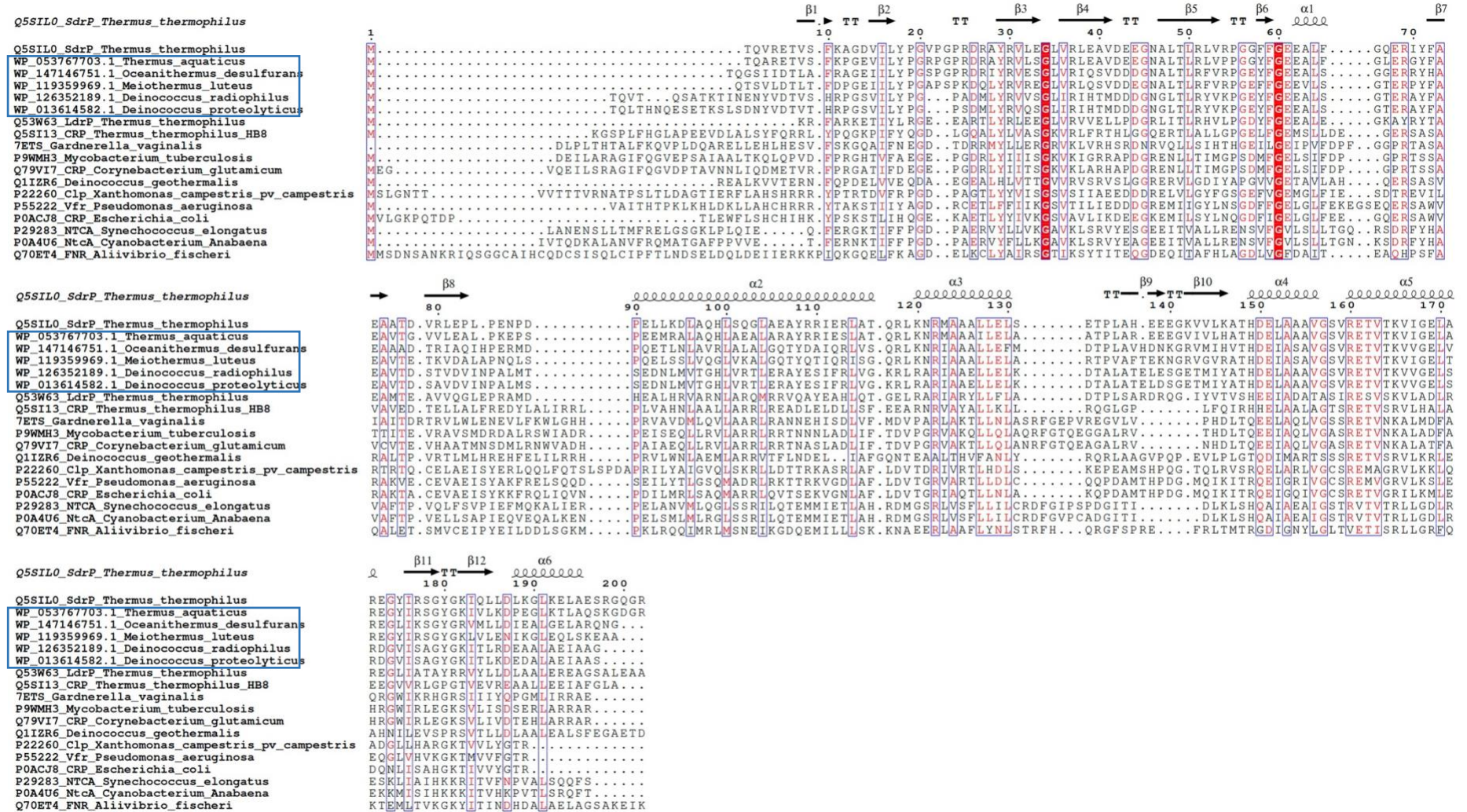
En la Figura 8A, se observan los residuos de aminoácidos conservados en la estructura de SdrP resultado del análisis en Consurf. En este se empleó el MSA de la figura 7B. Los residuos más conservados se identifican con el color morado y los menos conservados en color verde. De acuerdo con lo representado, los residuos de aminoácidos de SdrP G34, A75, R122, T148, A153, R160, E161 son los más conservados entre los ortólogos cristalizados. Estos dos últimos se encuentran en la hélice de unión al ADN y son clave para la interacción con este. Por otra parte, la Figura 8B, muestra los sitios conservados de SdrP usando el MSA donde solo se incluyeron los ortólogos del clado donde se encuentra SdrP (Recuadro de la figura 7A). Los sitios conservados aumentan G34, A75, R122, T148, A153, R160, E161, K165, Y180 y podemos observar que se incluye la K165 ubicada en la hélice de unión al ADN y la Y180 que se encuentra en el asa que conecta el motivo HTH.

## A. Árbol filogenético familia CRP/FNR

Tree scale: 1



## B. Alineamiento múltiple de secuencia miembros de la familia CRP/FNR



**Figura 7. Análisis de los ortólogos de SdrP en diferentes especies de bacterias.** A) Árbol filogenético de factores de transcripción de la familia CRP/FNR construido con Máxima Verosimilitud usando CIPRES. En azul se identifica a SdrP de *T. thermophilus* y en verde sus ortólogos cuya estructura se ha determinado experimentalmente. B) Alineamiento múltiple de secuencia de los ortólogos de SdrP con estructura resuelta y los ortólogos de SdrP del clado marcado en recuadro verde en el árbol filogenético. En recuadro azul se marcan los homólogos estructurales de SdrP cuya estructura no se ha resuelto experimentalmente. En rojo se identifica los aminoácidos conservados. El alineamiento se realizó con MAFFT y la figura fue generada con ESPript 2.2.

**Tabla 4.** Proteínas de la familia CRP/FNR y su comparación con SdrP

Organismo	ID uniprot	Proteína	PDB	Molécula efectora	aa*	% Identidad con SdrP (aa)	RMSD** Å
<i>Thermus thermophilus HB8</i>	Q5SIL0	SdrP	2ZCW	ND	202	100	0.0 (194/194)
<i>Thermus thermophilus HB8</i>	Q53W63	LdrP	3B02	ND	195	36.08	1.378 (146/146)
<i>Thermus thermophilus HB8</i>	Q5SID7	CRP	4EV0	cAMP	216	31.94	10.031 (147/147)
<i>Escherichia coli</i>	P0ACJ8	CRP	1O3T	cAMP	210	21.7	1.153 (178/178)
<i>Salmonella typhimurium</i>	P0A2T6	CRP	AlphaFold	cAMP	210	21.1	1.674 (131/131)
<i>Mycobacterium tuberculosis</i>	P9WMH3	CRP	4A2U	cAMP	224	25.8	2.141 (161/161)
<i>Pseudomonas aeruginosa</i>	P55222	Vfr	2OZ6	cAMP	214	24.9	1.565 (141/141)
<i>Xanthomonas campestris pv. campestris</i>	P22260	Clp	3IWZ	c-di-GMP	230	21.6	2.005 (130/130)
<i>Synechococcus elongatus</i>	P29283	NtcA	2XGX	2-Oxoglutarato	222	25.3	1.802 (137/137)
<i>Aliivibrio fischeri</i>	Q70ET4	FNR	5CVR	Fe/S	250	16.6	5.140 (127/127)
<i>Cyanobacterium Anabaena</i>	P0A4U6	NtcA	3LA2	2-Oxoglutarato	223	29.9	1.662 (133/133)
<i>Gardnerella vaginalis</i>	---	CRP	7ETS	ND	226	27.3	3.172 (158/158)
<i>Corynebacterium glutamicum</i>	Q79VI7	CRP	3R6S	cAMP	227	27.8	2.527 (152/152)
<i>Listeria monocytogenes</i>	P22262	Listeriolysin regulatory protein	2BEO	ND	237	17.7	9.733 (153/153)
<i>Deinococcus geothermalis</i>	Q11ZR6	Transcriptional regulator	3E97	ND	210	21.5	2.779 (127/127)

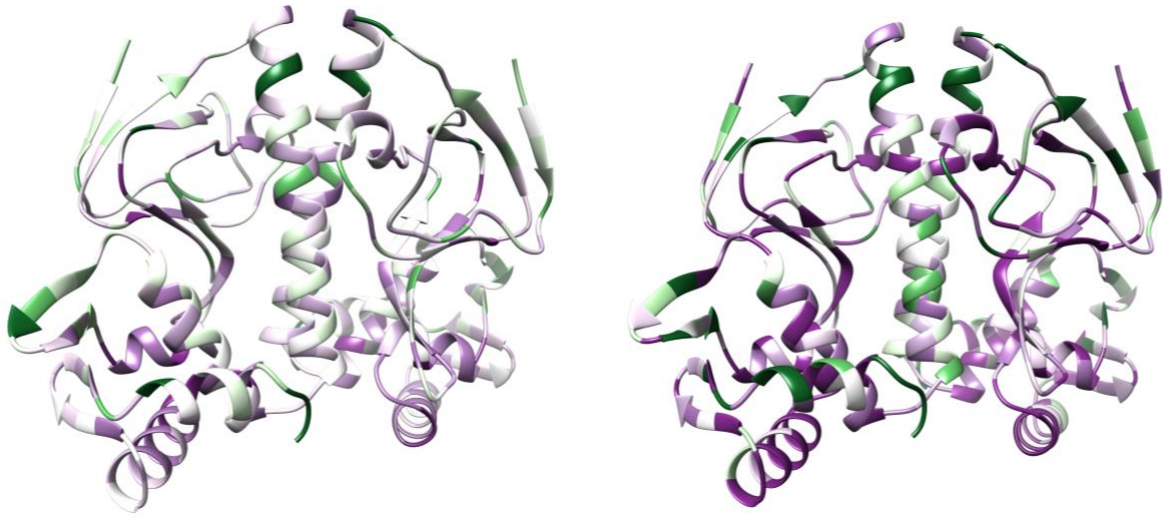
\*aa: aminiácidos

\*\*RMSD: root-mean-square deviation de las posiciones atómicas

ND: no determinado



**A. Sitios conservados familia CRP/FNR    B. Ortólogos cercanos a SdrP**



**Figura 8. Identificación de la conservación de residuos de aminoácidos en la estructura de SdrP.** A) Sitios conservados en SdrP de acuerdo con el análisis de Consurf usando el MSA de las secuencias empleadas para el árbol filogenético de la figura 7A. B) Sitios conservados en SdrP de acuerdo con el análisis de Consurf usando el MSA solamente de las secuencias del clado donde se ubica SdrP en el árbol filogenético de la figura 7A. En color morado se identifican los aminoácidos más conservados, en blanco los promedio y en verde los aminoácidos más variables.

Para determinar si estos residuos de aminoácidos tienen relevancia en la unión al ADN, se determinó la estructura cristalográfica de SdrP en complejo con ADN.

### **2.3.2 Características generales de la estructura del complejo SdrP-ADN**

Preparamos complejos de SdrP-ADN usando fragmentos de ADN de 30 y 38 pb con el sitio consenso reportado para CAP (Figura 9A y C)[67]. El fragmento de ADN de 30 pb dio cristales en el grupo espacial P 31 y el ADN de 38 pb en el grupo P6522. Los datos de difracción fueron colectados a una longitud de onda de 1.54 Å y la resolución observada fue de 2.99 Å y 3.1 Å para los complejos SdrP/ADN30pb y SdrP/ADN38pb, respectivamente (Tabla 5). El análisis de las estructuras del complejo SdrP-ADN de este estudio nos reveló evidencia importante sobre la interacción y los ángulos de doblez del ADN, la estructura global del complejo se muestra en la Figura 9B y D. Cada estructura contiene dos semicomplejos no simétricos (non-equivalent half-complexes), que reconstituyen al dímero de SdrP unido al sitio consenso de ADN. Ambas estructuras, que solo difieren en la longitud del ADN, comparten similitudes en conformaciones locales y generales de los ángulos del ADN, incluso en dos diferentes grupos espaciales, lo que nos hace pensar que estas similitudes son reales y no un artefacto de empaquetamiento del cristal.

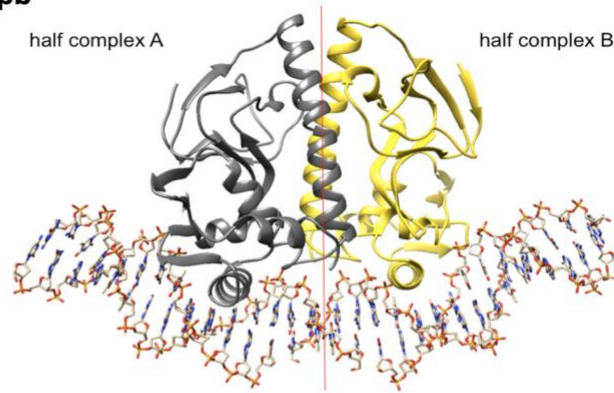
La estructura global de SdrP en complejo con el ADN tiene el mismo plegamiento que apo-SdrP (2ZCW)[54] y es altamente similar con un RMSD de 0.591 Å. En el alineamiento estructural del complejo con apo-SdrP solo difiere en las cadenas laterales que tienen contacto con el ADN. También, se comparó la estructura de SdrP con el factor de transcripción de la familia CRP/FNR más estudiado, la proteína CAP unida a cAMP o en complejo con ADN [56,72]. Las diferencias más notables entre estas dos proteínas son: 1) SdrP no tiene la sección de los aminoácidos del 1-14 del N-terminal de CAP y 2) SdrP tiene una hélice alfa extra en el extremo C-terminal que comprende los aminoácidos del 190-202 (Figura 7B). En el Protein Data Bank, se han depositado tres estructuras de CAP en complejo con el ADN que utilizamos como referencia en este trabajo; al hacer un alineamiento estructural del

**A. ADN consenso de 30 pb**

iSdrP30

```
-4 -3 -2 -1 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  
C G A A A A T T G T G A T C T A G A T C A C A A T T T T C G  
G C T T T T A A C A C T A G A T C T A G T G T T A A A A G C
```

**B. Complejo SdrP-ADN 30 pb**

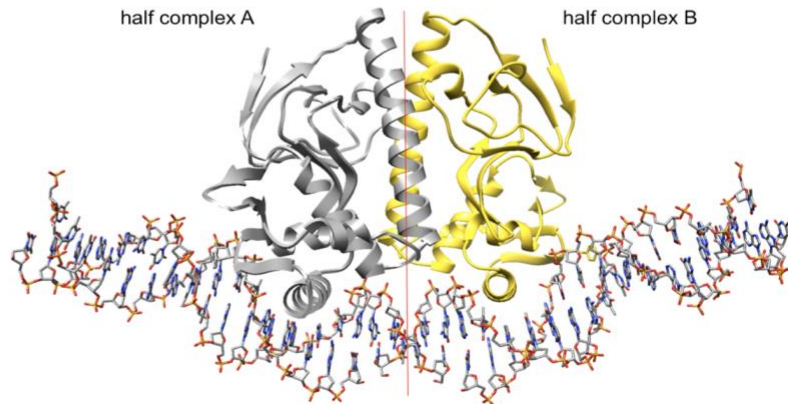


**C. ADN consenso de 38 pb**

iSdrP38

```
-8 -7 -6 -5 -4 -3 -2 -1 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  
A T T T C G A A A A T T G T G A T C T A G A T C A C A A T T T T C G A A A T  
T A A A G C T T T T A A C A C T A G A T C T A G T G T T A A A A G C T T T A
```

**D. Complejo SdrP-ADN 38 pb**



**Figura 9. Fragmentos de ADN usados para la cristalización del complejo y estructura del complejo de SdrP-ADN.** (a) Fragmento de ADN de 30 pb usado para la cristalización en el grupo espacial P31. (b) Estructura del complejo SdrP-ADN cristalizado en el grupo espacial P31. (c) Fragmento de ADN de 38 pb usado para la cristalización en el grupo espacial P6522. (d) Estructura del complejo SdrP-ADN cristalizado en el grupo espacial P6522. La estructura que se muestra de SdrP en (b) y (d) está representada en listones y fue generada usando Chimera.

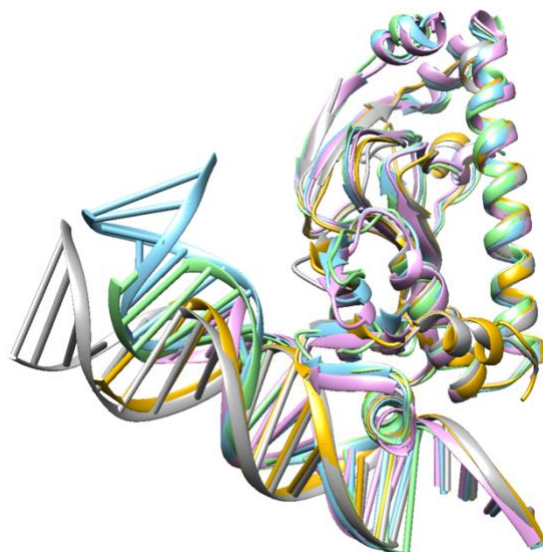
**Tabla 5. Recopilación de datos de estructura y estadísticas de refinamiento.**

	<b>SdrP-ADN30</b>	<b>SdrP-ADN38</b>
<b>A. Parámetros de colección de datos</b>		
Fuente de Rayos X	Instituto de Química, UNAM	
Temperatura (K)	100	
Grupo espacial	P 31	P 65 2 2
Parámetros de celda unitaria		
a (Å)	58.49	89.97
b (Å)	58.49	89.97
c (Å)	172.53	234.95
Resolución (Å)	2.99	3.1
Reflexiones observadas		
Reflexiones únicas	12107 (1038)	10880 (1051)
Completadas (%)	91.28 (77.49)	99.45 (98.57)
<b>B. Refinamiento</b>		
Factor-B promedio (Å)	29.02	75.60
R-factor (%)	0.2069	0.2719
R-free (%)	0.2538	0.3203

complejo SdrP-ADN con las de CAP resultó con un RMSD de 1.264 Å (1ZRC), 1.381 Å (1O3Q) y 1.285 Å (1O3T). Las principales diferencias las podemos observar en la conformación del ADN (Fig. 10). Donde observamos que el dobléz del ADN por SdrP es más suave y menos pronunciado que el reportado por CAP. De acuerdo con los resultados del software x3ADN, el semicomplejo de SdrP-ADN provoca un dobléz general (overall bend) de 43°. En el 2006, Napoli et al., determinaron que la unión de CAP al ADN consenso (PDB 1ZRC) genera un dobléz general de 48° en el semicomplejo de la cadena A y una curvatura más suave en el complejo medio de la cadena B de 38°. Nuestros resultados muestran que el complejo de SdrP provoca dobléz similar al del complejo medio de la cadena B de CAP [73].

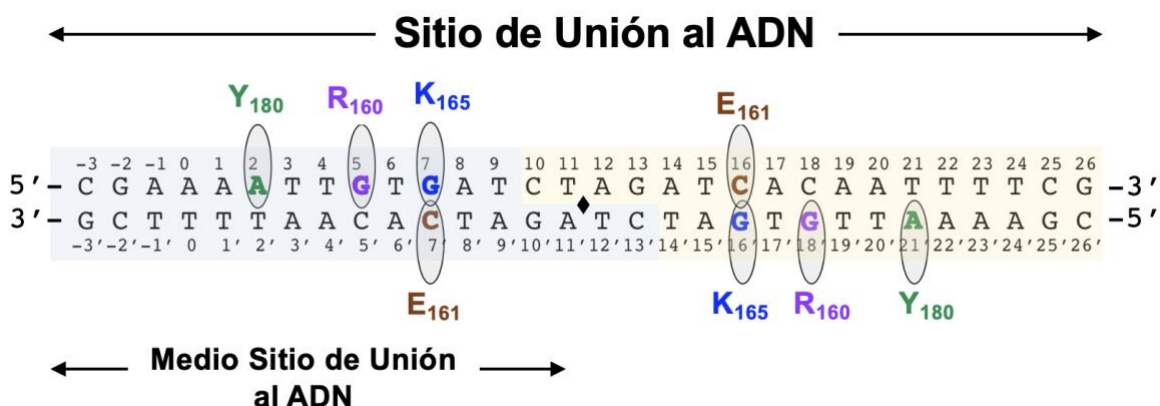
Como ya se ha mencionado, el reconocimiento del ADN por CAP ha sido el caso más estudiado de los factores de transcripción de la familia CRP/FNR; y se ha determinado que la interfase de interacción de proteína-ADN se caracteriza por: 1) un modo de reconocimiento directo (direct readout mode) a través de los residuos de aminoácidos en la hélice de reconocimiento y que forman interacciones con los nucleótidos del ADN. 2) El reconocimiento indirecto, que se describe en términos de desviaciones en la geometría del ADN con respecto a la estructura ideal o en términos de cambios en la geometría entre conjuntos de estructuras observadas[58], involucrando desviaciones en la estructura del ADN, incluyendo la generación de un dobléz y una curvatura a través de los ángulos “roll” y “twist” [74].

De acuerdo con la Figura 11, el reconocimiento directo entre SdrP y los pares de bases del ADN de 30 pb es: a) en el surco mayor del motivo TGTGA, el cual involucra la interacción de los residuos Arg-160, Glu-161 y Lys-165 localizados en la hélice  $\alpha_5$  de reconocimiento y b) en el surco menor a través de la Tyr-180 localizada en el asa adyacente a la hélice de reconocimiento. En la tabla 6, se reportan las distancias de los residuos que participan en el reconocimiento directo del ADN de 30 y 38 pb, los cuales fueron muy similares. En el surco mayor, la interacción de la Arg-160 se da a través del grupo guanidino de su cadena lateral, el cual forma un par de puentes de hidrógeno con los átomos O6 y N7 de la guanina

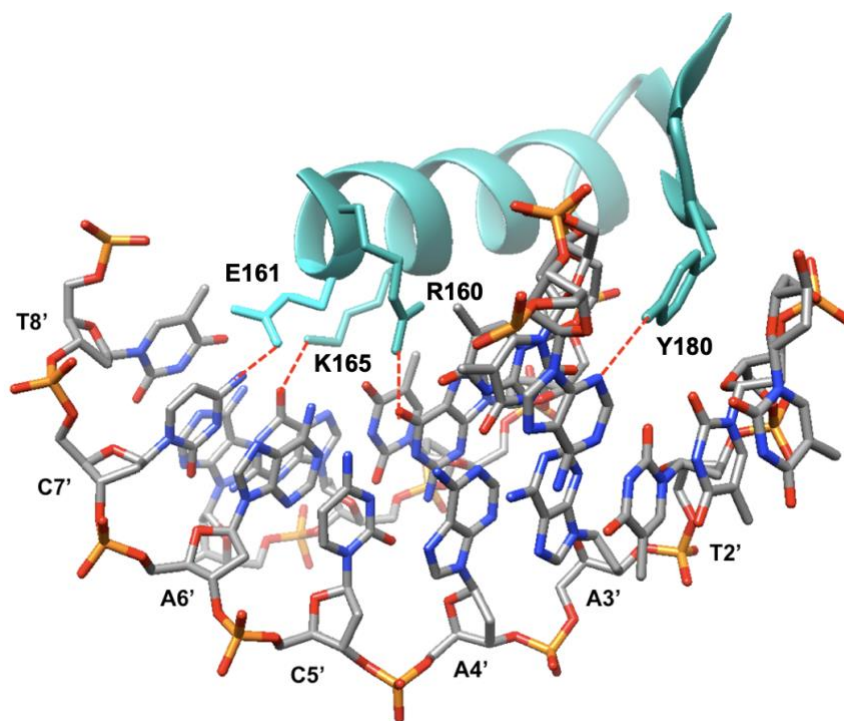


**Figura 10. Superposición de SdrP-ADN\_30pb (amarillo), SdrP-ADN\_38pb (gris), 1ZRC (azul), 1O3Q (Verde) y 1O3T (Rosa).** La estructura general de las proteínas homólogas es muy similar observándose un RMSD de 1.264 Å (1ZRC), 1.381 Å (1O3Q) y 1.285 Å (1O3T). Sin embargo, el ángulo de doblez del ADN de CAP es más pronunciado (48°) que el de SdrP (43°).

### A. ADN consenso de 30 pb



### B. Motivo HTH de SdrP en complejo con ADN consenso de 30 pb



**Figura 11. Interacciones directas de SdrP con el ADN observadas en el complejo SdrP-ADN.** A) Esquema que representa las interacciones directas de SdrP con el ADN de 30 pb. Las regiones sombreadas con azul y amarillo muestran los pares de oligonucleótidos con extremos cohesivos usados para formar el ADN dúplex. El rombo negro en el centro señala la región que divide el sitio de unión simétrico. En colores se muestran los residuos de aminoácidos que realizan contacto directo con el ADN y en ovalo la posición del ADN. B) Motivo hélice-giro-

hélice de SdrP (R160-R172), el asa adyacente donde se localiza la Y180 y las posiciones 2-8 del medio sitio de unión en el ADN. Como referencia, el pliegue primario se encuentra entre las posiciones 6 y 7 del medio sitio de unión del ADN. Los enlaces de hidrógeno entre las cadenas laterales de los aminoácidos de SdrP y las bases nitrogenadas del ADN se muestran en líneas punteadas rojas (se tomó como límite puentes de hidrógeno a 4 Å de distancia).

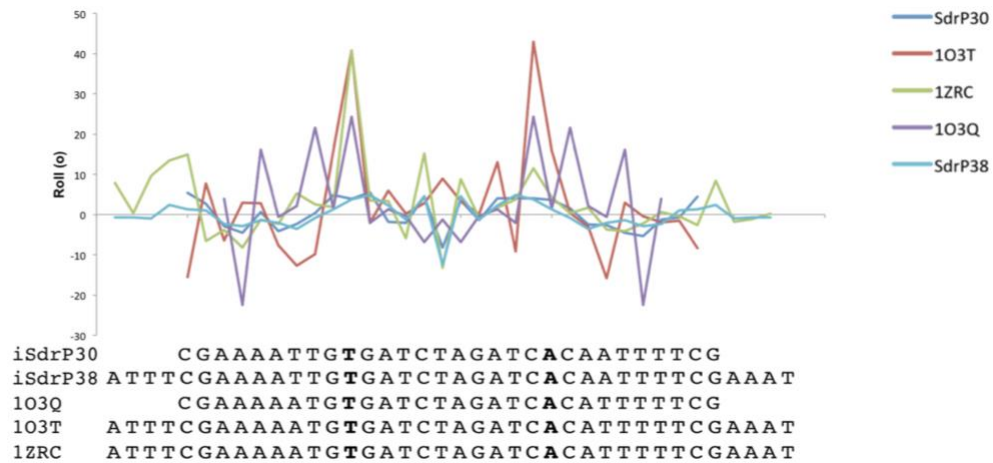
**Tabla 6.** Distancias moleculares entre los residuos de aminoácidos ( $a_1$ ) de SdrP que realizan el reconocimiento directo con las bases ( $N_2$ ) del ADN; las distancias medidas para pares de átomos específicos  $d(a_1, N_2)$  se muestran en Å.

ADN	$a_1$	$N_2$	Distancia, Å
30 pb	R160 ( $N_{\eta 1}$ )	G <sub>5</sub> (O6)	2.52
	R160 ( $N_{\eta 2}$ )	G <sub>5</sub> (N7)	3.24
	E161 ( $O_{\epsilon 1}$ )	C <sub>7</sub> (N4)	2.82
	K165 ( $N_{\zeta}$ )	G <sub>7</sub> (O6)	2.92
	Y180 (OH)	A <sub>2</sub> (N3)	2.85
38 pb	R160 ( $N_{\eta 1}$ )	G <sub>5</sub> (O6)	3.55
	R160 ( $N_{\eta 2}$ )	G <sub>5</sub> (N7)	3.73
	E161 ( $O_{\epsilon 1}$ )	C <sub>7</sub> (N4)	2.84
	K165 ( $N_{\zeta}$ )	G <sub>7</sub> (O6)	2.94
	Y180 (OH)	A <sub>2</sub> (N3)	2.67

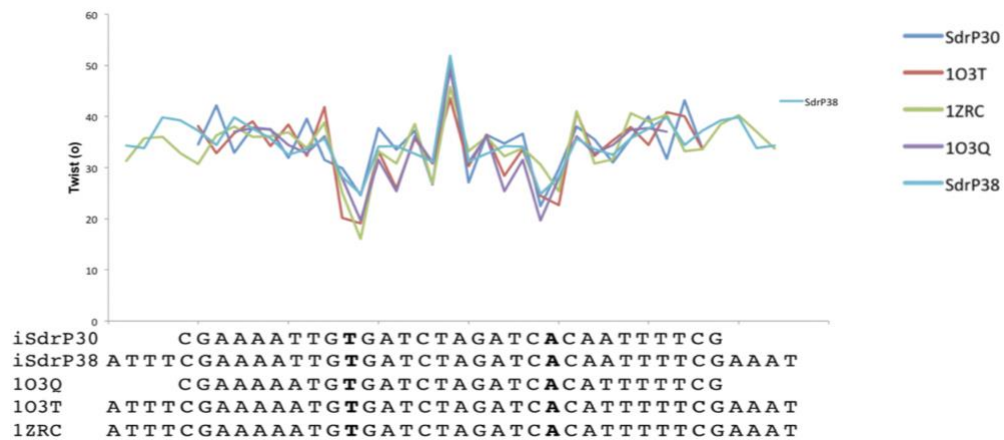


en la posición 5 (G5) a una distancia de 2.52 y 3.24 Å, respectivamente. La cadena lateral del residuo Glu-161 forma un puente de hidrógeno (2.82 Å) con el átomo N4 de la citosina en la posición 7 (C7) y finalmente, la cadena lateral de la Lys-165 forma un puente de hidrógeno (2.92 Å) con el O6 de la guanina en la posición 7 (G7). Respecto al surco menor, el grupo hidroxilo de la Tyr-180 forma un puente de hidrógeno (2.85 Å) con el átomo N3 de la adenina en la posición 2 (A2) (Figura 11B). Esta tirosina no se encuentra conservada en otros factores de transcripción de otros organismos cuya estructura cristalográfica se ha reportado excepto en LdrP (Figura 7B). A diferencia de SdrP, los contactos directos de CAP con los pares de bases del ADN son solamente en el surco mayor del motivo TGTGA e involucra los residuos R180, E181 y R185 de la hélice de reconocimiento de CAP [58,72,73,75,76]. Al tener SdrP una Lisina (K165) en la hélice de reconocimiento en lugar de la segunda arginina (R185) (Figura 7B), provoca que el número de contactos con el ADN en el surco mayor sea menor, mientras que en CAP la R185 forma dos puentes de hidrógeno con dos bases nitrogenadas (G7 y T8'), la K165 de SdrP solamente forma un puente de hidrógeno con la guanina (G7) [77]. Esto se debe principalmente a la naturaleza de la cadena lateral del aminoácido, a pesar de que los dos son aminoácidos con carga positiva, la arginina contiene el grupo guanidinio que se caracteriza por tres átomos de nitrógeno enlazados a un mismo átomo de carbono, en cambio la lisina solamente tiene un nitrógeno enlazado al carbono [78]. Para analizar el reconocimiento indirecto de SdrP en el ADN, se determinaron los parámetros de las hélices del ADN mediante los gráficos de los ángulos "roll" y "twist" (Figura 12). En la Figura 12A, se graficaron los ángulos "Roll" del paso de par de bases (base pair step) y se observa que SdrP genera el doblez principal en los dos sitios de unión, con ángulos de 6° en la posición 6T (medio complejo de la subunidad A) y 8° en la 17A (subunidad B), sin embargo, estos ángulos son aproximadamente siete veces menores que el determinado para la subunidad A de la estructura de CAP 1ZRC, donde se determinó el ángulo "Roll" de 41° pero es similar a la subunidad B de 12° [73]. En cuanto a los ángulos "Twist" (Figura 12B), se observa un patrón similar a los reportados con CAP. En SdrP determinamos un ángulo "twist" de aproximadamente 25° entre la posición 6 y 7 del medio sitio de

### A. Parámetro del ADN "roll"



### B. Parámetro del ADN "twist"



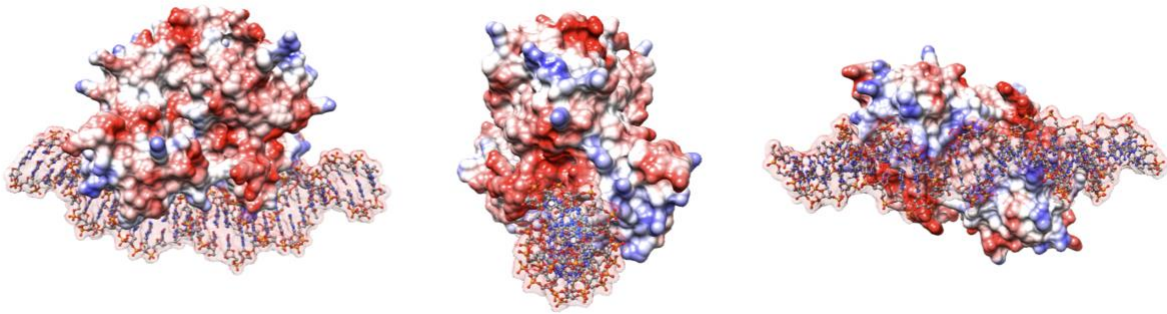
**Figura 12. Parámetros de "roll" y "twist" del ADN en complejo con SdrP y CAP.**

**A.** ADN "roll". **B.** ADN "twist". Se comparó los ángulos del ADN consenso de 30 y 38 pb en complejo con SdrP con lo reportado para CAP con la secuencia de ADN consenso (PDB: 1O3T, 1ZRC y 1O3Q). En negritas se encuentra el nucleótido donde se realiza el doblez principal. Los ángulos "roll" y "twist" del ADN con SdrP y CAP fueron determinados con x3ADN. Las secuencias de ADN y colores se muestran en la Figura.

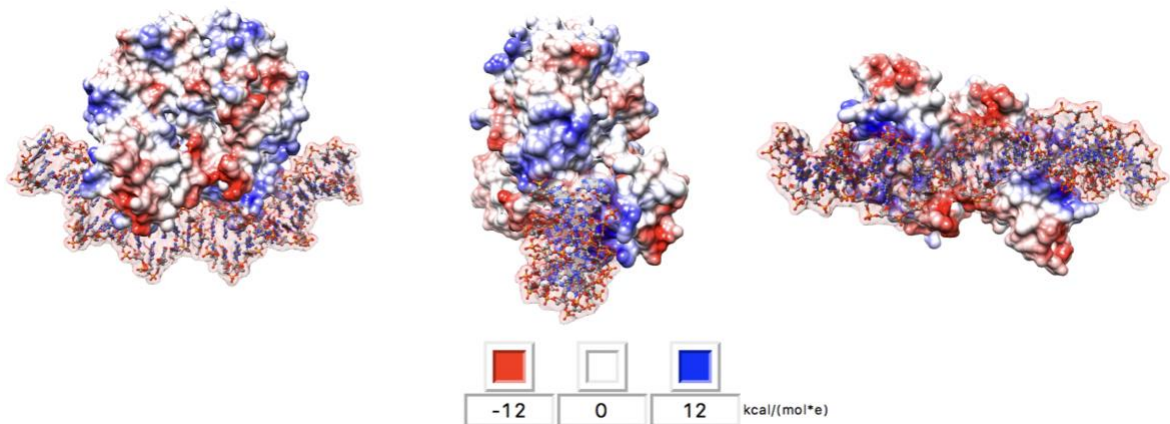
unión al ADN, similar al reportado para CAP con la secuencia consenso, de aproximadamente  $20^\circ$  en la misma posición [58]. Considerando estos dos ángulos, podemos observar que SdrP genera un dobléz global más suave que el de CAP, pero el ADN tiene una simetría mayor que el de 1ZRC y 1O3T. Al superponer las cadenas de ADN del complejo medio A y B, determinamos un RMSD de 0.249 Å, 4.6 veces más similar que el del PDB 1ZRC (1.154 Å). El cambio observado en las interacciones indirectas, pueden ser resultado de las interacciones directas de la K165 con la 7G del medio sitio de unión al ADN como ya se mencionó anteriormente, esto conforme a lo reportado para CAP, donde al sustituir el E181 por aspártico, se sigue formando el puente de hidrógeno, es decir, se mantiene la interacción directa sin embargo, se elimina la interacción indirecta con la base 6T dando como resultado la eliminación del dobléz primario [72].

Otra interacción que se puede dar en el reconocimiento indirecto entre este tipo de complejos es por interacción electrostática. La superficie global de SdrP exhibe una distribución equilibrada de cargas positivas y negativas, en la Figura 13A se puede observar como la sección del C-terminal (H149-R172) tienen carga positiva (color azul) que le permite interaccionar con el ADN que tiene carga negativa (color rojo). En CAP (Fig.13B), se observa que presenta la sección del C-terminal con carga positiva en la región de unión al ADN (R180-N192) y además se ha reportado que realiza un segundo dobléz por medio de interacciones electrostáticas entre las cargas positivas de los aminoácidos de CAP (K22, K26, K44 y K166) y las cargas negativas de los fosfatos del ADN [74] (Figura 13B). Sin embargo, en el caso de SdrP, como podemos observar en la Figura 13A, en los lados del dímero la carga de los residuos de aminoácidos son negativos (G13-E151), lo que provoca una repulsión de las cargas negativas de los fosfatos del ADN, dando como resultado que el dobléz general del ADN sea menor. Esto se observó en 2006 al sustituir los aminoácidos cargados positivamente en CAP (K22, K26, K44 y K166) por residuos con carga neta neutra, se reportó que se redujo el ángulo de dobléz en solución de aproximadamente  $5^\circ$  por residuo por medio complejo [74].

### A. Complejo SdrP-ADN



### B. Complejo CAP-cAMP-ADN



**Figura 13. Superficie electrostática del complejo SdrP-ADN (A) y CAP-cAMP-ADN (B).** La estructura se muestra en tres orientaciones diferentes, relacionada por una rotación de 90° en el eje “x” y “y”, Rojo indica las regiones cargadas negativamente (-12kT/e) y azul indica las regiones cargadas positivamente (+12kT/e). La superficie sin carga es mostrada en blanco. La imagen fue generada usando la herramienta Coulombic Surface Coloring en Chimera.

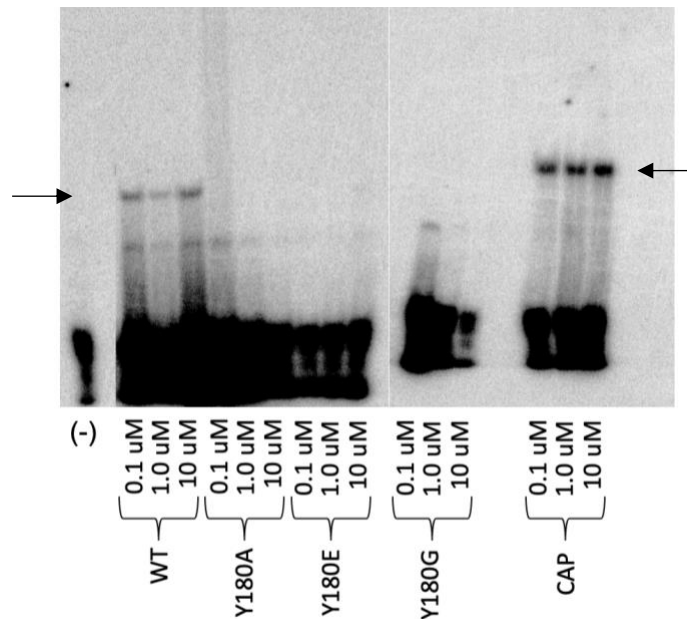
Nosotros concluimos que las interacciones directas de los aminoácidos de la hélice de reconocimiento así como las interacciones indirectas por cargas electrostáticas, provoca que tanto los ángulos “roll” en el sitio del pliegue primario así como la curva general sea diferente al reportado para el complejo de CAP-ADN [73,79].

### **2.3.3 Y180 es un residuo externo al motivo de reconocimiento (hélice-giro-hélice) que participa en el reconocimiento del ADN**

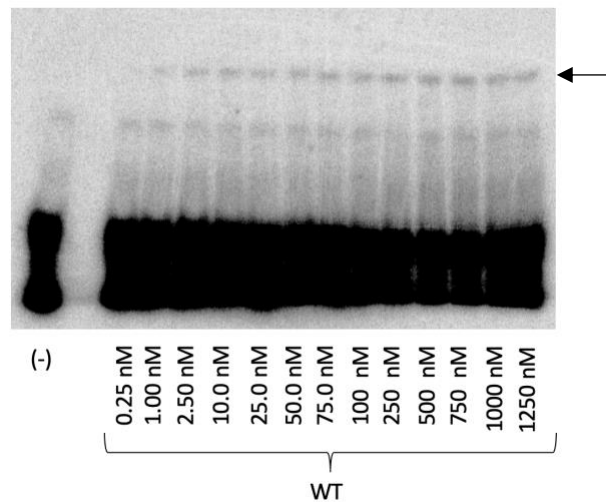
Como mencionamos previamente, en la estructura de SdrP en complejo con ADN, determinamos que existe un contacto directo con el surco menor del ADN a través de la Y180 localizada en el asa adyacente a la hélice de reconocimiento. Para verificar el papel de este residuo en la interacción, realizamos las mutantes Y180G Y180A y Y180E y analizamos la unión al ADN mediante EMSA. En la Figura 14 se observa la movilidad electroforética del complejo proteína-ADN, donde la proteína WT forma el complejo con el ADN a concentraciones de 0.1, 1.0 y 10  $\mu$ M de proteína, este complejo también se pudo observar con la proteína CAP-cAMP la cual usamos como control positivo. Sin embargo, al sustituir la Y180 por alanina, glicina y ácido glutámico no se observa la formación del complejo en las condiciones evaluadas. La alanina y glicina son los aminoácidos más pequeños, cuya cadena lateral es un grupo metilo y un átomo de hidrógeno, respectivamente y pertenecen al grupo de aminoácidos neutros, por lo que las mutantes Y180A y Y180G no forman el puente de hidrógeno con el ADN en la adenina de la posición 2 que observamos en la Y180. El ácido glutámico forma parte del grupo de aminoácidos polares cuya cadena lateral puede formar puentes de hidrógeno, sin embargo, tiene carga negativa, por lo que en la mutante Y180E se esperaba un impedimento estérico con las cargas negativas de los grupos fosfatos. Es interesante que a pesar de que se mantienen los aminoácidos R160, E161 y K165, presentes en la hélice de reconocimiento, la sustitución de la Y180 no forma el complejo con el ADN en las concentraciones de proteína utilizados.

Para determinar la constante de afinidad de SdrP WT por el ADN consenso, se realizó un EMSA con concentración constante de ADN y concentraciones crecientes

**A. EMSA SdrP WT, Y180A, Y180E, Y180G y CAP con ADN consenso**



**B. EMSA SdrP WT con ADN consenso**



**Figura 14. Ensayo de movilidad electroforética del complejo SdrP-ADN WT, Y180A, Y180E, Y180G.** A) Ensayo de EMSA con 1.0 nM de ADN consenso y concentraciones de 0.1, 1.0 y 10 mM de SdrP WT y mutantes. B) Ensayo de EMSA con 1.0 nM de ADN consenso y concentraciones crecientes de SdrP WT, se observa formación de complejo a partir de 1 nM de proteína. La flecha indica la formación del complejo proteína/ADN.

de proteína de 0.25 nM a 1250 nM. En la Figura 14 se observa el resultado de la movilidad electroforética para calcular la  $K_d$  y donde se observa que SdrP se une al ADN desde concentraciones muy bajas (1 nM), sin embargo, al hacer el análisis en Prism 9, los datos resultantes del EMSA fueron ambiguos y no se pudo determinar la  $K_d$ . En un artículo reciente, un grupo de trabajo determinó que la  $K_d$  de SdrP por la secuencia consenso es de 3.5 nM, una alta afinidad similar a la de otras proteínas de la familia CRP/FNR como CAP de *E. coli* con una  $K_d$  de 1.0 nM, [80]. Este valor nos sugiere que en el ensayo que realizamos era necesario aumentar los puntos de la curva de concentración de proteína en valores por debajo de 1 nM y no manejar valores tan altos de proteína como los que se emplearon.

Uno de los hallazgos más relevantes de nuestro trabajo, fue el papel de la Tyr180 de SdrP en la unión al ADN. Hasta la fecha, SdrP es el primer factor de transcripción de la familia CRP/FNR en el que se ha reportado un aminoácido fuera de la hélice de reconocimiento que tiene contacto directo con el ADN. Sin embargo, hay reportes de proteínas que realizan contacto con el ADN que incluyen tirosinas, como el regulador transcripcional FatR de *Bacillus subtilis* (*B. Subtillis*), que pertenece a la familia TetR y que contiene el motivo Helix-Turn-Helix [81]. FatR tiene una Tirosina en la posición 45 que se encuentra altamente conservada y es fosforilada por la kinasa PtkA, esta fosforilación interrumpe la interacción de la proteína con el ADN y se considera un mecanismo general para la inactivación de los reguladores que desplazan ácidos grasos en bacterias [81]. En el caso de *T. thermophilus* HB8, se ha reportado un análisis fosfoproteómico donde identificaron 48 proteínas que son fosforiladas: 30 en serina, 12 en treonina y 4 en tirosina, no obstante el estudio se realizó durante la fase logarítmica de crecimiento y no se identificó a la proteína SdrP [82]. La adición local de una carga negativa mediante una fosforilación, puede afectar negativa o positivamente la unión a ADN de los factores de transcripción a sus secuencias regulatorias, participando de esta manera en la regulación de la respuesta celular[83]. Queda pendiente determinar si la proteína SdrP es fosforilada por las quinasas de *T. thermophilus* como un mecanismo de regulación de este factor de transcripción.

## **2.4 Conclusiones y perspectivas**

Hemos determinado la estructura de SdrP en complejo con ADN e identificamos los residuos de aminoácidos que hacen contacto directo con este. Determinamos que los residuos de aminoácidos R160, E161 y K165 hacen contacto directo con los nucleótidos del surco mayor y la Y180, un residuo externo al motivo de reconocimiento, con el surco menor; interesantemente, esta interacción no se ha reportado para otro miembro de la familia CRP/FNR. El análisis de los parámetros generales del ADN en complejo con SdrP mostró que la curvatura del ADN presenta un ángulo de doblez menor al reportado para CAP esto como resultado de las interacciones directas, indirectas y por cargas electrostáticas. Como perspectivas de este trabajo, se propone determinar la posible fosforilación del residuo Y180 como mecanismo de regulación de SdrP.



## Referencias

- [1] X. Du, Y. Li, Y.L. Xia, S.M. Ai, J. Liang, P. Sang, X.L. Ji, S.Q. Liu, Insights into protein–ligand interactions: Mechanisms, models, and methods, *Int. J. Mol. Sci.* 17 (2016) 1–34. <https://doi.org/10.3390/ijms17020144>.
- [2] J. Reinders, A. Sickmann, *Modificomics: Posttranslational modifications beyond protein phosphorylation and glycosylation*, *Biomol. Eng.* 24 (2007) 169–177. <https://doi.org/10.1016/j.bioeng.2007.03.002>.
- [3] M. Mann, O.N. Jensen, Proteomic analysis of post-translational modifications, *Nat. Biotechnol.* 21 (2003) 255–261. <https://doi.org/10.1038/nbt0303-255>.
- [4] O.N. Jensen, Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry, *Curr. Opin. Chem. Biol.* 8 (2004) 33–41. <https://doi.org/10.1016/j.cbpa.2003.12.009>.
- [5] Y. Shi, Y. Zhang, S. Lin, C. Wang, J. Zhou, D. Peng, Y. Xue, dbPSP 2.0, an updated database of protein phosphorylation sites in prokaryotes, *Sci. Data.* 7 (2020) 1–9. <https://doi.org/10.1038/s41597-020-0506-7>.
- [6] A. Lass, R. Zimmermann, M. Oberer, R. Zechner, Lipolysis – A highly regulated multi-enzyme complex mediates the catabolism of cellular fat stores, *Prog. Lipid Res.* 50 (2011) 14–27. <https://doi.org/10.1016/j.plipres.2010.10.004>.
- [7] D. Sun, X. Zhou, C. Liu, J. Zhu, Y. Ru, W. Liu, J. Liu, Fnr Negatively Regulates Prodigiosin Synthesis in *Serratia* sp. ATCC 39006 During Aerobic Fermentation, *Front. Microbiol.* 12 (2021) 1–13. <https://doi.org/10.3389/fmicb.2021.734854>.
- [8] C.-H. Lai, C.-Y. Chou, L.-Y. Ch'ang, C.-S. Liu, Identification of Novel Human Genes Evolutionarily Conserved in *Caenorhabditis elegans* by Comparative Proteomics, *Genome Res.* 10 (2000) 703–713. <https://doi.org/10.1101/gr.10.5.703>.
- [9] C. Lefèvre, F. Jobard, F. Caux, B. Bouadjar, A. Karaduman, R. Heilig, H. Lakhdar, A. Wollenberg, J.-L. Verret, J. Weissenbach, M. Özgüç, M. Lathrop, J.-F. Prud'homme, J. Fischer, Mutations in CGI-58, the Gene Encoding a

- New Protein of the Esterase/Lipase/Thioesterase Subfamily, in Chanarin-Dorfman Syndrome, *Am. J. Hum. Genet.* 69 (2001) 1002–1012.  
<https://doi.org/10.1086/324121>.
- [10] N. Kulminskaya, M. Oberer, Protein-protein interactions regulate the activity of Adipose Triglyceride Lipase in intracellular lipolysis, *Biochimie.* 169 (2020) 62–68. <https://doi.org/10.1016/j.biochi.2019.08.004>.
- [11] L. Yu, Y. Li, A. Gris , H. Wang, CGI-58: Versatile Regulator of Intracellular Lipid Droplet Homeostasis, in: X.-C. Jiang (Ed.), *Lipid Transf. Lipoprotein Metab. Cardiovasc. Dis.*, Springer Nature, Singapore, 2020: pp. 197–222. [https://doi.org/10.1007/978-981-15-6082-8\\_13](https://doi.org/10.1007/978-981-15-6082-8_13).
- [12] K.A. Zierler, R. Zechner, G. Haemmerle, Comparative gene identification-58/ $\alpha/\beta$  hydrolase domain 5: more than just an adipose triglyceride lipase activator?, *Curr. Opin. Lipidol.* 25 (2014) 102–109. <https://doi.org/10.1097/MOL.000000000000058>.
- [13] A. Gruber, I. Cornaciu, A. Lass, M. Schweiger, M. Poeschl, C. Eder, M. Kumari, G. Schoiswohl, H. Wolinski, S.D. Kohlwein, R. Zechner, R. Zimmermann, M. Oberer, The N-terminal Region of Comparative Gene Identification-58 (CGI-58) Is Important for Lipid Droplet Binding and Activation of Adipose Triglyceride Lipase, *J. Biol. Chem.* 285 (2010) 12289–12298. <https://doi.org/10.1074/jbc.M109.064469>.
- [14] M.A. Sanders, H. Zhang, L. Mladenovic, Y.Y. Tseng, J.G. Granneman, Molecular Basis of ABHD5 Lipolysis Activation, *Sci. Rep.* 7 (2017) 42589. <https://doi.org/10.1038/srep42589>.
- [15] T.O. Eichmann, M. Kumari, J.T. Haas, R. V Farese, R. Zimmermann, A. Lass, R. Zechner, Studies on the Substrate and Stereo/Regioselectivity of Adipose Triglyceride Lipase, Hormone-sensitive Lipase, and Diacylglycerol-O-acyltransferases, *J. Biol. Chem.* 287 (2012) 41446–41457. <https://doi.org/10.1074/jbc.M112.400416>.
- [16] V. Subramanian, A. Rothenberg, C. Gomez, A.W. Cohen, A. Garcia, S. Bhattacharyya, L. Shapiro, G. Dolios, R. Wang, M.P. Lisanti, D.L. Brasaemle, Perilipin A Mediates the Reversible Binding of CGI-58 to Lipid Droplets in

- 3T3-L1 Adipocytes, *J. Biol. Chem.* 279 (2004) 42062–42071.  
<https://doi.org/10.1074/jbc.M407462200>.
- [17] A. Boeszoermenyi, H.M. Nagy, H. Arthanari, C.J. Phillip, H. Lindermuth, R.E. Luna, G. Wagner, R. Zechner, K. Zangger, M. Oberer, Structure of a CGI-58 Motif Provides the Molecular Basis of Lipid Droplet Anchoring, *J. Biol. Chem.* 290 (2015) 26361–26372. <https://doi.org/10.1074/jbc.M115.682203>.
- [18] J.G. Granneman, H.H. Moore, R. Krishnamoorthy, M. Rathod, Perilipin Controls Lipolysis by Regulating the Interactions of AB-hydrolase Containing 5 (Abhd5) and Adipose Triglyceride Lipase (Atgl), *J. Biol. Chem.* 284 (2009) 34538–34544. <https://doi.org/10.1074/jbc.M109.068478>.
- [19] A. Sahu-Osen, G. Montero-Moran, M. Schittmayer, K. Fritz, A. Dinh, Y.-F. Chang, D. McMahon, A. Boeszoermenyi, I. Cornaciu, D. Russell, M. Oberer, G.M. Carman, R. Birner-Gruenberger, D.L. Brasaemle, CGI-58/ABHD5 is phosphorylated on Ser239 by protein kinase A: control of subcellular localization, *J. Lipid Res.* 56 (2015) 109–121.  
<https://doi.org/10.1194/jlr.M055004>.
- [20] M.A. Sanders, F. Madoux, L. Mladenovic, H. Zhang, X. Ye, M. Angrish, E.P. Mottillo, J.A. Caruso, G. Halvorsen, W.R. Roush, P. Chase, P. Hodder, J.G. Granneman, Endogenous and Synthetic ABHD5 Ligands Regulate ABHD5-Perilipin Interactions and Lipolysis in Fat and Muscle, *Cell Metab.* 22 (2015) 851–860. <https://doi.org/10.1016/j.cmet.2015.08.023>.
- [21] C. Notredame, D.G. Higgins, J. Heringa, T-coffee: a novel method for fast and accurate multiple sequence alignment, *J. Mol. Biol.* 302 (2000) 205–217.  
<https://doi.org/10.1006/jmbi.2000.4042>.
- [22] A. Drozdetskiy, C. Cole, J. Procter, G.J. Barton, JPred4: a protein secondary structure prediction server, *Nucleic Acids Res.* 43 (2015) W389–W394.  
<https://doi.org/10.1093/nar/gkv332>.
- [23] G. Yachdav, E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Hönigschmid, A. Schafferhans, M. Roos, M. Bernhofer, L. Richter, H. Ashkenazy, M. Punta, A. Schlessinger, Y. Bromberg, R. Schneider, G. Vriend, C. Sander, N. Ben-Tal, B. Rost, PredictProtein—an open resource

- for online prediction of protein structural and functional features, *Nucleic Acids Res.* 42 (2014) W337–W343. <https://doi.org/10.1093/nar/gku366>.
- [24] O. Conchillo-Solé, N.S. de Groot, F.X. Avilés, J. Vendrell, X. Daura, S. Ventura, AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides, *BMC Bioinformatics.* 8 (2007) 65. <https://doi.org/10.1186/1471-2105-8-65>.
- [25] P.H. Lee, X.X. Huang, B.T. Teh, L.M. Ng, TSA-CRAFT: A Free Software for Automatic and Robust Thermal Shift Assay Data Analysis, *SLAS Discov.* 24 (2019) 606–612. <https://doi.org/10.1177/2472555218823547>.
- [26] K.A. Bakar, S.R. Feroz, A critical view on the analysis of fluorescence quenching data for determining ligand–protein binding affinity, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 223 (2019) 117337. <https://doi.org/10.1016/j.saa.2019.117337>.
- [27] I. Jarmoskaite, I. Alsadhan, P.P. Vaidyanathan, D. Herschlag, How to measure and evaluate binding affinities, *Elife.* 9 (2020) 1–34. <https://doi.org/10.7554/ELIFE.57264>.
- [28] N.J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, *Nat. Protoc.* 1 (2007) 2876–2890. <https://doi.org/10.1038/nprot.2006.202>.
- [29] A. Micsonai, F. Wien, É. Bulyáki, J. Kun, É. Moussong, Y.-H. Lee, Y. Goto, M. Réfrégiers, J. Kardos, BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra, *Nucleic Acids Res.* 46 (2018) W315–W322. <https://doi.org/10.1093/nar/gky497>.
- [30] C. Park, S. Marqusee, Analysis of the stability of multimeric proteins by effective  $\Delta G$  and effective  $m$ -values, *Protein Sci.* 13 (2004) 2553–2558. <https://doi.org/10.1110/ps.04811004>.
- [31] S.E. Jackson, A.R. Fersht, Folding of Chymotrypsin Inhibitor 2.1. Evidence for a Two-State Transition, *Biochemistry.* 30 (1991) 10428–10435. <https://doi.org/10.1021/bi00107a010>.
- [32] J. Yang, Y. Zhang, I-TASSER server: new development for protein structure

- and function predictions, *Nucleic Acids Res.* 43 (2015) W174–W181.  
<https://doi.org/10.1093/nar/gkv342>.
- [33] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera - A visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612.  
<https://doi.org/10.1002/jcc.20084>.
- [34] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66 (2010) 486–501.  
<https://doi.org/10.1107/S0907444910007493>.
- [35] T.J. Dolinsky, P. Czodrowski, H. Li, J.E. Nielsen, J.H. Jensen, G. Klebe, N.A. Baker, PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations, *Nucleic Acids Res.* 35 (2007) 522–525. <https://doi.org/10.1093/nar/gkm276>.
- [36] N. Sánchez de Groot, I. Pallarés, F.X. Avilés, J. Vendrell, S. Ventura, Prediction of “hot spots” of aggregation in disease-linked polypeptides, *BMC Struct. Biol.* 5 (2005) 18. <https://doi.org/10.1186/1472-6807-5-18>.
- [37] R. Graña-Montes, R. Sant’Anna de Oliveira, S. Ventura, Protein aggregation profile of the human kinome, *Front. Physiol.* 3 (2012) 438.  
<https://doi.org/10.3389/fphys.2012.00438>.
- [38] P. Hofer, A. Boeszoermenyi, D. Jaeger, U. Feiler, H. Arthanari, N. Mayer, F. Zehender, G. Rechberger, M. Oberer, R. Zimmermann, A. Lass, G. Haemmerle, R. Breinbauer, R. Zechner, K. Preiss-Landl, Fatty Acid-binding Proteins Interact with Comparative Gene Identification-58 Linking Lipolysis with Lipid Ligand Shuttling, *J. Biol. Chem.* 290 (2015) 18438–18453.  
<https://doi.org/10.1074/jbc.M114.628958>.
- [39] K. Phillips, A. Hernandez de la Peña, The Combined Use of the Thermofluor Assay and ThermoQ Analytical Software for the Determination of Protein Stability and Buffer Optimization as an Aid in Protein Crystallization, *Curr. Protoc. Mol. Biol.* 94 (2011) 10.28.1-10.28.15.  
<https://doi.org/10.1002/0471142727.mb1028s94>.
- [40] P. Evans, K. Wyatt, G.J. Wistow, O.A. Bateman, B.A. Wallace, C. Slingsby,

- The P23T cataract mutation causes loss of solubility of folded  $\gamma$ D-crystallin, *J. Mol. Biol.* 343 (2004) 435–444. <https://doi.org/10.1016/j.jmb.2004.08.050>.
- [41] A.P. Golovanov, G.M. Hautbergue, S.A. Wilson, L.Y. Lian, A simple method for improving protein solubility and long-term stability, *J. Am. Chem. Soc.* 126 (2004) 8933–8939. <https://doi.org/10.1021/ja049297h>.
- [42] L. Reinhard, A. Geerlof, J. Mueller-dieckmann, M.S. Weiss, G.W. Vej, F.-G. Cedex, Optimization of protein buffer cocktails using Thermofluor, *Acta Crystallogr. Sect. F Structural Biol. Commun.* F69 (2013) 209–214. <https://doi.org/10.1107/S1744309112051858>.
- [43] J.X. Huang, G. Lee, K.E. Cavanaugh, J.W. Chang, M.L. Gardel, R.E. Moellering, High throughput discovery of functional protein modifications by Hotspot Thermal Profiling, *Nat. Methods.* 16 (2019) 894–901. <https://doi.org/10.1038/s41592-019-0499-3>.
- [44] L. Selzer, R. Kant, J.C.Y. Wang, B. Bothner, A. Zlotnick, Hepatitis B virus core protein phosphorylation sites affect capsid stability and transient exposure of the C-terminal domain, *J. Biol. Chem.* 290 (2015) 28584–28593. <https://doi.org/10.1074/jbc.M115.678441>.
- [45] E.G. Worrall, L. Worrall, E. Blackburn, M. Walkinshaw, T.R. Hupp, The Effects of Phosphomimetic Lid Mutation on the Thermostability of the N-terminal Domain of MDM2, *J. Mol. Biol.* 398 (2010) 414–428. <https://doi.org/10.1016/j.jmb.2010.03.023>.
- [46] Z.H. Jebessa, K.D. Shanmukha, M. Dewenter, L.H. Lehmann, C. Xu, F. Schreiter, D. Siede, X. Gong, B.C. Worst, G. Federico, S.W. Sauer, T. Fischer, L. Wechselberger, O.J. Müller, S. Sossalla, C. Dieterich, P. Most, H.-J. Gröne, C. Moro, M. Oberer, G. Haemmerle, H.A. Katus, J. Tyedmers, J. Backs, The lipid-droplet-associated protein ABHD5 protects the heart through proteolysis of HDAC4, *Nat. Metab.* 1 (2019) 1157–1167. <https://doi.org/10.1038/s42255-019-0138-4>.
- [47] C.C. Lord, G. Thomas, J.M. Brown, Mammalian alpha beta hydrolase domain (ABHD) proteins: Lipid metabolizing enzymes at the interface of cell signaling and energy metabolism, *Biochim. Biophys. Acta - Mol. Cell Biol.*

- Lipids. 1831 (2013) 792–802. <https://doi.org/10.1016/j.bbalip.2013.01.002>.
- [48] A.P. Yamniuk, H. Ishida, D. Lippert, H.J. Vogel, Thermodynamic effects of noncoded and coded methionine substitutions in calmodulin, *Biophys. J.* 96 (2009) 1495–1507. <https://doi.org/10.1016/j.bpj.2008.10.060>.
- [49] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism, *Biochim. Biophys. Acta - Proteins Proteomics.* 1751 (2005) 119–139. <https://doi.org/10.1016/j.bbapap.2005.06.005>.
- [50] G. Montero-Moran, J.M. Caviglia, D. McMahon, A. Rothenberg, V. Subramanian, Z. Xu, S. Lara-Gonzalez, J. Storch, G.M. Carman, D.L. Brasaemle, CGI-58/ABHD5 is a coenzyme A-dependent lysophosphatidic acid acyltransferase, *J. Lipid Res.* 51 (2010) 709–719. <https://doi.org/10.1194/jlr.M001917>.
- [51] P. Sindrewicz, X. Li, E.A. Yates, J.E. Turnbull, L.Y. Lian, L.G. Yu, Intrinsic tryptophan fluorescence spectroscopy reliably determines galectin-ligand interactions, *Sci. Rep.* 9 (2019) 1–12. <https://doi.org/10.1038/s41598-019-47658-8>.
- [52] N. Rivera-Gómez, S. Lorenzo, E. Pérez-Rueda, Diversity and distribution of transcription factors : their partner domains play an important role in regulatory plasticity in bacteria, *Microbiology.* 157 (2011) 2308–2318. <https://doi.org/10.1099/mic.0.050617-0>.
- [53] H. Körner, H.J. Sofia, W.G. Zumft, Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: Exploiting the metabolic spectrum by controlling alternative gene programs, *FEMS Microbiol. Rev.* 27 (2003) 559–592. [https://doi.org/10.1016/S0168-6445\(03\)00066-4](https://doi.org/10.1016/S0168-6445(03)00066-4).
- [54] Y. Agari, A. Kashihara, S. Yokoyama, S. Kuramitsu, A. Shinkai, Global gene expression mediated by *Thermus thermophilus* SdrP, a CRP/FNR family transcriptional regulator, *Mol. Microbiol.* 70 (2008) 60–75. <https://doi.org/10.1111/j.1365-2958.2008.06388.x>.
- [55] A. Shinkai, S. Kira, N. Nakagawa, A. Kashihara, S. Kuramitsu, S. Yokoyama, Transcription Activation Mediated by a Cyclic AMP Receptor Protein from *Thermus thermophilus* HB8 □ §, 189 (2007) 3891–3901.

<https://doi.org/10.1128/JB.01739-06>.

- [56] J.M. Passner, S.C. Schultz, T.A. Steitz, Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Å resolution, *J. Mol. Biol.* 304 (2000) 847–859.  
<https://doi.org/10.1006/jmbi.2000.4231>.
- [57] H. Sharma, S. Yu, J. Kong, J. Wang, T. a Steitz, Structure of apo-CAP reveals that large conformational changes are necessary for DNA binding., *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 16604–16609.  
<https://doi.org/10.1073/pnas.0908380106>.
- [58] S. Chen, J. Vojtechovsky, G.N. Parkinson, R.H. Ebright, H.M. Berman, Indirect Readout of DNA Sequence at the Primary-kink Site in the CAP-DNA Complex : DNA Binding Specificity Based on Energetics of DNA Kinking, (2001). <https://doi.org/10.1006/jmbi.2001.5089>.
- [59] N. Popovych, S.-R. Tzeng, M. Tonelli, R.H. Ebright, C.G. Kalodimos, Structural basis for cAMP-mediated allosteric control of the catabolite activator protein., *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 6927–6932.  
<https://doi.org/10.1073/pnas.0900595106>.
- [60] Y. Agari, S. Kuramitsu, A. Shinkai, X-ray crystal structure of TTHB099, a CRP/ FNR superfamily transcriptional regulator from *Thermus thermophilus* HB8, reveals a DNA-binding protein with no required allosteric effector molecule, *Proteins.* (2012) 1490–1494. <https://doi.org/10.1002/prot.24049>.
- [61] H. Takano, M. Kondo, N. Usui, T. Usui, H. Ohzeki, R. Yamazaki, M. Washioka, A. Nakamura, T. Hoshino, W. Hakamata, T. Beppu, K. Ueda, Involvement of CarA / LitR and CRP / FNR Family Transcriptional Regulators in Light-Induced Carotenoid Production in, 193 (2011) 2451–2459.  
<https://doi.org/10.1128/JB.01125-10>.
- [62] Y. Feng, Y. Zhang, R.H. Ebright, Structural basis of transcription activation, 352 (2016).
- [63] Y. Agari, S. Kuramitsu, A. Shinkai, Identification of novel genes regulated by the oxidative stress-responsive transcriptional activator SdrP in *Thermus thermophilus* HB8, *FEMS.* (2010) 127–134. <https://doi.org/10.1111/j.1574->



6968.2010.02133.x.

- [64] G. Soberón-Chávez, L.D. Alcaraz, E. Morales, G.Y. Ponce-Soto, L. Servín-González, The transcriptional regulators of the CRP family regulate different essential bacterial functions and can be inherited vertically and horizontally, *Front. Microbiol.* 8 (2017) 1–8. <https://doi.org/10.3389/fmicb.2017.00959>.
- [65] M.A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees, 2010 Gatew. Comput. Environ. Work. GCE 2010. (2010). <https://doi.org/10.1109/GCE.2010.5676129>.
- [66] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, *Nucleic Acids Res.* 44 (2016) W344–W350. <https://doi.org/10.1093/NAR/GKW408>.
- [67] R.H. Ebright, Y.W. Ebright, A. Gunasekera, N. Brunswick, *Nucleic Acids Research*, 17 (1989) 16–18.
- [68] X.J. Lu, W.K. Olson, 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, *Nucleic Acids Res.* 31 (2003) 5108–5121. <https://doi.org/10.1093/nar/gkg680>.
- [69] X.J. Lu, W.K. Olson, 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures, *Nat. Protoc.* 3 (2008) 1213–1227. <https://doi.org/10.1038/nprot.2008.104>.
- [70] P.L. García-Medel, N. Baruch-Torres, A. Peralta-Castro, C.H. Trasviña-Arenas, A. Torres-Larios, L.G. Brieba, Plant organellar DNA polymerases repair double-stranded breaks by microhomology-mediated end-joining, *Nucleic Acids Res.* 47 (2019) 3028–3044. <https://doi.org/10.1093/nar/gkz039>.
- [71] O. Carugo, S. Pongor, A normalized root-mean-square distance for comparing protein three-dimensional structures, *Protein Sci.* 10 (2008) 1470–1473. <https://doi.org/10.1110/ps.690101>.
- [72] S. Chen, A. Gunasekera, X. Zhang, T.A. Kunkel, R.H. Ebright, H.M. Berman, Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA

- complex: alteration of DNA binding specificity through alteration of DNA kinking, *J. Mol. Biol.* 314 (2001) 75–82.  
<https://doi.org/10.1006/jmbi.2001.5090>.
- [73] A.A. Napoli, C.L. Lawson, R.H. Ebright, H.M. Berman, Indirect Readout of DNA Sequence at the Primary-kink Site in the CAP – DNA Complex : Recognition of Pyrimidine-Purine and Purine-Purine Steps, (2006) 173–183.  
<https://doi.org/10.1016/j.jmb.2005.12.051>.
- [74] G.B. Koudelka, S.A. Mauro, M. Ciubotaru, Indirect Readout of DNA Sequence by Proteins: The Roles of DNA Sequence-Dependent Intrinsic and Extrinsic Forces, in: *Prog. Nucleic Acid Res.*, 2006: pp. 143–177.  
[https://doi.org/10.1016/S0079-6603\(06\)81004-4](https://doi.org/10.1016/S0079-6603(06)81004-4).
- [75] S.C. Schultz, G.C. Shields, T.A. Steitz, S.C. Schultz, G.C. Shields, T.A. Steitz, Crystal Structure of a CAP-DNA Complex : The DNA Is Bent by 90°, 253 (1991) 1001–1007.
- [76] G. Parkinson, C. Wilson, A. Gunasekera, Y.W. Ebright, R.E. Ebright, H.M. Berman, Structure of the CAP-DNA Complex at 2.5 Å Resolution : A Complete Picture of the Protein-DNA Interface, 1 (1996) 395–408.
- [77] C.L. Lawson, D. Swigon, K.S. Murakami, S.A. Darst, H.M. Berman, R.H. Ebright, Catabolite activator protein: DNA binding and transcription activation, *Curr. Opin. Struct. Biol.* 14 (2004) 1–11.  
<https://doi.org/10.1016/j.sbi.2004.01.012>.
- [78] C.T. Armstrong, P.E. Mason, J.L.R. Anderson, C.E. Dempsey, Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels, *Sci. Rep.* 6 (2016) 1–10.  
<https://doi.org/10.1038/srep21759>.
- [79] Y. Wang, J. Chen, W. Tang, Y. Zhang, X. Li, Rapamycin inhibits the proliferation of endothelial cells in hemangioma by blocking the mTOR-FABP4 pathway, *Biomed. Pharmacother.* 85 (2017) 272–279.  
<https://doi.org/10.1016/j.biopha.2016.11.021>.
- [80] J.L. Teague, J.K. Barrows, C.A. Baafi, M.W. Van Dyke, Discovering the dna-binding consensus of the thermus thermophilus hb8 transcriptional regulator

- ttha1359, *Int. J. Mol. Sci.* 22 (2021). <https://doi.org/10.3390/ijms221810042>.
- [81] A. Derouiche, V. Bidnenko, R. Grenha, N. Pignonneau, Interaction of bacterial fatty-acid-displaced regulators with DNA is interrupted by tyrosine phosphorylation in the helix-turn-helix domain, *41* (2013) 9371–9381. <https://doi.org/10.1093/nar/gkt709>.
- [82] Y. Takahata, M. Inoue, K. Kim, Y. Iio, M. Miyamoto, R. Masui, Y. Ishihama, S. Kuramitsu, Close proximity of phosphorylation sites to ligand in the phosphoproteome of the extreme thermophile *Thermus thermophilus* HB8, (2012) 1414–1430. <https://doi.org/10.1002/pmic.201100573>.
- [83] T. Garcia-Garcia, S. Poncet, A. Derouiche, L. Shi, I. Mijakovic, M.F. Noirot-Gros, Role of protein phosphorylation in the regulation of cell cycle and DNA-related processes in bacteria, *Front. Microbiol.* 7 (2016) 1–11. <https://doi.org/10.3389/fmicb.2016.00184>.

**Artículos publicados**

**Miriam Livier Llamas-García**, Edgar D. Páez-Pérez, Claudia G. Benítez-Cardoza, Gabriela M. Montero-Morán, and Samuel Lara-González. 2022. “Improved Stability of Human CGI-58 Induced by Phosphomimetic S237E Mutation”. ACS Omega, Vol. 7: 12643–12653. DOI: 10.1021/acsomega.1c06872.

Edgar D. Páez-Pérez, **Miriam Livier Llamas-García**, Claudia G. Benítez-Cardoza, Gabriela M. Montero-Morán, and Samuel Lara-González. 2020. “**Bioinformatic Analysis and Biophysical Characterization Reveal Structural Disorder in G0S2 Protein.**” ACS Omega. Vol. 5: 25841-25847 DOI: 10.1021/acsomega.0c03171.