



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

**POSGRADO EN CIENCIAS EN BIOLOGIA MOLECULAR**

**Construcción de secuencias en tándem para  
análisis de interacción proteína-ADN en  
herramientas de molécula única**

Tesis que presenta

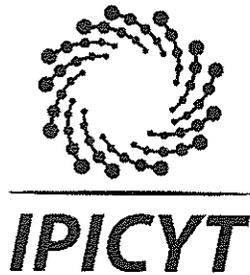
**L.B.G. Paulina Janeth Cázares Samaniego**

Para obtener el grado de

**Maestra en Ciencias en Biología Molecular**

**Director de la Tesis:  
Dr. Samuel Lara González**

San Luis Potosí, S.L.P., Septiembre de 2017



## Constancia de aprobación de la tesis

La tesis ***“Construcción de secuencias en tándem para análisis de interacción proteína-ADN en herramientas de molécula única”*** presentada para obtener el Grado de Maestra en Ciencias en Biología Molecular fue elaborada por Paulina Janeth Cázares Samaniego y aprobada el **quince de septiembre de dos mil diecisiete** 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 de la tesis

**Dr. Guillermo Pastor Palacios**  
Miembro del Comité Tutorial

**Dr. Braulio Gutiérrez Medina**  
Miembro del Comité Tutorial

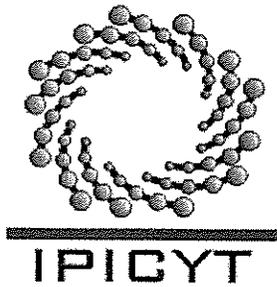


### **Créditos Institucionales**

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 (No. de registro 590396) 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. Parte del equipo utilizado fue adquirido con el proyecto CONACyT INFRA No. 204373.



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

## Acta de Examen de Grado

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 170 del Libro Primero de Actas de Exámenes de Grado del Programa de Maestría en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 15 días del mes de septiembre del año 2017, se reunió a las 10:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

<b>Dr. Braulio Gutiérrez Medina</b>	<b>Presidente</b>	<b>IPICYT</b>
<b>Dr. Guillermo Pastor Palacios</b>	<b>Secretario</b>	<b>IPICYT</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:

**MAESTRA EN CIENCIAS EN BIOLOGÍA MOLECULAR**

sustentó la C.

**Paulina Janeth Cázares Samaniego**

sobre la Tesis intitulada:

*Construcción de secuencias en tándem para análisis de interacción proteína-ADN en herramientas de molécula única*

que se desarrolló bajo la dirección de

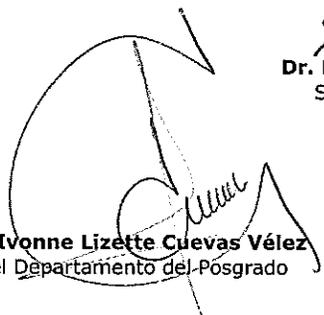
**Dr. Samuel Lara González**

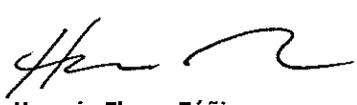
El Jurado, después de deliberar, determinó

**APROBARLA**

Dándose por terminado el acto a las 11:40 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico 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 15 días del mes de septiembre de 2017.

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

  
**Dr. Horacio Flores Zúñiga**  
Secretario Académico



## **Dedicatorias**

Me gustaría que estas líneas sirvieran para expresar mi más profundo y sincero agradecimiento a todas aquellas personas que con su ayuda han colaborado en la realización del presente trabajo.

En especial, por la comprensión, paciencia y el ánimo recibidos de mi familia, amigos y Daniel.

A todos ellos, muchas gracias.

## **Agradecimientos**

Al Consejo Nacional de Ciencia y Tecnología (CONACYT).

Al Instituto Potosino de Investigación Científica y Tecnológica (IPICYT).

Al Dr. Samuel Lara González por la dirección y apoyo en este trabajo.

A los miembros de mi comité tutorial, Dr. Braulio Gutiérrez Medina y Dr. Guillermo Pastor Palacios, por las aportaciones en la tesis.

## Contenido

Constancia de aprobación de la tesis	ii
Créditos institucionales	iii
Acta de examen	iv
Dedicatorias	v
Agradecimientos	vi
Lista de tablas	viii
Lista de figuras	ix
Anexos	x
Resumen	xi
Abstract	xii
<b>INTRODUCCIÓN</b>	<b>1</b>
<b>MATERIALES Y METODOS</b>	
<i>Materiales</i>	8
<i>Construcción del vector en tándem pBlueScript/CONCAP*2</i>	8
<i>Construcción del vector en tándem pBlueScript/LacNat*3</i>	9
<i>Expresión y purificación de la proteína CAP</i>	10
<i>Expresión y purificación de la proteína SdrP</i>	11
<i>Dispersión dinámica de luz (DLS)</i>	11
<i>Ensayo de cambio en la movilidad electroforética del ADN(EMSA)</i>	12
<i>Análisis bioinformático de ortólogos putativos de la proteína Clp</i>	12
<b>RESULTADOS</b>	
<i>Construcción del vector en tándem pBlueScript/CONCAP*2</i>	14
<i>Construcción del vector en tándem pBlueScript/LacNat*3</i>	14
<i>Expresión y purificación de la proteína CAP</i>	18
<i>Expresión y purificación de la proteína SdrP</i>	18
<i>Interacción de CAP y SdrP con los secuencias LacNat y CONOCAP por dispersión dinámica de luz (DLS)</i>	22
<i>Ensayo de cambio en la movilidad electroforética del ADN(EMSA)</i>	26
<i>Análisis bioinformático de ortólogos putativos de la proteína Clp</i>	30
<b>DISCUSIÓN</b>	<b>31</b>
<b>REFERENCIAS</b>	<b>35</b>

## Lista de tablas

<b>Tabla 1.</b> Oligonucleótidos empleados en la construcción de los vectores.	38
<b>Tabla 2.</b> Concentración total de proteína empleada en EMSA	39
<b>Tabla 3.</b> Análisis de las secuencias seleccionadas.	40

## Lista de figuras

<b>Figura 1.</b> Torsión intrínseca del ADN LacNat y CONCAP.	4
<b>Figura 2.</b> Análisis comparativo de las proteínas CAP, Clp y SdrP	7
<b>Figura 3.</b> Construcción del vector pET28/CONCAP	15
<b>Figura 4.</b> Construcción del vector pBlueScript/CONCAP*2	16
<b>Figura 5.</b> Construcción del vector pET28/LacNat	17
<b>Figura 6.</b> Construcción del vector pBlueScript/LacNat*3	19
<b>Figura 7.</b> Expresión y purificación de la proteína CAP	20
<b>Figura 8.</b> Expresión y purificación de la proteína SdrP	21
<b>Figura 9.</b> Dispersión dinámica de luz (DLS) del complejo SdrP-CONCAP	23
<b>Figura 10.</b> DLS de los complejos CAP-LacNat y CAP-CONCAP sin cAMP	24
<b>Figura 11.</b> DLS de los complejos CAP-LacNat y CAP-CONCAP con cAMP	25
<b>Figura 12.</b> Ensayo de cambio en la movilidad electroforética del ADN	27
<b>Figura 13.</b> Análisis de los complejos SdrP-LacNat y SdrP-CONCAP	28
<b>Figura 14.</b> Parámetros de selección de secuencias homólogas de Clp	29

## **Anexos**

<b>Figura S1.</b> Vectores contruidos con el inserto CONCAP	S-1
<b>Figura S2.</b> Vectores contruidos con el inserto LacNat	S-2
<b>Figura S3.</b> Análisis de los sitios de unión de SdrP	S-3
<b>Anexo I.</b> Protocolo de PCR a partir de producto de ligación	S-4

## Resumen

### **Construcción de secuencias en tándem para análisis de interacción proteína-ADN en herramientas de molécula única**

La interacción proteína-ADN tiene un rol determinante en diversos mecanismos biológicos, como son la regulación de la transcripción de genes, la replicación y recombinación sitio específica. Entre las proteínas que llevan a cabo esta función, destacan los factores de transcripción. La gran mayoría de los estudios realizados de la interacción proteína-ADN se han llevado a cabo con estrategias tradicionales, las cuales llegan a enmascarar el comportamiento de moléculas que representen un porcentaje pequeño dentro la muestra. En años recientes, el empleo de herramientas de molécula única, como son la microscopia de fuerza atómica y las pinzas ópticas, ha facilitado el estudio de estas moléculas debido al uso de partículas individuales. El presente trabajo se enfoca en estudiar las proteínas CAP, Clp y SdrP, que pertenecen a la familia CRP/FNR, ampliamente distribuida en bacterias. Estos factores de transcripción tienen la capacidad de interaccionar con los mismos sitios de unión, el sitio de unión nativo del promotor de lactosa (LacNat) y un sitio consenso (CONCAP). Los tres objetivos principales que nos planteamos fueron los siguientes. El diseño y construcción de secuencias repetidas en tándem que contenga los sitios de unión LacNat y CONCAP, para su uso en trabajos posteriores de molécula única. Evaluar la interacción de CAP y SdrP con las secuencias construidas LacNat y CONCAP de un solo repetido, mediante EMSA y DLS. Finalmente, un análisis bioinformático para determinar secuencias ortólogas de Clp con características más favorables para su expresión y purificación. Los resultados constan de la obtención de las construcciones en tándem con dos y tres copias para la secuencia CONCAP y LacNat, respectivamente. La purificación de las proteínas CAP y SdrP, con un rendimiento final de 26 mg y 20 mg por litro de cultivo, respectivamente, y una pureza superior al 90%. Así mismo, los análisis de interacción preliminares indican que CAP y SdrP son capaces de interaccionar con las secuencias CONCAP y LacNat, con un solo repetido. La constante de disociación ( $K_d$ ) preliminar de SdrP por la secuencia CONCAP fue 0.23 nM y de 0.74 nM para LacNat, ambas estimadas por EMSA. Por último, se seleccionaron 10 ortólogos putativos cuya predicción de éxito en procesos de expresión y purificación fue más favorable, comparada con la secuencia de *Xanthomonas campestris* que ha sido cristalizada.

Palabras clave: Proteína activadora por catabolito (CAP), interacción proteína/ADN, dispersión dinámica de luz (DLS), ensayo de cambio en la movilidad electroforética (EMSA).

## Abstract

### **Construction of tandem repeats of DNA fragments for analysis of protein-DNA interactions with single-molecule methods.**

The protein-DNA interaction has a determining role in various biological mechanisms, such as the regulation of gene transcription, replication and site-specific recombination. Among the proteins that perform this function, the transcription factors stand out. In bacteria the family of CRP / FNR transcription factors is one of the most widely distributed. Because of the differences in their interaction mechanisms and their structural similarity, we are interested in three members of this family, CAP, Clp and SdrP. These proteins have the ability to interact with the same binding sites, the native lactose promoter binding site (LacNat) and a consensus site designed in the 1980s (CONCAP). The vast majority of studies of the protein-ADN interaction have been carried out with traditional strategies, which come to hide the behavior of molecules representing a small percentage within the sample. In recent years, the use of single-molecule tools, such as atomic force microscopy and optical tweezers, has facilitated the study of these molecules due to the use of individual particles. The three main objectives we proposed were as follows. The design and construction of repeated sequences in tandem containing the LacNat and CONCAP binding sites, for use in subsequent single molecule works. To evaluate the interaction of CAP and SdrP with the sequences constructed LacNat and CONCAP of a single repeated, by EMSA and DLS. Finally, a bioinformatic analysis to determine orthologous sequences of Clp with characteristics more favorable for its expression and purification. Our results show that the tandem integration of two and three copies of CONCAP and LacNat sequences, respectively, were successfully cloned in pBlueScript vector. We obtained protein concentrations of 26 mg and 20 mg per liter of culture for CAP and SdrP, respectively, and a purity greater than 90% for both proteins. Also, preliminary interaction analyzes indicate that CAP and SdrP are capable of interacting with the CONCAP and LacNat sequences when exposed to an individual copy of the binding site. A preliminary dissociation constant ( $K_d$ ) of 0.23 nM for CONCAP and 0.74 nM for LacNat was determined for SdrP transcription factor, estimated by EMSA. Finally, 10 putative orthologues were selected whose prediction of success in expression and purification processes was more favorable.

Key words: Catabolite activator protein (CAP), protein-ADN interaction, Dynamic Light Scattering (DLS), Electrophoretic Mobility Shift Assay (EMSA).

## INTRODUCCIÓN

La interacción proteína-ADN puede definirse como el reconocimiento de una secuencia nucleotídica por un dominio proteico. Esta interacción tiene un rol determinante en diversos mecanismos biológicos<sup>1-3</sup>. La unión sitio específica de proteína con el ADN es un proceso crucial en la regulación de la transcripción de genes, así como para la replicación y recombinación sitio específica<sup>4,5</sup>. Un ejemplo, es la RNA polimerasa que interacciona con secuencias específicas del promotor para iniciar la transcripción, mientras que muchos factores de transcripción interaccionan con el ADN para regular los niveles de transcripción<sup>6,7</sup>. En bacterias se conocen aproximadamente 30,642 factores de transcripción de unión a ADN agrupados en 39 familias, siendo la familia CRP/FNR una de las más ampliamente distribuidas<sup>8</sup>.

La familia CRP/FNR fue propuesta en 1983 cuando las proteínas CRP y FNR fueron reconocidas como homólogos<sup>9</sup>, siendo estas mismas proteínas las que darían el nombre a la familia. Actualmente, se conocen un gran número de miembros, cada uno implicado en la regulación de diversos procesos celulares. Las proteínas de esta familia típicamente actúan como activadores de la transcripción, pero también pueden ser represores importantes, y responder a diversas señales intracelulares y exógenas, como AMPc, condiciones anaeróbicas, estado redox, estrés oxidativo y nitrosante, monóxido de carbono, óxido nítrico o temperatura<sup>10,11</sup>. Los miembros de la familia presentan, en su mayoría, un tamaño constante entre los 230 y 250 aminoácidos. Estas proteínas forman dímeros con subunidades idénticas, donde cada subunidad está constituida por dos dominios, un dominio de unión al ADN el cual presenta un motivo hélice-giro-hélice (HTH) localizado en el extremo C-terminal y un dominio efector, que se encuentra en el extremo N-terminal, capaz de interaccionar con diversas moléculas propias de la célula para activar o reprimir la actividad de los factores de transcripción<sup>11,12</sup>. Dentro de esta familia destaca la proteína CAP de *E. coli* debido a que fue el primer factor de transcripción en tener su estructura tridimensional determinada, además de haber sido objeto de extensos estudios biofísicos, bioquímicos y genéticos<sup>13</sup>.

CAP regula la expresión de cerca de 200 genes en su mayoría involucrados en el metabolismo energético de la célula<sup>12,14</sup>. CAP actúa en función de los niveles

de glucosa en el medio, cuando estos niveles disminuyen, la concentración de AMPc aumenta, interaccionando específicamente con el dominio efector de CAP. Esta interacción provoca cambios alostéricos en la proteína promoviendo la transición de un estado inactivo a uno activo<sup>8,15-17</sup>. CAP es un dímero conformado por dos subunidades idénticas compuestas por 209 residuos cada una, con una masa total de 45 kDa y con un motivo HTH de unión a ADN, el cual contiene a la hélice F<sup>13,17</sup>. La estructura tridimensional de CAP en complejo con ADN ha sido determinada por análisis de difracción de rayos X con una resolución de 2.5 Å<sup>18</sup>.

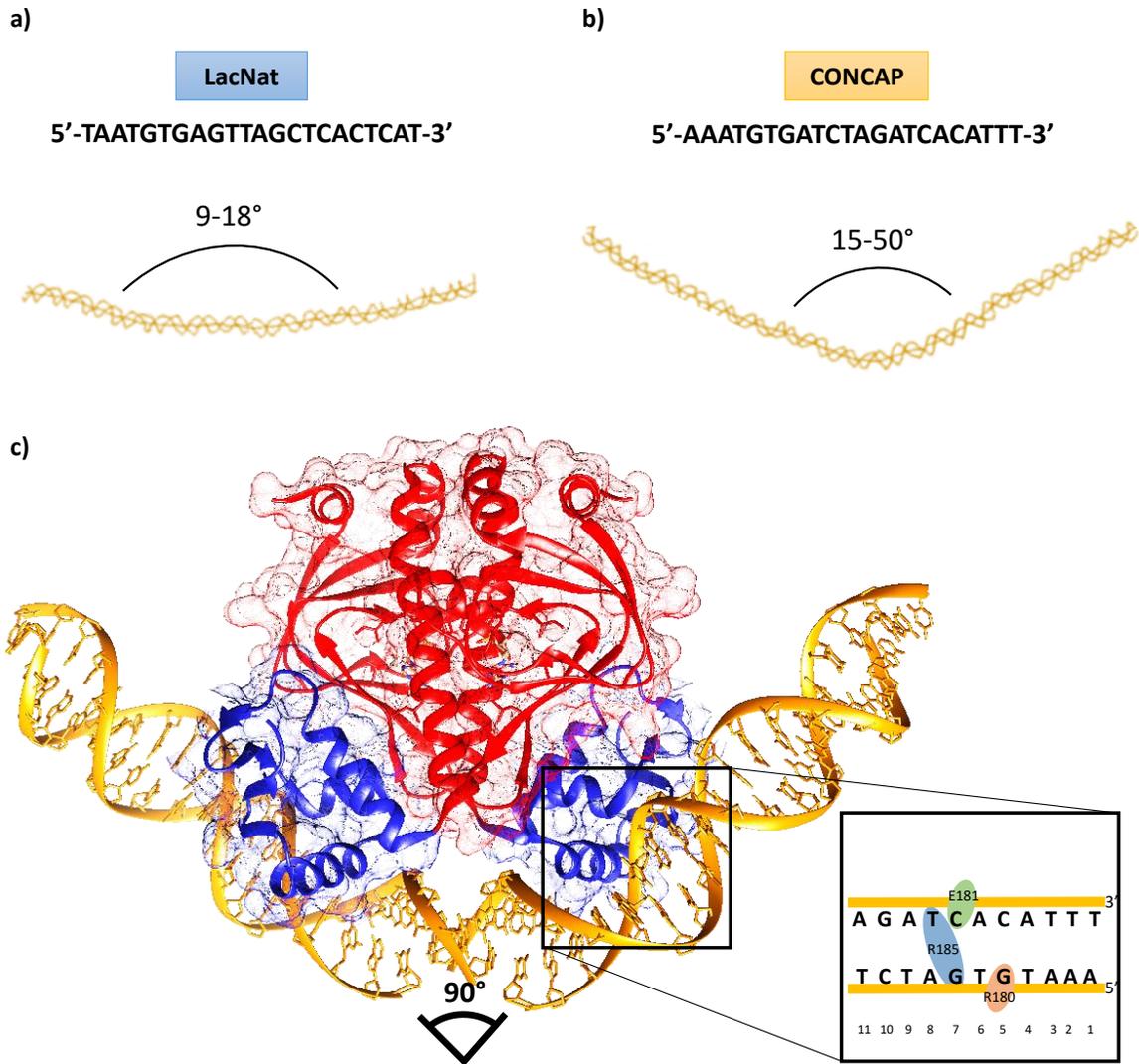
Para este factor de transcripción se han estudiado dos secuencias de unión en el ADN principalmente, una secuencia nativa que proviene del sitio de unión en el operón de lactosa que consta de 22 pb, 5'-TAATGTGAGTTAGCTCACTCAT-3' (LacNat) y una secuencia consenso diseñada a partir de más de 20 sitios de unión de CAP conocidos hasta ese momento<sup>16</sup>, la cual tiene una longitud de 22 pb y la particularidad de ser palíndroma, 5'-AAATGTGATCTAGATCACATTT-3' (CONCAP). Se han realizado estudios de interacción proteína-ADN con fragmentos de 40 pb de los sitios de unión LacNat y CONCAP con la proteína CAP mediante ensayos de unión en filtros de nitrocelulosa, con los que se determinó que CAP tiene un constante de disociación de 10.75 nM y de 0.0238 nM para la secuencia LacNat y CONCAP respectivamente, siendo CAP 450 veces más afín hacia la secuencia consenso que hacia la nativa<sup>16</sup>. Así mismo, se han realizado ensayos de cambio de movilidad electroforética del ADN (EMSA) en el cual evaluaron la interacción de CAP con el sitio LacNat empleando fragmentos de ADN de 219 pb a partir de los cuales se calculó una constante de disociación de  $1.5 \pm 0.3 \times 10^{-4} \text{ s}^{-1}$  en geles de poliacrilamida al 3.5 % y de  $3.5 \pm 0.4 \times 10^{-4} \text{ s}^{-1}$  en geles al 15 %<sup>19</sup>.

Se ha analizado el comportamiento de las secuencias LacNat y CONCAP en solución en ausencia de proteína. Para el caso de LacNat se realizaron ensayos hidrodinámicos, microscopía electrónica, así como ensayos de ciclación *in silico*. Mientras que la secuencia CONCAP se analizó mediante ensayos de ciclación, ensayos de cambio en la movilidad electroforética (EMSA) y transferencia de energía de resonancia luminiscente (FRET). Mediante estos análisis se determinó que por su naturaleza dinámica presentan torsiones intrínsecas que van de 9-18° en el caso de LacNat<sup>20</sup> y de 15-50° en CONCAP<sup>13</sup> (Fig. 1a y b). Sin embargo, cuando

la forma activa de la proteína se encuentra presente en la solución, la torsión del ADN aumenta hasta alcanzar los 90°, esto debido a la fuerza aplicada por la proteína al interactuar con el surco mayor del ADN. De acuerdo a la estructura del complejo CAP-ADN obtenida por cristalografía de rayos X, la interacción se da por el reconocimiento directo de los aminoácidos Arg180, Glu181 y Arg185 de la hélice F con los nucleótidos G5, C7 y G8 de la secuencia consenso<sup>13</sup> (Fig. 1c).

En resumen, CAP es un ejemplo clásico de un activador de la transcripción, el cual requiere de AMPc para adquirir su conformación activa; esta forma activa induce que la hélice F rote cerca de 60° y pueda interactuar con el ADN.

Por otro lado, se ha descrito un factor de transcripción dentro de la misma familia cuyo mecanismo de activación es contrario al presentado por CAP, esta proteína proviene de *Xanthomonas campestris* y se conoce como Clp (CAP-like Protein). Este factor de transcripción regula la expresión de aproximadamente 300 genes<sup>21-24</sup>. Estos genes se agrupan en trece categorías: enzimas extracelulares, síntesis de exopolisacáridos, multiresistencia a fármacos y desintoxicación, síntesis flagelar y motilidad, reacciones hipersensibles y sistema de patogenicidad, captación de hierro, metabolismo de aminoácidos y proteínas, ciclo del TCA, respiración aerobia y anaerobia, reguladores transcripcionales, componentes y transportadores de membrana, metabolismo de ácidos grasos y proteínas hipotéticas<sup>25</sup>. Es un homodímero, cada subunidad está conformada por 230 residuos, con una masa de 52 kDa<sup>24,26</sup>. Clp es homólogo a CRP con 45% de identidad de secuencia aminoacídica, se ha demostrado que no requiere de la unión de ningún efector para mantener su forma activa, por el contrario la unión del dinucleótido cíclico de GMP (c-di-GMP) induce cambios conformacionales importantes en la proteína que inhiben la capacidad de unión al ADN de este factor<sup>23,24</sup>, por lo que el aumento de la concentración intracelular de c-di-GMP regula negativamente la expresión de genes de virulencia y atenúa la virulencia bacteriana<sup>27</sup>. La estructura tridimensional de Clp fue obtenida por análisis de difracción de rayos X con una resolución de 2.3 Å<sup>24</sup>. Los aminoácidos característicos de esta proteína ya han sido determinados, siendo clave para la interacción con el c-di-GMP un ácido glutámico en la posición 99 (E99)<sup>27</sup>. Cuando este aminoácido es



**Figura 1.** Torsión intrínseca del ADN LacNat y CONCAP.

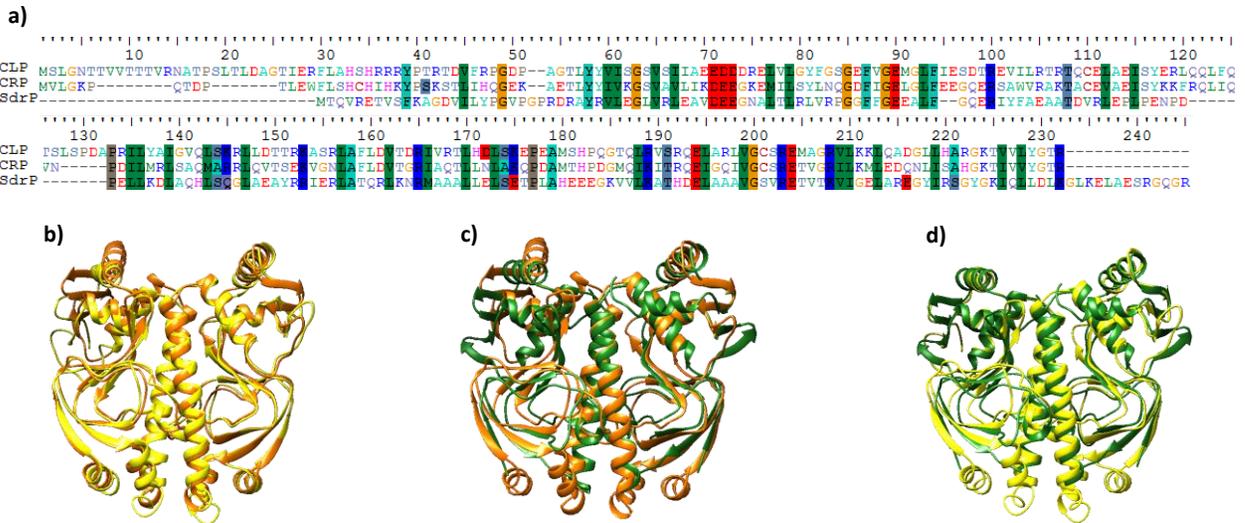
a) Sitio de unión LacNat, donde se observa la torsión intrínseca del ADN en solución acuosa y en ausencia de proteína, estos datos fueron determinados por difracción eléctrica transitoria. b) Sitio de unión CONCAP, donde se observa la torsión intrínseca del ADN en solución acuosa y en ausencia de proteína determinada mediante EMSA, FRET y ensayos de ciclación. c) Torsión inducida por la proteína CAP al interactuar con el surco mayor del ADN, en rojo se muestra el dominio efector de la proteína, y en azul el dominio de unión a ADN con motivo HTH. En la ampliación se muestran los aminoácidos específicos que interactúan con el surco mayor del ADN así como los nucleótidos que reconocen. Se empleó la estructura cristalizada de la proteína CAP con ADN almacenada en la Base de datos de proteínas (PDB) con ID 1G6N y visualizada en CHIMERA.

reemplazado por una serina (S) como se encuentra en su homólogo CAP, se priva completamente la capacidad de interacción entre Clp y el c-di-GMP *in vitro*, por lo que la proteína pierde la sensibilidad a la inhibición por esta molécula<sup>28</sup>. En 2005, se reportaron 2 sitios de unión putativos para Clp, 5'-CGATGTGATCGGTGCGGCAAT-3' y 5'-TTCTGTGGGGACGATCACACCA-3' pero hasta el momento no se han reportado complejos proteína-ADN cristalizados para esta proteína<sup>22</sup>, ni se han determinado sus constantes de disociación con el ADN. Sin embargo, se sabe que Clp es capaz de interaccionar con el sitio de unión del promotor de lactosa<sup>29</sup> así como con la secuencia CONCAP<sup>30,31</sup>.

Así mismo, se ha descrito un miembro de la familia CRP/FNR proveniente de *Thermus thermophilus* que no presenta regulación alostérica, llamado SdrP. Este se encuentra presente en la fase estacionaria de la célula, si bien SdrP no es un factor esencial para la viabilidad celular, interviene en la supervivencia de la misma bajo condiciones de estrés. SdrP está implicado en el suministro de energía y nutrientes, el control redox y la poliadenilación del mRNA. Al igual que los factores antes mencionados, también es un homodímero cuyas subunidades están formadas por 202 residuos aminoacídicos, con una masa de 44 kDa<sup>32-34</sup>. SdrP es homólogo a CRP y CLP con identidad de secuencia del 29 % y 28 %, respectivamente. La estructura tridimensional de SdrP se determinó por análisis de difracción de rayos X con una resolución de 1.5 Å<sup>34</sup>, donde se observó que SdrP tiene residuos con cadenas laterales grandes como Arg50, Glu60, glu61, Tyr70 y Tyr170 en el dominio efector, por lo que la cavidad reportada para la interacción con AMPc no existe en esta proteína, así mismo, ensayos realizados para evaluar la actividad en presencia de distintos catabolitos como AMPc y 2-oxoglutarato, indicaron que SdrP no requiere de ningún efector para permanecer activo, ni posee la capacidad de unirse a inhibidores conocidos para la familia CRP/FNR<sup>32,34</sup>. Hasta la fecha se han descrito al menos 16 sitios de unión para SdrP a partir de los cuales se obtuvo la secuencia consenso 5'-WWGTGA(N5-7)ACACWW-3'<sup>32,33</sup>, la cual es muy similar a la reportada para CAP. En estudios previos realizados en nuestro laboratorio, se observó la capacidad de esta proteína para interaccionar con la secuencia CONCAP<sup>35</sup> y utilizando el complejo SdrP-CONCAP se han obtenido cristales capaces de difractar a 2.5 Å de resolución con fragmentos de ADN de 30 pb (datos no publicados).

CAP, Clp y SdrP difieren sustancialmente en la secuencias de aminoácidos que las conforman así como en la actividad que desempeñan dentro de la célula. Sin embargo, si se analiza su estructura tridimensional sobreponiéndolas entre sí, son virtualmente idénticas respecto a la estructura cristalizada de la forma activa de CAP (Fig. 2). Así mismo, tienen la capacidad de interactuar con la misma secuencia de ADN<sup>16,26,29-31,35</sup>. En décadas recientes, el implemento de herramientas biofísicas para el análisis de interacción proteína-ADN ha permitido estudiar fenómenos que no eran directamente accesibles con herramientas tradicionales<sup>36-39</sup>. Herramientas como las pinzas ópticas y la microscopia de fuerza atómica ofrecen la oportunidad de estudiar procesos como la flexión inducida por una proteína al ADN, el ajuste molecular inducido por el reconocimiento entre proteínas y ADN, así como la transducción de energía mecanoquímica de los motores moleculares de unión a ADN<sup>40</sup>.

Este trabajo tiene tres objetivos principales, siendo el primero el diseño y construcción de secuencias con tres repeticiones idénticas en tándem de los sitios de unión LacNat y CONCAP, las cuales serán usadas en ensayos posteriores con pinzas ópticas para determinar las fuerzas de interacción proteína-ADN. El segundo, analizar la capacidad de los factores de transcripción de interés (CAP, Clp y SdrP) de interactuar con las secuencias en tándem. Y por último, la identificación de posibles ortólogos de la proteína Clp que presenten mejores características para optimizar su expresión y purificación.



**Figura 2.** Análisis comparativo de las proteínas CAP, Clp y SdrP

a) Alineamiento de las secuencias aminoacídica de las proteínas CAP, Clp y SdrP en BioEdit, los aminoácidos remarcados indican aquellos que se encuentran conservados en las tres secuencias b) Sobreposición de CAP (naranja) y Clp (amarillo), RMSD entre 153 átomos de 1.066Å c) Sobreposición de CAP y SdrP, RMSD entre 100 átomos de 1.178Å d) Sobreposición de Clp y SdrP (verde) RMSD entre 98 átomos de 1.053Å. Esquemas y valores obtenidos empleando las estructuras cristalizadas de la forma activa de cada proteína almacenadas en la Base de datos de proteínas (PDB) con ID (CRP:1G6N, Clp:3IWZ, SdrP:2ZCW) y analizados con CHIMERA.

## MATERIALES Y MÉTODOS

**Materiales.** Las enzimas empleadas fueron compradas a Thermo Fisher Scientific (Waltham, Massachusetts, USA), AMPc y el resto de los reactivos son de Sigma-Aldrich (St. Louis, MO, USA).

**Construcción del vector pBlueScript/CONCAP\*2.** Las condiciones generales para llevar a cabo la reacción de PCR son 0.5  $\mu$ M de cada oligonucleótido, 200  $\mu$ M de cada dNTP y 0.5 unidades de Taq ADN polimerasa en un volumen final de 50  $\mu$ L, en párrafos posteriores serán descritos cambios puntuales a esta reacción. Para construir el vector en tándem pBlueScript/CONCAP\*2 fue necesaria la construcción inicial del vector pET28/CONCAP. La secuencia CONCAP fue sintetizada de novo por Integrated ADN Technologies, IDT (Coralville, Iowa) y dividida en dos partes a las que se le nombró gBlock G1 y gBlock G2. A partir de los gBlocks G1 y G2 (Fig S-1), se armó la secuencia consenso de unión de CAP. Para lo cual se realizó una reacción de PCR limitante, usando los oligonucleótidos CAP38Fext, 910\_IntRev, 910\_IntFor, CAP38Rext (Tabla 1), donde los oligonucleótidos externos (CAP38Fext y CAP38Rext) se encontraban diluidos en una relación molar 1:10 respecto a su contraparte, lo cual permitió la amplificación de la cadena molde de G1 y la cadena codificante de G2. A continuación, se realizó una reacción de PCR de fusión con el producto de cada PCR limitante en concentraciones equimolares, se corrieron 8 ciclos de amplificación sin oligonucleótidos, posteriormente se amplificó con los oligonucleótidos externos durante 30 ciclos. Una vez obtenido, el producto de fusión que consta de 910 pb, tanto el vector pET28 como el inserto (producto de PCR de fusión) fueron digeridos con las condiciones siguientes: 1X NEBuffer 2.1, 10 unidades de las enzimas NdeI y HindIII, y 360 ng de ADN en un volumen final de 20  $\mu$ L. Una vez realizada la reacción de digestión, se llevó a cabo una reacción de ligación de los mismos, cuyas condiciones se muestran a continuación: 1X amortiguador T4 ligasa, 5 unidades de la enzima T4 ligasa, vector e inserto digerido en una relación 1:4 M en un volumen final de 10  $\mu$ L.

Una vez construido el vector pET28/CONCAP, se transformaron por choque termino células calcio-competentes de *E. coli* Top10 y las transformantes se analizaron por PCR de colonia. Las clonas positivas se aislaron y el ADN plasmídico

de interés fue purificado por columna empleando el Kit GeneJet Plasmid Miniprep (Thermo Scientific), la ausencia de mutaciones dentro del inserto se verificó mediante secuenciación.

Los insertos para la construcción en tándem fueron obtenidos a partir de una reacción de PCR anidada, utilizando como molde el vector pET28/CONCAP descrito anteriormente. Se emplearon los oligonucleótidos HindIII\_4sF/XhoI\_4sR y EcoRI\_4sF/HindIII\_4sR para obtener los insertos CONCAP\*1 y CONCAP\*2, respectivamente. Estos insertos fueron subclonados secuencialmente dentro del vector pBlueScript II KS entre los sitios de restricción XhoI-HindIII y EcoRI-HindIII.

**Construcción del vector pBlueScript/LacNat\*3.** Para la construcción del vector en tándem pBlueScript/LacNat\*3 fue necesaria la construcción inicial del vector pET28/LacNat. El inserto LacNat de 1031pb, en cuyo centro se encuentra el sitio de unión nativo de la proteína CAP, fue construido a partir de la amplificación por reacción de PCR oligo-específica del operón de lactosa de la cepa BL21 Start de *E. coli*. El cual, fue clonado entre los sitios de restricción NdeI y HindIII del vector pET28. Este vector se empleó para transformar células calcio competentes de *E. coli* Top10, las colonias obtenidas se analizaron mediante un ensayo de PCR de colonia. A partir de las colonias positivas, se purificó el vector como se menciona anteriormente y se verificó por secuenciación que la construcción fuese correcta.

Para llevar a cabo la construcción del tándem, fueron obtenidos tres insertos por reacción de PCR anidada a partir del vector pET28/LacNat previamente descrito. Se emplearon los oligonucleótidos XhoI\_4sF/HindIII\_4sR, EcoRI\_4sF/HindIII\_4R y HindIII\_4sF/HindIII\_4sR para amplificar los insertos LacNat\*1, LacNat\*2 y LacNat\*3, respectivamente. Estos insertos fueron subclonados secuencialmente en el vector pBlueScript II KS entre los sitios de restricción XhoI-HindIII, EcoRI-HindIII y HindIII-HindIII. El arreglo correcto de las secuencias en tándem se verificó mediante una reacción de PCR a partir de la reacción de ligación, con el siguiente protocolo. Se realizó una dilución 1:10 con H<sub>2</sub>O MiliQ de 1 µL de la reacción de ligación para amplificar las secuencias recombinantes (vector e inserto) empleando oligonucleótidos externos propios del vector pBlueScript II KS. La reacción consiste en 0.5 µM de cada oligo, 200 µM de cada dNTP y 0.5 unidades de Taq ADN polimerasa un volumen final de 20 µl, la

cual fue llevada a cabo en el termociclador T100 Thermal Cycler (Bio-Rad) (Véase ANEXO A para protocolo detallado).

**Expresión y purificación de CAP.** El vector pET28-PPS/CAP fue construido en nuestro grupo en un trabajo previo, para realizar ensayos de cristalización <sup>30</sup>. Para la expresión de CAP, se transformó por choque térmico células calcio-competentes de la cepa BL21 Star con el vector previamente mencionado. A partir de una sola colonia, se creció un pre-inoculo de 50 mL en medio LB con 50 µg/ml de kanamicina a 37°C por 16 h a 190 rpm. El pre-inoculo se utilizó para inocular 1 L de medio fresco LB con 50 µg/ml de kanamicina haciendo una dilución 1:20, éste se incubó a 37 °C a 190 rpm hasta alcanzar una densidad óptica (OD<sub>600</sub>) de 0.6-0.8. La inducción de la expresión de la proteína se realizó con una concentración final de 1mM de IPTG a 37 °C por 5 h en agitación constante (190 rpm). Para recuperar el pellet celular, el cultivo fue centrifugado a 10,000 rpm por 10 min a 4 °C. Se retiró el sobrenadante y el botón celular fue almacenado a -20 °C hasta su uso.

Para la purificación de la proteína, el botón celular fue resuspendido en 25 mL de amortiguador de lisis (50 mM Tris pH 8.0, 300 nM NaCl y 10 mM Imidazol) y sonicado en hielo con una amplitud de 50%, durante 7 min 30 seg (15 ON, 45 OFF). Una vez lisadas las células, se centrifugaron a 12,000 rpm por 10 min a 4 °C, separando el sobrenadante de los restos celulares. Con el sobrenadante se realizó una cromatografía de afinidad a níquel (IMAC) en una columna Ni-NTA (Qiagen) previamente equilibrada con el amortiguador de lisis. La elución se realizó con un gradiente escalonado de imidazol de 20, 200 y 500mM, utilizando 3 volúmenes de columna en cada paso. Cada una de las fracciones donde eluyó la proteína fueron colectadas, y se analizaron por electroforesis en gel de poliacrilamida desnaturizante (SDS-PAGE) en geles al 12%. La etiqueta de histidinas fue removida con 5 U/mL de *Prescission protease* (PPS) (Ge Healthcare Life Science) incubándose a 4°C por 16 h. La proteína sin la etiqueta de histidinas se concentró por ultrafiltración a un volumen de 2.5 mL (Vivaspin Turbo 10 kDa, Sartorius). Una vez concentrada, fue desalada en la columna PD-10 empleando el amortiguador 20mM Tris pH 8 y 2mM EDTA. Para la cromatografía de afinidad a heparina, la muestra (5 mL) fue inyectada en la columna HiTrap Heparina acoplada a un equipo AKTA-FPLC (GE Healthcare). La columna fue equilibrada con un amortiguador A

(10 mM Tris pH 8, 0.1 mM EDTA) y se empleó un gradiente lineal de 0-100% del amortiguador B (10 mM Tris pH 8, 0.1 mM EDTA, 1M NaCl) en 20 volúmenes de columna para eluir la proteína. El frente fue colectado, así como fracciones de 3 mL con un flujo constante de 2 mL/min. Se determinaron las fracciones correspondientes a la proteína con el cromatograma de elusión. Una vez obtenidas las fracciones, se realizó una electroforesis SDS-PAGE en geles de poliacrilamida al 12%. Posteriormente se cuantificó la concentración de proteína por espectrofotometría usando el coeficiente de extinción molar teórico ( $20065 \text{ M}^{-1} \text{ cm}^{-1}$ ). La proteína fue concentrada a 2 mg/mL por ultrafiltración y se hicieron alícuotas de proteína de 500  $\mu\text{L}$  cada una, las cuales fueron almacenadas a 4 °C.

**Expresión y purificación de SdrP.** El vector pET28-PPS/SdrP fue construido en nuestro grupo en un trabajo previo, para cristalizar el complejo SdrP/ADN<sup>35</sup>. La expresión de la proteína SdrP se realizó siguiendo un protocolo similar al descrito previamente para CAP, con las siguientes modificaciones. La inducción de la expresión de SdrP se llevó a cabo a 28 °C por 12 h. Para el primer paso de purificación de SdrP se empleó el protocolo previamente descrito. Sin embargo, para el segundo paso, se realizó una cromatografía de intercambio iónico en la columna Hi Trap Q acoplada a un equipo AKTA-FPLC. La columna fue equilibrada con un amortiguador A (20 mM Tris pH 8, 2 mM EDTA) y se empleó un gradiente lineal de 0-50% del amortiguador B (20 mM Tris pH 8, 2 mM EDTA, 1M NaCl) en 20 volúmenes de columna para eluir la proteína. Se recolectaron fracciones de 3 ml con un flujo constante de 2 ml/min. Posterior a los ensayos de purificación, la concentración de proteína fue cuantificada por espectrofotometría usando el coeficiente de extinción molar teórico ( $8940 \text{ M}^{-1} \text{ cm}^{-1}$ ). La proteína fue concentrada a 1 mg/mL por ultrafiltración, se hicieron alícuotas de 500  $\mu\text{L}$  y almacenó a 4 °C.

**Dispersión dinámica de luz (DLS).** Para este ensayo se analizaron muestras de proteína (SdrP, CAP), ADN (CONCAP) y complejos proteína-ADN (SdrP-CONCAP, SdrP-LacNat, CAP-CONCAP, CAP-LacNat). El volumen final de las muestras de trabajo fue de 40  $\mu\text{L}$ . Las proteínas SdrP y CAP fueron llevadas a una concentración final de 1 mg/mL y 0.8 mg/mL, respectivamente. Para el ADN CONCAP se empleó una concentración final de 200 ng. Para los complejos SdrP-

CONCAP y SdrP-LacNat se emplearon concentraciones crecientes de ADN (200 ng, 600 ng y 4,000 ng totales) así como 1 mg/mL de proteína. Para los complejos CAP-CONCAP y CAP-LacNat en su forma inactiva, se usaron concentraciones crecientes de cada ADN (50 ng, 100 ng, 150 ng y 200 ng) y una concentración constante de 0.8 mg/mL de proteína. Los complejos CAP-ADN fueron activados saturando la solución de trabajo con 300  $\mu$ M de AMPc, se emplearon concentraciones de 25 ng y 50 ng finales de ADN (CONCAP/LacNat) así como 0.8 mg/mL de CAP. El equipo usado para llevar a cabo estos ensayos fue Zetasizer APS (Malvern Instruments). Las muestras fueron pre-incubadas a 25 °C por 5 min en la celda del equipo. La calibración del láser se realizó automáticamente por el equipo. Se realizaron 3 lecturas por muestra, cada lectura consta de 15 mediciones con una duración de 40 seg cada una, espaciadas por 2 min de incubación.

**Ensayo de cambio en la movilidad electroforética del ADN (EMSA).** Para los ensayos EMSA se empleó una concentración constante de 160 ng (10.27 nM) de ADN, el cual consta de 800 pb obtenidas por PCR a partir de los vectores pET28/LacNat y pET28/CONCAP con los oligonucleótidos HindIII\_4sF y XhoI\_4sR. La proteína SdrP se utilizó en un gradiente de 0.375 a 374.4 ng, preparados por dilución a partir de tres soluciones “stock” con la siguiente concentración,  $1.4 \times 10^{-7}$  M,  $1.4 \times 10^{-8}$  M y  $1.4 \times 10^{-9}$  M (Tabla 2). Las muestras que contienen el complejo SdrP/ADN con concentración constante de ADN y concentración variable de SdrP, fueron incubadas a 37°C durante 30 min, antes de ser resueltas por electroforesis. Para la electroforesis se utilizaron geles de poliacrilamida nativos (native PAGE) al 5% con amortiguador TBE. Los geles se pre-corrieron a 100 V por 40 min, antes de cargar las muestras. La electroforesis se realizó a 70 V por 6 horas. Los geles fueron teñidos con bromuro de etidio y desteñidos con agua destilada, y reteñidos con azul de Coomassie para visualizar los controles proteicos. La documentación de los geles se llevó a cabo en el equipo Gel Doc XR+ de Bio-Rad. Los datos obtenidos se analizaron en Image Lab 6 de Bio-Rad y se calcularon las constantes de disociación en OriginPro 8.

**Análisis bioinformático de ortólogos putativos de la proteína Clp.** Se realizó un BLASTP para buscar secuencias relacionadas con la proteína Clp (3IWZ) en la base de datos NCBI. El BLASTP se llevó a cabo con los parámetros

predeterminados por el programa. A partir de las secuencias obtenidas se aplicaron filtros para seleccionar ortólogos de Clp. El tamaño de las secuencias fue delimitado en  $\pm 5$  aminoácidos respecto a los 230 residuos de Clp, se delimitaron las secuencias al orden de las *Xanthomodales*, las secuencias repetidas identificadas por un alineamiento múltiple de secuencias llevado a cabo en el programa Bioedit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) fueron eliminadas, de igual forma se analizó la presencia de los aminoácidos característicos de la proteína Clp en las secuencias ortólogas putativas. Algunos de los residuos característicos analizados fueron R201, E202, R206 que participan en la interacción con el ADN y D70, R154, R166, D170 que están asociados con la interacción del c-di-GMP. Se analizó el contenido aminoacídico en el programa Protparam (<http://web.expasy.org/protparam/s>), el coeficiente de agregación en el servidor Aggrescan (<http://bioinf.uab.es/aggrescan/>), la solubilidad en el servidor SPpred (<http://crdd.osdd.net:8081/sppred/submit.jsp>), así como la predicción de la clonación, expresión, purificación y cristalización de dichas proteínas en los servidores CrysaliS I y II (<http://nmrcen.xmu.edu.cn/crysalis/>) así como en XtalPred-RF (<http://ffas.burnham.org/XtalPred-cgi/xtal.pl>).

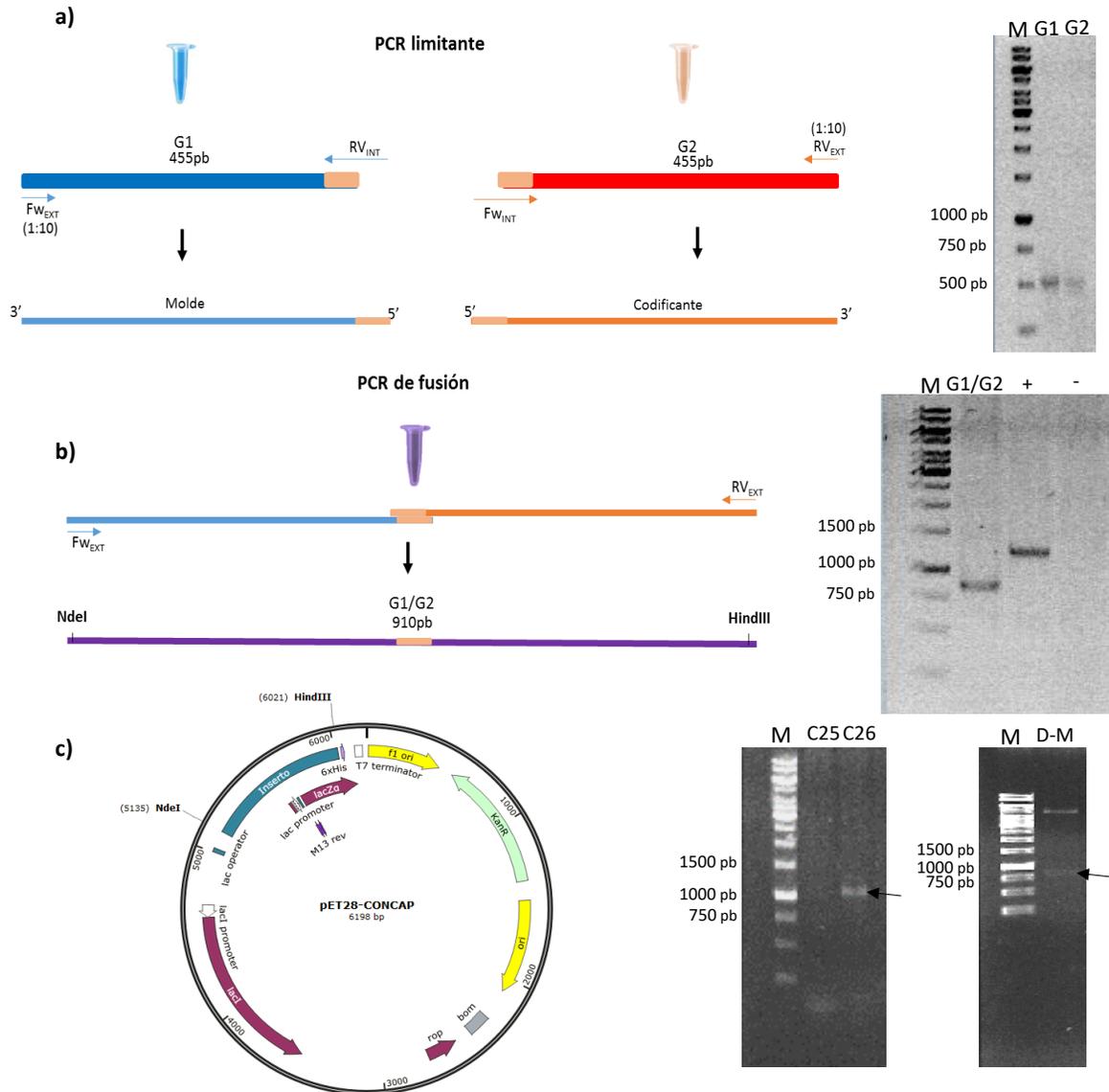
## RESULTADOS

**Construcción del vector en tándem pBlueScript/CONCAP\*2.** La secuencia diseñada para las construcciones consta de 910 pb, y proviene del promotor de lactosa en cuyo centro tiene integradas las 22 pb del sitio de unión consenso de la proteína CAP. El vector pET28/CONCAP fue construido a partir de la reacción de PCR de fusión de los gBLOCKs G1 y G2. Para llevarla a cabo, se realizó una reacción de PCR limitante, para obtener hebras monocatenarias de ambos gBlocks de 455 pb (Fig. 3a). Para el ensayo de PCR de fusión se mezclaron en una misma reacción las hebras monocatenarias obtenidas previamente, obteniendo el inserto complejo de 910 pb (Fig. 3b). Se clonó en el vector pET28 entre los sitios de restricción NdeI y HindIII. Se seleccionó la clona positiva por PCR de colonia y se comprobó el ensamblaje adecuado por digestión de vector así como por secuenciación (Fig. 3c). Este vector se nombró pET28/CONCAP.

Después de verificar la construcción pET28/CONCAP por secuenciación, se emplearon los oligonucleótidos HindIII\_4sF, XhoI\_4sR, EcoRI\_4sF, HindIII\_4sR (Tabla 1), para construir los insertos necesarios para hacer el vector en tándem. Cada inserto fue subclonado de forma secuencial en el vector de clonación pBlueScript II KS, entre los sitios de restricción XhoI-HindIII, y HindIII-EcoRI. Este vector fue llamado pBlueScript/CONCAP\*2 (Fig. 4).

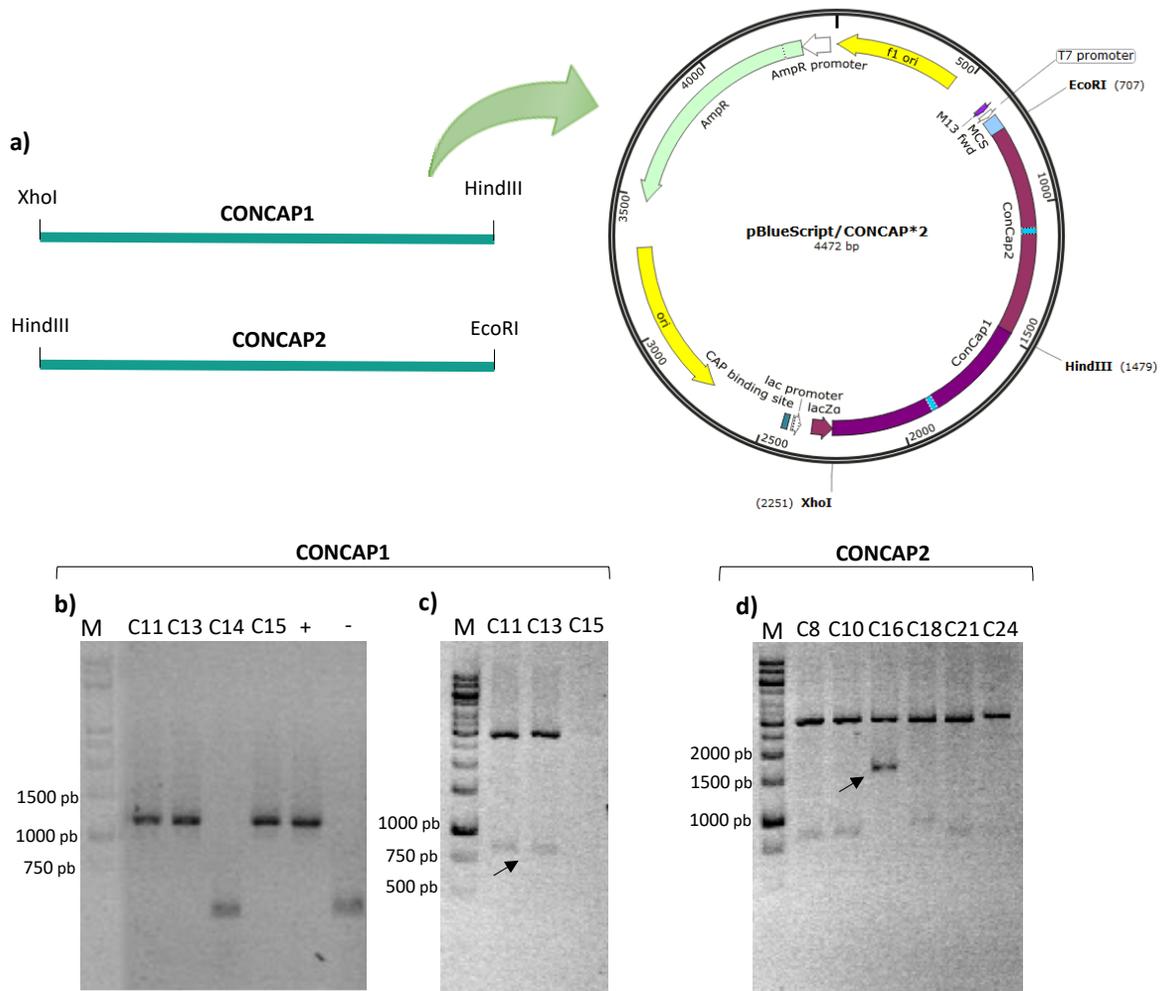
**Construcción del vector en tándem pBlueScript/LacNat\*3.** El promotor de lactosa se amplificó mediante PCR a partir del ADN genómico de *E. coli*, obteniendo un amplicón de 1031 pb en cuyo centro se encuentra el sitio de unión a CAP nativo, empleando los oligonucleótidos CAP38Fext y NEWR (Fig. 5a). El amplicón fue purificado y clonado dentro del vector pET28 entre los sitios de restricción NdeI y HindIII (Fig. 5b). Se seleccionó la clona positiva por PCR de colonia. A partir de la digestión de vector purificado se comprobó el ensamblaje correcto del inserto (Fig. 5c) así mismo, el vector fue secuenciado. Este vector se nombró pET28/LacNat. Los resultados obtenidos por secuenciación mostraron que construcción no presentaba mutaciones.

Los oligonucleótidos HindIII\_4sF, XhoI\_4sR, EcoRI\_4sF, HindIII\_4sR fueron empleados para obtener un amplicón de 800 pb, los cuales fueron digeridos con las



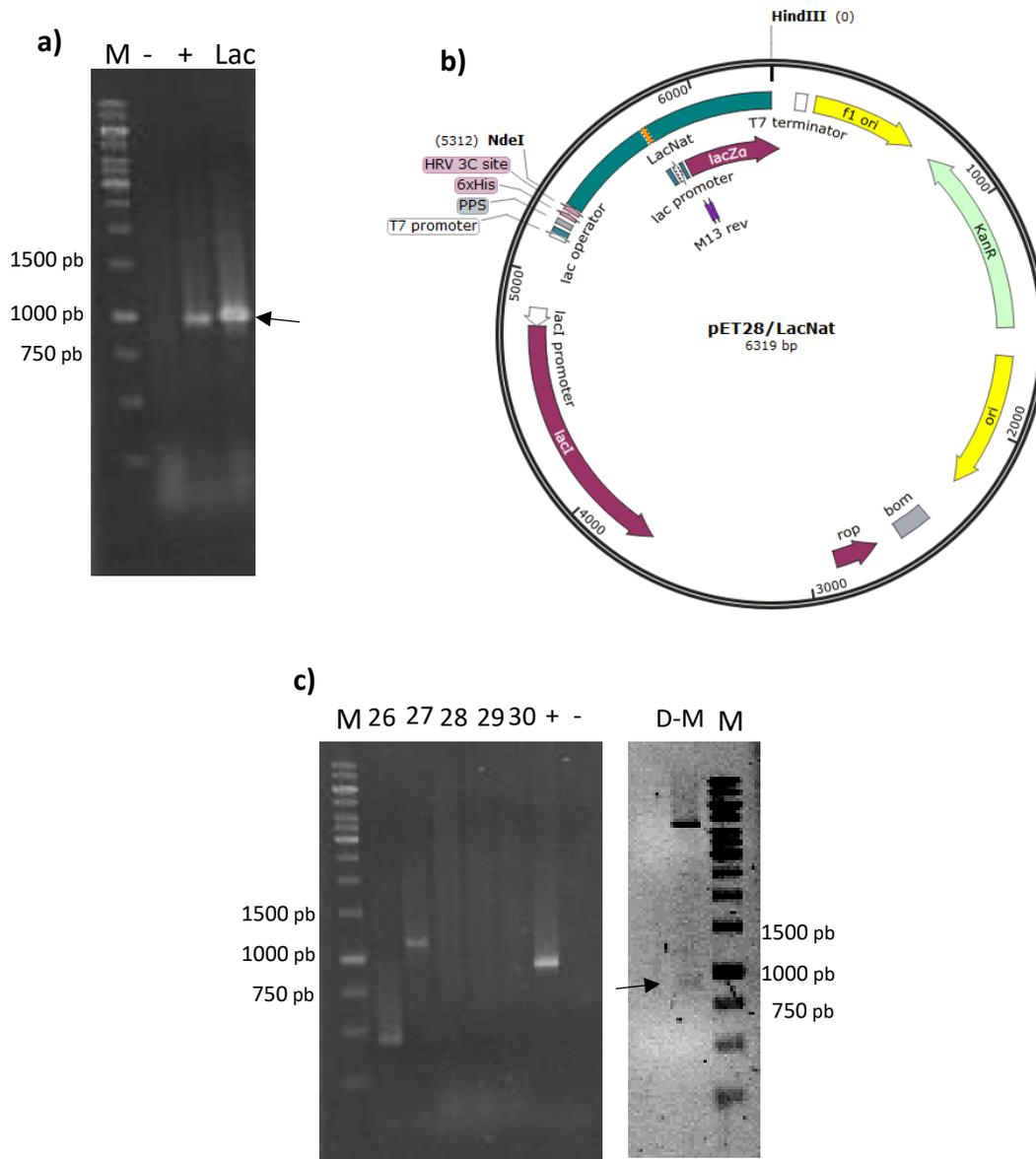
**Figura 3.** Construcción del vector pET28/CONCAP

a) PCR limitante: los oligonucleótidos externos señalados en la imagen como Fw<sub>ext</sub> (CAP38Fext) y Rv<sub>ext</sub> (CAP38Rext) deben estar diluidos a concentración final 1:10 respecto a su contraparte, obteniendo mayoritariamente hebras monocatenarias, el gel a la derecha muestra los gBlocks G1 y G2 de 455 pb. b) PCR de fusión de los gBlocks G1 y G2, el gel muestra el inserto fusionado (G1/G2) de 910pb, el control positivo (+) y el negativo (-). c) La secuencia fusionada se clonó en un vector pET28 entre los sitios de restricción NdeI y HindIII. Gel al 1% que muestra la PCR de colonia de las células transformadas con el vector, en el carril C26 aparece el amplicón de interés (1147 pb), el carril D-M muestra la liberación del inserto (886 pb) una vez digerido con las enzimas HindIII y NdeI el vector purificado.



**Figura 4.** Construcción del vector pBlueScript/CONCAP\*2

a) Esquema del vector pBlueScript/CONCAP\*2, con los insertos clonados entre los sitios de restricción XhoI-EcoRI. b) PCR de colonia para evaluar la presencia de pBlueScript/CONCAP\*1; los carriles C11, C13 y C15 muestran el amplicón de interés (~1000 pb). c) Digestión con las enzimas XhoI-HindIII del vector pBlueScript/CONCAP\*1 purificado, el carril C13 muestra el inserto de interés (~800pb). d) Digestión con las enzimas XhoI-EcoRI de vectores purificados a partir de células transformadas con pBlueScript/CONCAP\*2, el carril C16 indica la construcción correcta (~1600pb).



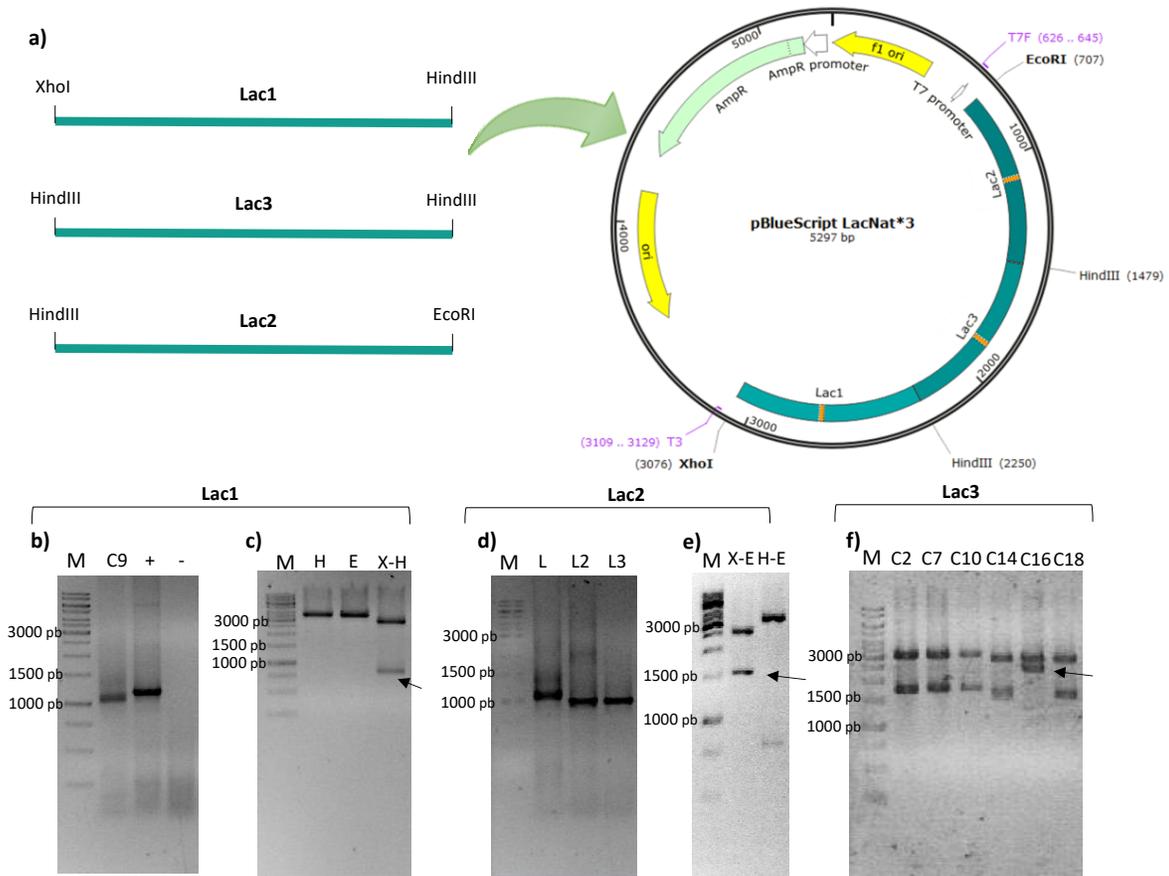
**Figura 5.** Construcción del vector pET28/LacNat

a) Electroforesis en gel de agarosa 1%, donde se muestra el producto de PCR (~1031 pb) obtenido a partir del operón de lactosa y etiquetado como Lac, los controles positivo (+) y negativo (-). b) Esquema del vector pET28/LacNat el cual tiene clonado el inserto LacNat entre los sitios de restricción NdeI y HindIII. c) Gel de agarosa al 1% (izq.), donde se analizaron las colonias transformadas mediante PCR de colonia, en el carril 27 se observa el vector completo. Gel de agarosa al 1% (der.) donde se comprobó el ensamblaje adecuado por digestión del vector, el carril D-M se observa la liberación del inserto (~1000 pb)

enzimas de restricción correspondientes. El inserto fue subclonado en el vector de clonación pBlueScript II KS, donde se clonaron tres insertos en tándem, entre los sitios de restricción XhoI-HindIII, HindIII-HindIII y HindIII-EcoRI. Este vector fue llamado pBlueScript/LacNat\*3 (Fig. 6). Para asegurar la obtención del tándem, se realizaron pasos de purificación en cada etapa de construcción del vector.

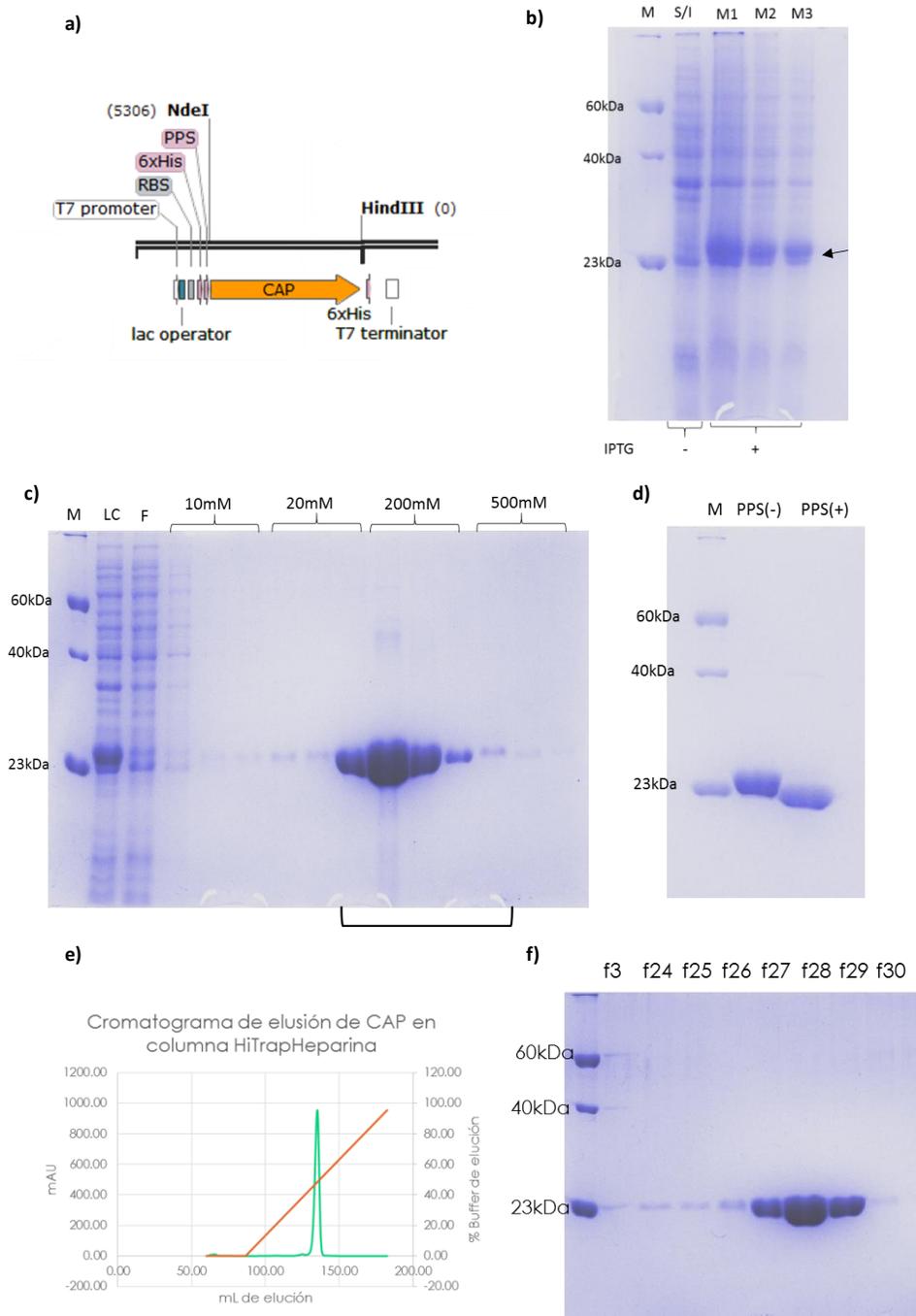
**Expresión y purificación de la proteína CAP.** Para llevar a cabo los estudios de interacción se requiere la proteína purificada, por ello, se transformó la cepa BL21 Star con la construcción pET28-PPS/CAP (Fig. 7a). En las pruebas de expresión de CAP se observa una banda diferencial correspondiente a la proteína de interés en los carriles M1, M2 y M3 (Fig. 7b). En el primer paso de purificación por cromatografía de afinidad a níquel, la proteína eluyó de la tercera fracción de 20mM a la primera fracción de 500 mM de imidazol (Fig. 7c). La evaluación del corte de la proteasa PPS mostró la remoción total de la etiqueta de histidinas, donde la proteína sin etiqueta se encuentra en el carril PPS(+) (Fig. 7d). Durante el segundo paso de purificación se observó una fuerte afinidad por la heparina eluyendo entre los 400-500mM de NaCl (Fig. 7e). Se determinó por análisis de densitometría una pureza del 98% de la proteína CAP, obteniendo un rendimiento final de 26 mg por litro de cultivo.

**Expresión y purificación de la proteína SdrP.** La cepa BL21 Star fue transformada con la construcción pET28-PPS/SdrP (Fig. 8a). La evaluación de la expresión de SdrP, mostró una banda diferencial correspondiente a ~25 kDa, en los carriles M1 y M2 (Fig. 8b). En el primer paso de purificación de SdrP, la proteína eluyó de la tercera fracción de 20 mM a la primera fracción de 500mM de imidazol (Fig. 8c). En la figura 8d, se observa la remoción completa de la etiqueta de histidinas, donde la proteína sin etiqueta tiene un peso aproximado de 22 kDa. En el segundo paso de purificación se obtuvo un pico de elución entre los 180-280mM de NaCl (Fig. 8e). Por análisis de densitometría se determinó una pureza del 90%, obteniendo un rendimiento final de 20 mg por litro de cultivo.



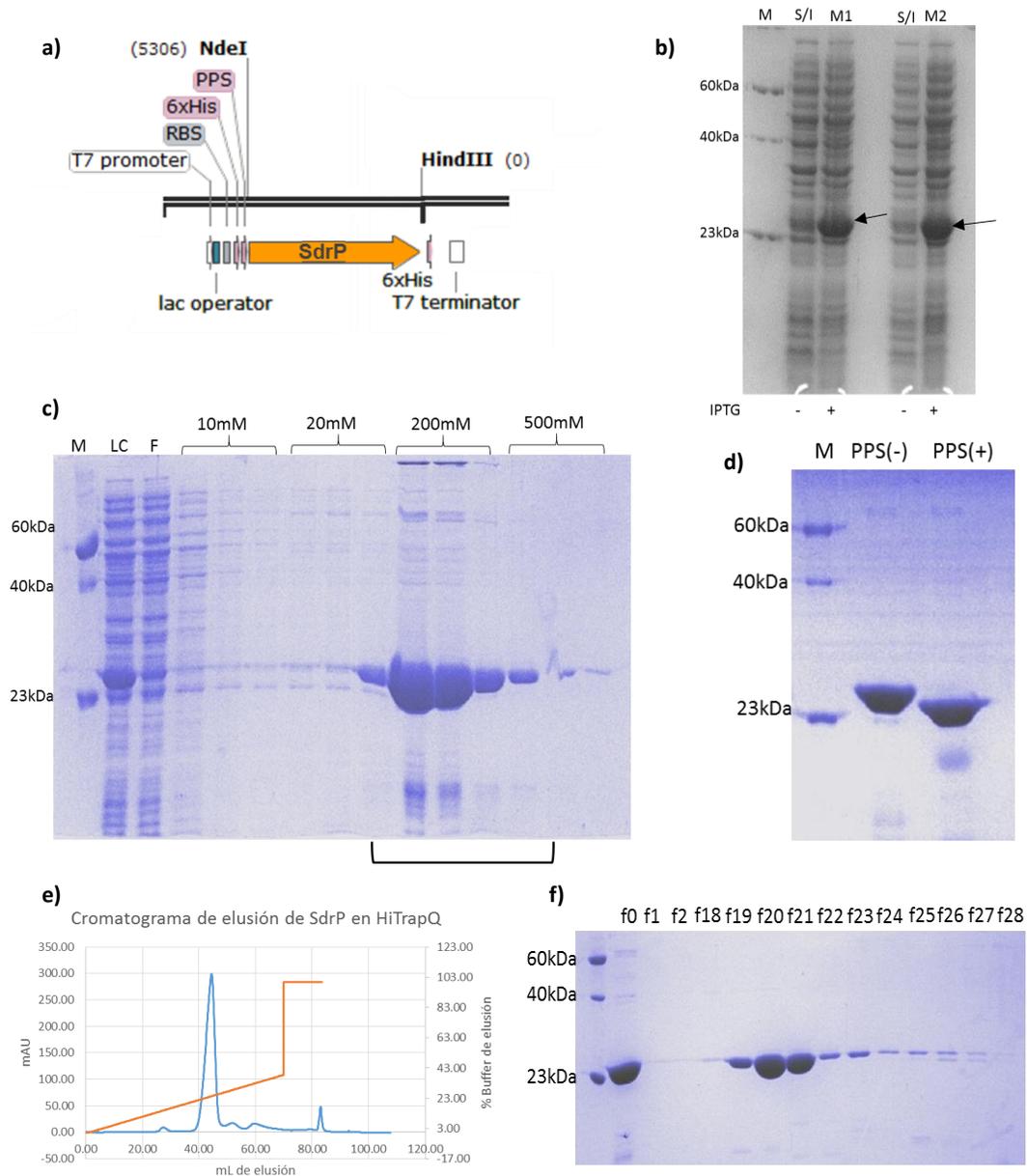
**Figura 6.** Construcción del vector pBlueScript/LacNat\*3

a) Esquema del vector pBlueScript/LacNat\*3, con los insertos clonados entre los sitios de restricción XhoI-HindIII-EcoRI. b) PCR de colonia del inserto Lac1 en el vector pBlueScript, donde se observa que la colonia 9 (C9) contiene el inserto de interés de ~1100 pb c) Digestión del vector para confirmar el ensamblaje correcto del vector, donde la digestión con HindIII (carril H) y EcoRI (carril E), muestra un tamaño de ~3700 pb, mientras que con XhoI-EcoRI (carril X-H) libera un inserto de ~800 pb d) PCR a partir de ligación de tres ligaciones (L, L2, L3) con el inserto Lac2 en el vector pBlueScript/LacNat\*1, donde no se aprecia el tamaño real del amplicón. e) Digestión de vector pBlueScript/LacNat\*2 purificado con XhoI-EcoRI (carril X-E), donde se observa la liberación de los insertos Lac1 y Lac2 (~1600pb) y al digerir con HindIII-EcoRI (carril H-E) libera el inserto Lac2 (~800pb). f) Digestión con XhoI-EcoRI del vector pBlueScript/LacNat\*3 purificado, la flecha en el carril C16 señala la liberación de los tres insertos (~2400pb).



**Figura 7.** Expresión y purificación de la proteína CAP

a) Esquema del vector pET28-PPS/CAP. b) Gel al 12% para evaluar la expresión de CAP, donde se muestra el marcador (M), tiempo cero (S/I), y el cultivo de tres matraces que fueron inducidos con IPTG (M1, M2, M3). c) Gel al 12% de la purificación por cromatografía de afinidad a níquel, se observa la elución de CAP en su mayoría en 200mM de imidazol. d) Gel para evaluar el corte de histidinas con PPS. e) Cromatograma de la columna HiTrap Heparin, donde la proteína eluyó entre 400-500mM de NaCl. f) Gel al 12% con las fracciones eluidas de la columna de heparina para corroborar el estado de la proteína.



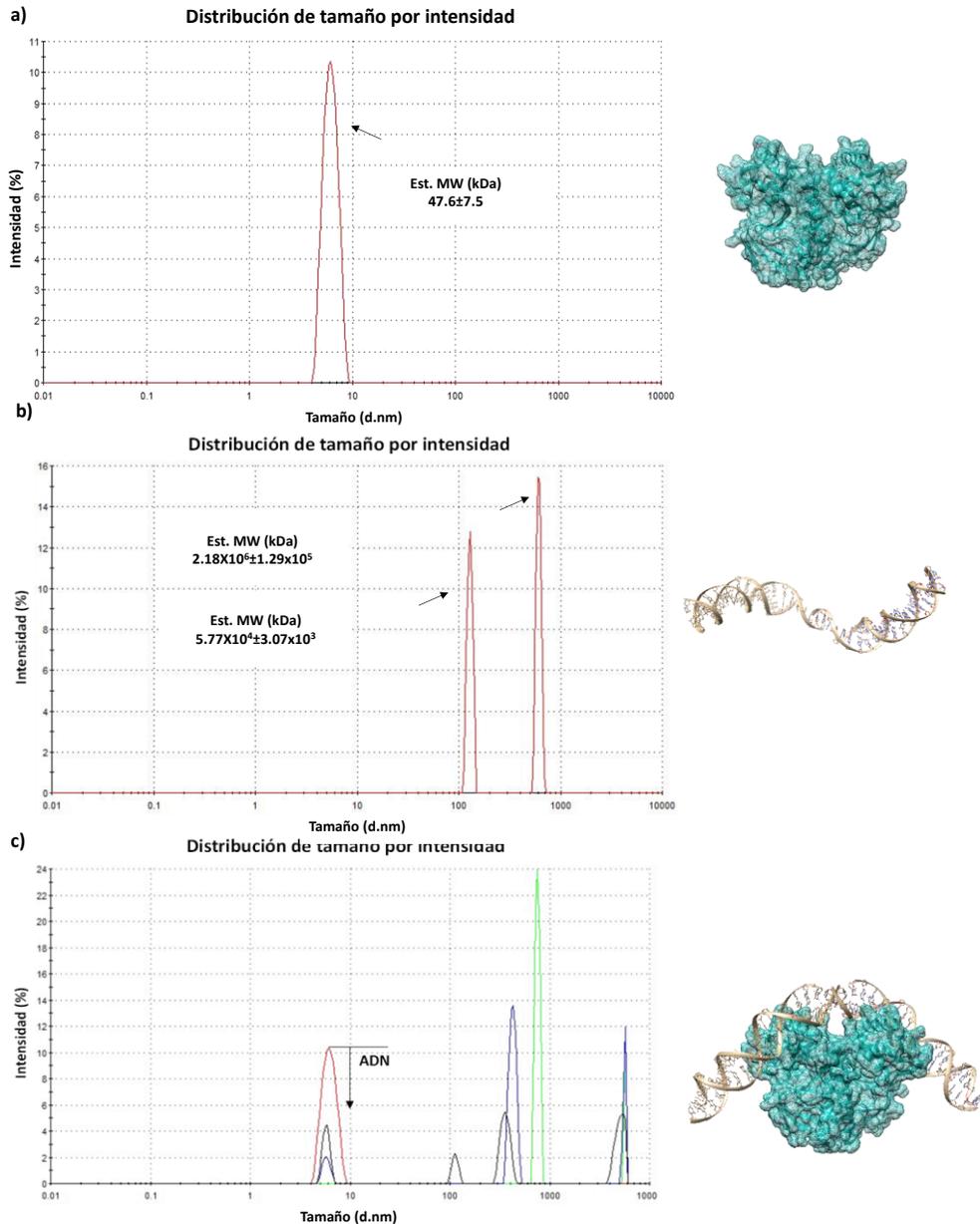
**Figura 8.** Expresión y purificación de la proteína SdrP

a) Esquema del vector pET28-PPS/SdrP. b) Gel al 12% para evaluar la expresión de SdrP, donde se observa el marcador (M), el tiempo cero (S/I) de los matraces 1 y 2, así como el tiempo inducido de cada uno (M1, M2). c) Gel al 12% de la purificación por cromatografía de afinidad a níquel, se observa la elución de SdrP en su mayoría en 200mM de imidazol. d) Gel al 12% para evaluar el corte de la cola de histidinas con la proteasa PPS. e) Cromatograma de la columna HiTrap Q, donde la proteína eluyó entre 180-280mM de NaCl. f) Gel al 12% con las fracciones eluidas de la columna HiTrap Q para corroborar el estado de la proteína.

**Interacción de CAP y SdrP con los secuencias LacNat y CONCAP por dispersión dinámica de luz (DLS).** En este ensayo se analizaron dos controles, la proteína SdrP y ADN por separado, así como una concentración fija de SdrP en presencia de tres concentraciones distintas del ADN CONCAP. Al medir SdrP en ausencia de ADN se obtiene un solo pico cuyo diámetro promedio es de  $6.205 \pm 0.9820$  nm, lo que corresponde a un peso molecular de  $47.6 \pm 7.5$  kDa, esto coincide con el peso molecular esperado para SdrP que es de 44 kDa (Fig. 9a). Por otro lado, al realizar el experimento con el ADN en ausencia de proteína, se observan dos picos con diámetros promedio entre  $129 \pm 7.671$  nm y  $608.8 \pm 32.27$  nm, lo que corresponde a  $5.77 \times 10^4 \pm 3.07 \times 10^3$  kDa y  $2.18 \times 10^6 \pm 1.29 \times 10^5$  kDa respectivamente (Fig. 9b). Sin embargo, al analizar la proteína SdrP en presencia de ADN, se observa que la intensidad de SdrP disminuye considerablemente al incrementar la concentración de ADN en solución, hasta el punto de desaparecer el pico de la proteína, a la par aparecen picos de mayor tamaño que incrementan su intensidad (Fig. 9c).

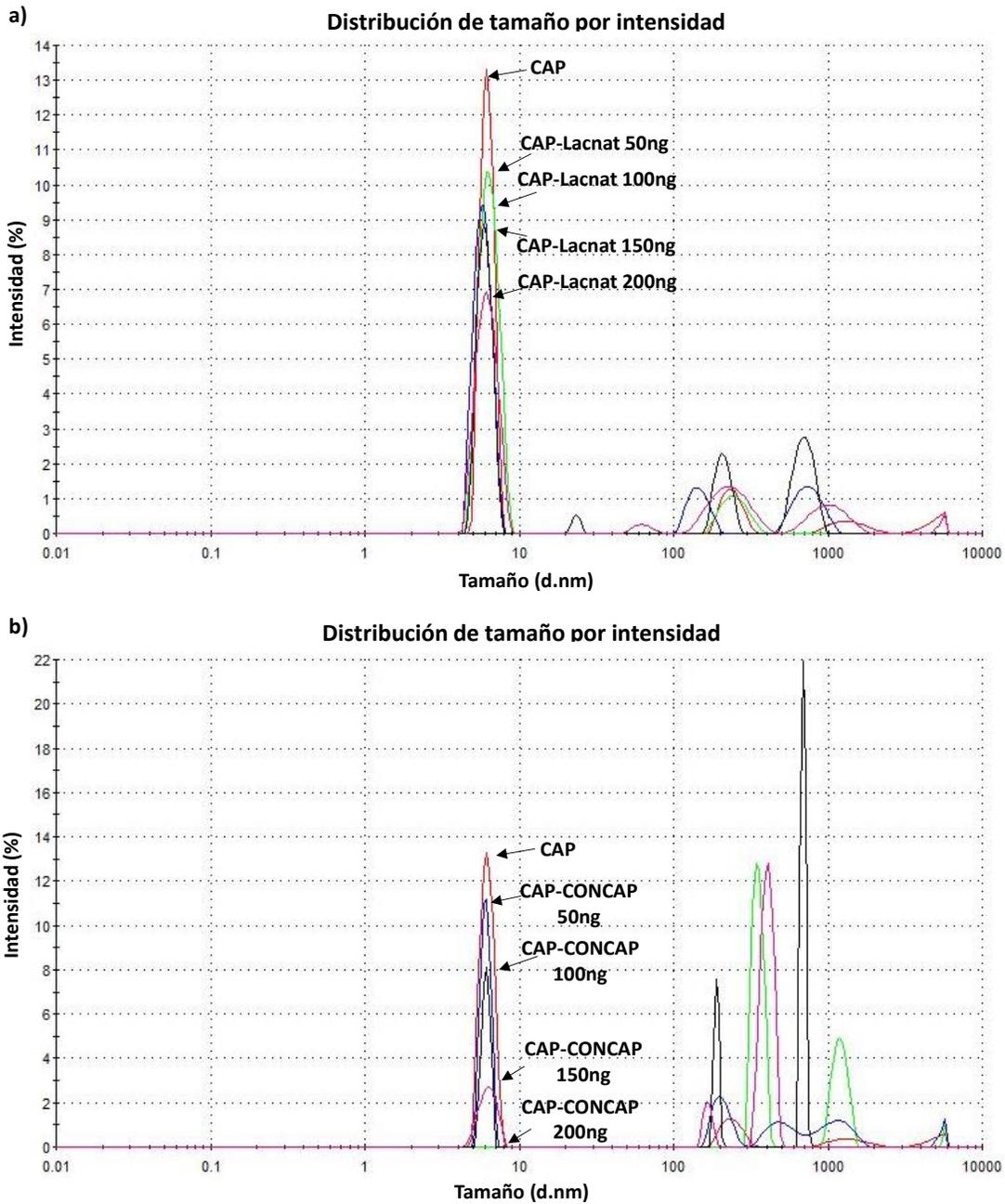
Al realizar el análisis de la proteína CAP inactiva en ausencia de ADN, se observa un único pico cuyo diámetro promedio fue de  $5.890 \pm 0.7513$  nm, lo que equivale a un peso molecular a  $42.2 \pm 4.5$  kDa, esto va de acuerdo al peso molecular esperado de CAP que es 45 kDa. Con el complejo CAP-LacNat en ausencia de AMPc se observa que el pico de la proteína mantiene su diámetro promedio respecto al control sin ADN y conforme se incrementa la concentración de ADN la intensidad de la proteína en un 47 % (Fig. 10a). Sin embargo, al realizar el mismo ensayo con el complejo CAP-CONCAP sin AMPc, la intensidad de la proteína se ve disminuida de 62 a 85 % entre los 100 y 150 ng de ADN (Fig. 10b).

De igual forma, se repitió este ensayo en presencia de concentraciones saturantes de AMPc ( $300 \mu\text{M}$ ), dejando a la proteína CAP en su forma activa en presencia de sus sitios de interacción LacNat y CONCAP a una concentración de 25 y 50 ng totales, donde se observó una disminución de la intensidad del pico correspondiente a la proteína a una menor concentración, siendo para la concentración de 25 ng totales de LacNat una disminución del ~80 % en los primeros 30 minutos de interacción proteína-ADN (Fig. 11a) y para 50 ng totales un decremento de la intensidad del 100 % en los primeros 10 minutos de medición



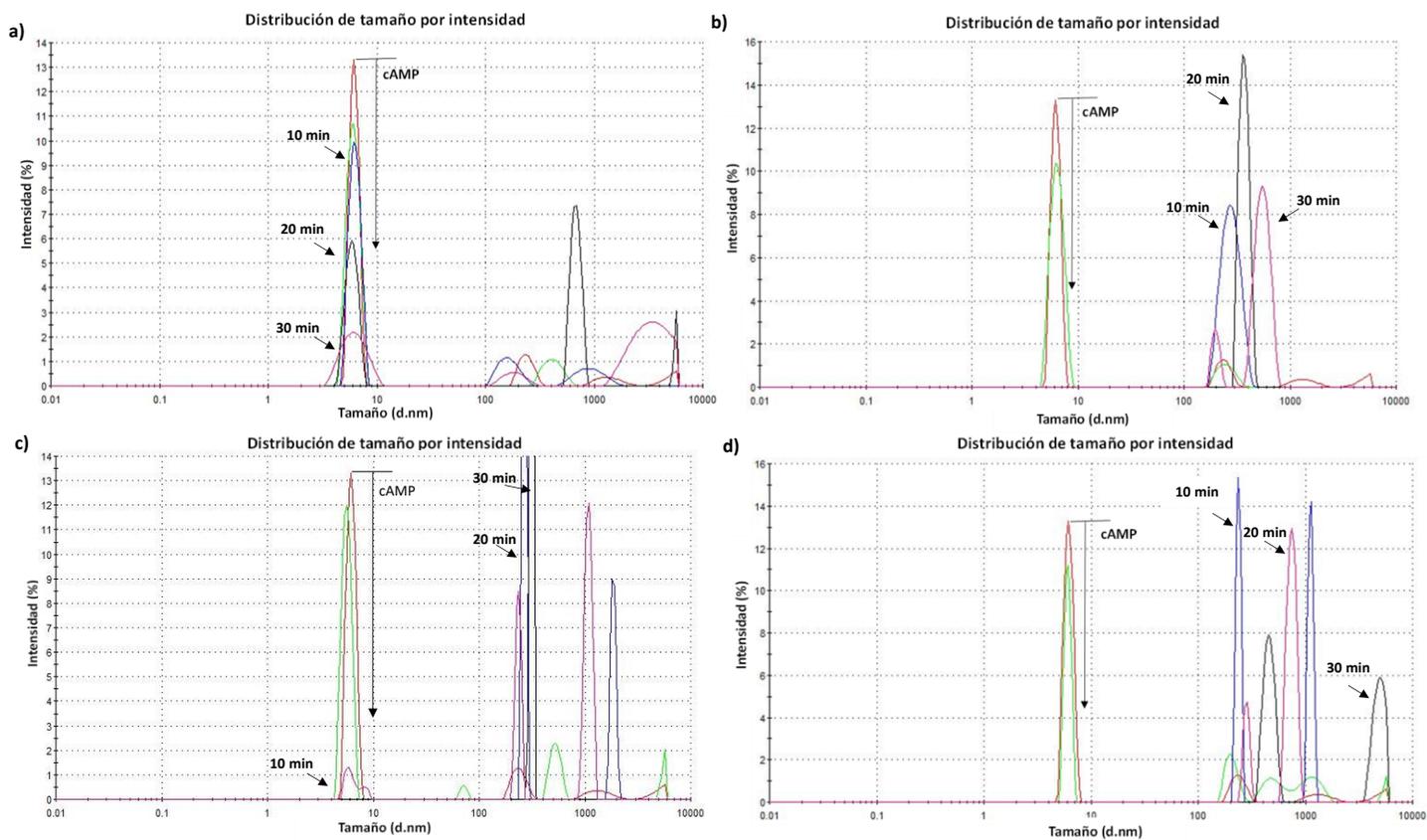
**Figura 9.** Dispersión dinámica de luz (DLS) del complejo SdrP-CONCAP

a) Gráfico de dispersión dinámica de luz con 1mg/mL de la proteína SdrP a temperatura ambiente, el modelo representa a SdrP dimérica, el modelo corresponde a la estructura cristalizada de SdrP (2ZCW). b) Gráfico obtenido del ADN CONCAP1 (800pb) en ausencia de proteína; el modelo obtenido en CHIMERA representa el dsADN con la secuencia consenso al centro. c) Gráfico obtenido con el complejo SdrP-CONCAP a concentraciones crecientes de ADN (200ng, 600ng, 4000ng). En rojo se indica la proteína en ausencia de ADN, en gris SdrP con 200ng totales de ADN, en azul SdrP con 600ng totales de ADN y en verde SdrP con 4000ng totales de ADN. El modelo obtenido en CHIMERA representa la interacción entre la proteína SdrP y el ADN CONCAP.



**Figura 10.** DLS de los complejos CAP-LacNat y CAP-CONCAP sin cAMP

a) Gráfico del complejo CAP-LacNat a 25°C con la proteína inactiva en presencia de concentraciones crecientes de ADN, la línea roja muestra la proteína sin ADN, la línea verde indica una concentración de 50 ng, la azul de 100 ng, el negra 150 ng y el rosa 200 ng totales b) Gráfico del complejo CAP-CONCAP a 25°C con la proteína inactiva en presencia de concentraciones crecientes de ADN, la línea roja muestra a CAP sin ADN, en azul se observa una concentración de 50 ng, en negro de 100 ng, en rosa 150 ng y en verde 200 ng totales.



**Figura 11.** DLS de los complejos CAP-LacNat y CAP-CONCAP con cAMP

a) Gráfico del complejo CAP-LacNat con 25ng totales de ADN, donde la línea roja muestra la intensidad de la proteína sola, la línea verde el complejo CAP-LacNat con CAP inactiva, las líneas azul, negra y rosa indican el complejo con la proteína activada por cAMP. b) Gráfico del complejo CAP-LacNat con 50 ng finales de ADN, la línea roja muestra la intensidad de la proteína sola, la línea verde el complejo CAP-LacNat con proteína inactiva, las líneas azul, negra y rosa indican el complejo con la proteína activada. c) Gráfico del complejo CAP-CONCAP con 25 ng totales de ADN, donde la línea roja muestra la intensidad de la proteína sola, la verde el complejo CAP-CONCAP con proteína inactiva, las líneas rosa, azul y negra muestran el complejo con CAP activa. d) Gráfico del complejo CAP-CONCAP con 50 ng totales de ADN, donde la línea roja muestra la intensidad de la proteína sola, la verde el complejo CAP-CONCAP con proteína inactiva, las líneas azul, rosa y negra muestran el complejo con CAP activa.

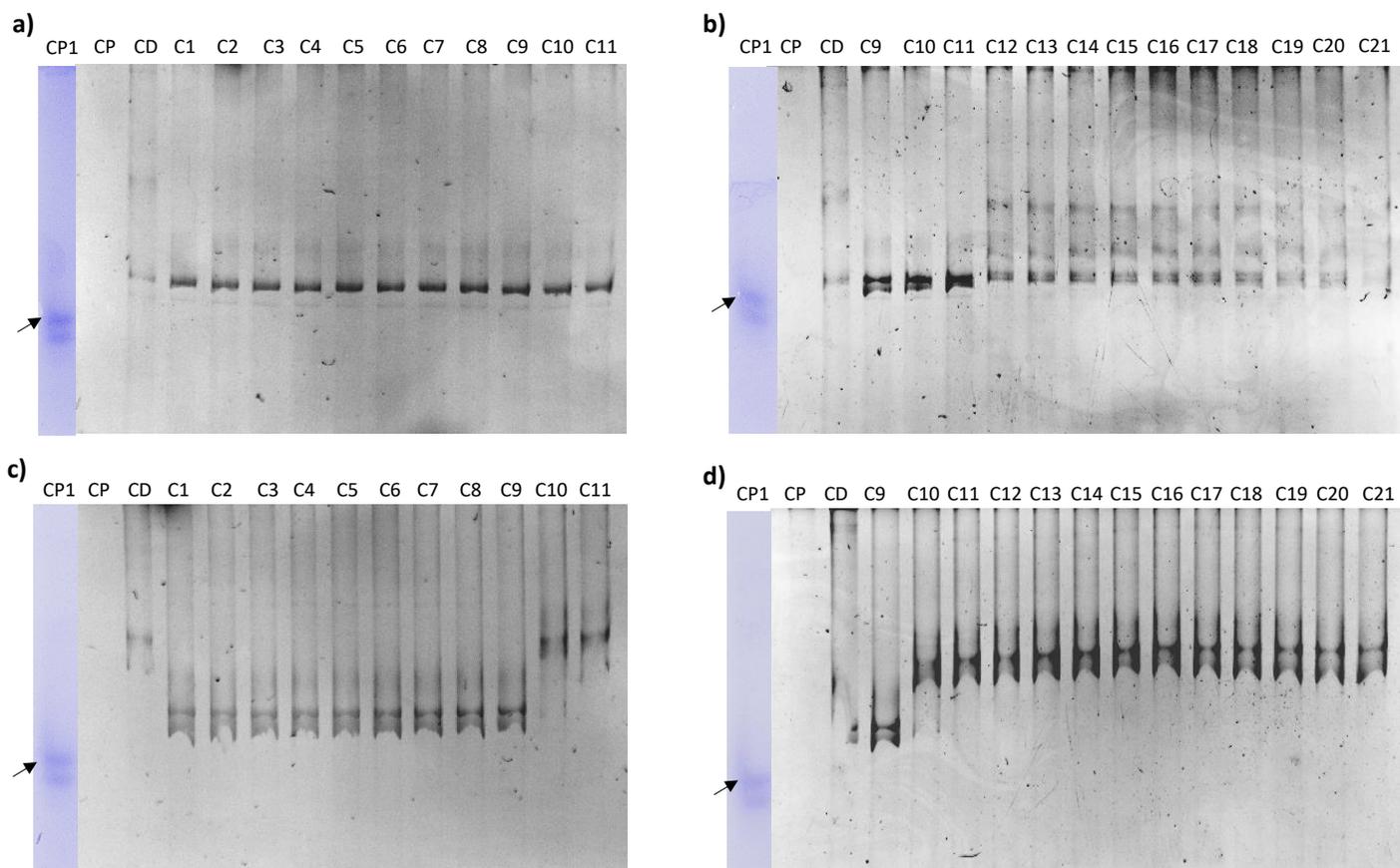
(Fig. 11b). Para el complejo CAP-CONCAP activo, con 25 ng totales de ADN se observó una reducción del 95 % de la intensidad del pico de la proteína en los primeros 10 minutos (Fig. 11c), y con 50 ng totales este cambio fue inmediato (Fig. 11d).

**Ensayo de cambio en la movilidad electroforética del ADN (EMSA).** Los resultados obtenidos mostraron que para el ADN LacNat el cambio en la movilidad fue evidente entre 31.62 ng (0.295 nM) y 126.48 ng (1.18 nM) de proteína que corresponden a los carriles C10 y C13 (Fig. 12a y 12b). Sin embargo, el patrón de bandeado obtenido no permitió dilucidar si esta estimación era la adecuada puesto que se obtuvieron múltiples bandas. Mientras que para el ADN CONCAP el retardo en la movilidad fue a partir de los 12.64 ng (0.118 nM), carril C7, con el punto de saturación cerca de 31.62 ng (0.295 nM), carril C10, de proteína (Fig. 12c y 12d). Así mismo fueron calculados los Rf en cada imagen de los complejos, obteniendo con ello una curva de interacción para cada uno (Fig. 13a), donde se aprecia un incremento notorio en la curva de SdrP-LacNat a partir de 0.295 nM de proteína mientras que para SdrP-CONCAP este cambio se da en 0.118 nM, posterior a ello, se realizó un ajuste preliminar con la ecuación de Boltzmann (Fig. 13b),

$$I = A1 + \frac{(A2-A1)}{1+e^{(x-x_0)/dx}} \quad 41$$

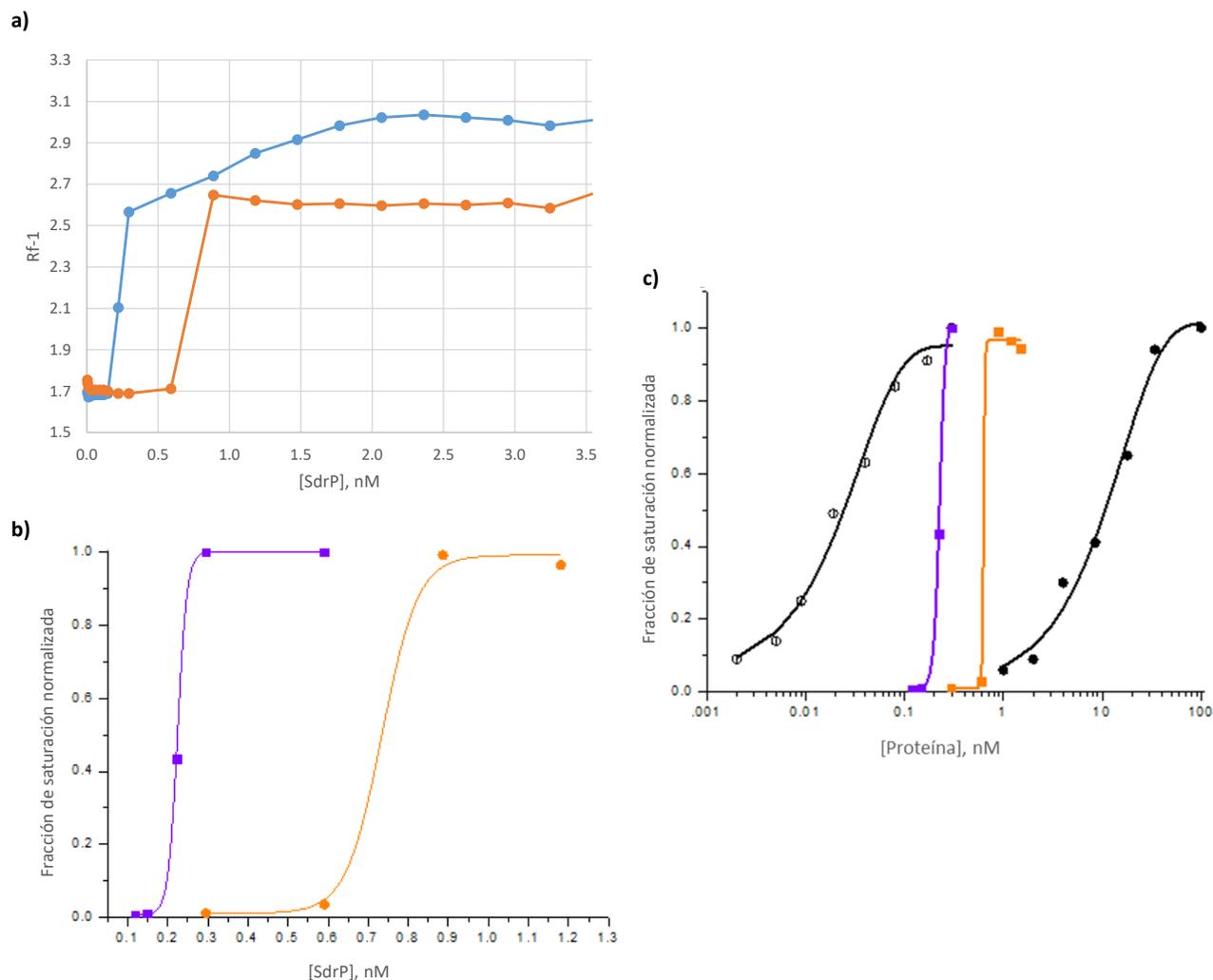
con la cual se obtuvo una constante de disociación estimada para la secuencia CONCAP de 0.2269 nM y para la secuencia LacNat de 0.7347 nM.

**Análisis bioinformático de ortólogos putativos de la proteína Clp.** Con el propósito de identificar posibles secuencias ortólogas de la proteína Clp, se realizó un análisis bioinformático de las secuencias relacionadas depositadas en la base de datos National Center for Biotechnology Information (NCBI). En la búsqueda inicial se obtuvieron 20, 712 secuencias reportadas hasta el 14 de enero del 2017. De las secuencias obtenidas sólo se seleccionamos aquellas provenientes de organismos del orden *Xanthomadales* por ser las más cercanas a la especie de interés (*X. campestris*), reduciendo con ello a 626 secuencias. A partir de estas últimas, se acotó la búsqueda seleccionando aquellas secuencias con un tamaño de  $230 \pm 5$  residuos aminoacídicos, resultando en 548 secuencias. De éstas se



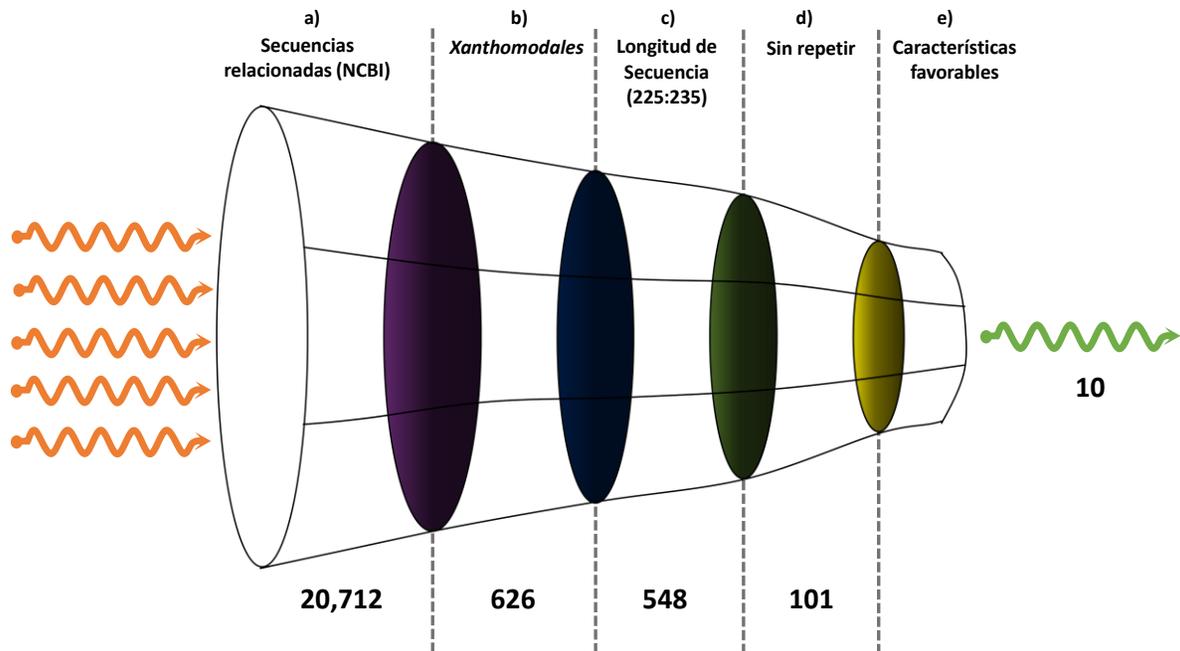
**Figura 12.** Ensayo de cambio en la movilidad electroforética del ADN

a) Gel al 5%, donde se evalúa la movilidad del complejo SdrP-LacNat, donde CP1 es el control con proteína teñido con azul de Comassie, CP es el control de proteína, CD es el control de ADN, C1 a C11 indican las concentraciones crecientes de proteína empleada (Tabla 2) teñido con bromuro de etidio. b) Gel al 5 % donde se observa la movilidad del complejo SdrP-LacNat con concentraciones mayores de proteína c) Gel al 5 %, se evalúa la movilidad del complejo SdrP-CONCAP, donde donde CP1 es el control con proteína teñido con azul de Comassie, CP es el control de proteína, CD es el control de ADN, C1 a C11 indican las concentraciones crecientes de proteína empleada teñido con bromuro de etidio. d) Gel al 5% se observa la movilidad del complejo SdrP-CONCAP con concentraciones mayores de proteína. Las flechas señalan la proteína dimérica.



**Figura 13.** Análisis de los complejos SdrP-LacNat y SdrP-CONCAP

a) Comparación de las curvas de interacción de los complejos SdrP-LacNat y SdrP-CONCAP, donde se muestra en naranja la curva de SdrP-LacNat y en azul SdrP-CONCAP obtenidas con el inverso del Rf de los geles de EMSA. Los análisis de las imágenes fueron realizados en el software Image Lab 6 de Bio Rad b) Ajuste de los datos obtenidos en EMSA a la ecuación de Boltzmann, en azul se muestra la curva de SdrP-CONCAP, y en naranja Sdrp-LacNat. El ajuste fue realizado con OriginPro 8. c) Gráfico de comparación las constantes reportadas por Ebright y col. de los complejos CAP-CONCAP (⊕) y CAP-LacNat (●) contra las constantes preliminares obtenidas para los complejos SdrP-CONCAP (■) y SdrP-LacNat (■).



**Figura 14.** Parámetros de selección de secuencias homólogas de Clp

a) Secuencias relacionadas con la secuencia de Clp de *Xanthomonas campestris* (3IWZ) de la base de datos National Center for Biotechnology Information (NCBI). b) Primer criterio de selección donde se redujo la búsqueda a solo secuencias pertenecientes al orden *Xanthomodales*. c) Selección de secuencias por tamaño con un rango de  $\pm 10$  de la secuencia original (230pb). d) Filtro para descartar secuencias repetidas. e) Determinación de secuencias con características favorables para la estabilidad, purificación y cristalización de la proteína mediante análisis con softwares online.

eliminaron las secuencias repetidas teniendo con ello un total de 101 secuencias candidatas (Fig. 14).

Los resultados obtenidos por el análisis de las secuencias mostraron que 100 de las secuencias son solubles, 94 presentaban un bajo índice de agregación, 31 tienen todos los aminoácidos característicos conservados, 25 presentaban un solo cambio, 10 presentaban dos cambios, y el resto 3 o más cambios.

Con estos datos, se seleccionaron las secuencias que cumplieron con los siguientes criterios. Que presenten todos los aminoácidos característicos conservados; que tengan un score de éxito igual o mayor a 0.54 en los servidores Crysalis I y II; así como, secuencias marcadas como solubles en SPpred y con un bajo índice de agregación (menor o igual que -3) en Aggrescan, entre otros (Tabla 3). Con ello, se lograron obtener 10 secuencias que presentaban características más favorables que la secuencia de referencia de Clp.

## DISCUSIÓN

Durante la construcción de los vectores en tándem se tuvieron complicaciones puesto que al ser secuencias grandes que son totalmente complementarias entre sí, se presentaron dificultades al momento de clonar los vectores dentro de las células hospederas<sup>42</sup>, dando como resultado la obtención de vectores incompletos o vacíos. Este problema fue solucionado al emplear células calcio-competentes de una cepa de *E. coli* Top10, la cual tiene un genotipo F- *mcrA*  $\Delta(mrr-hsdRMS-mcrBC)$   $\Phi80/lacZ\Delta M15$   $\Delta lacX74$  *recA1* *araD139*  $\Delta(ara\ leu)$  7697 *galU galK rpsL* (StrR) *endA1 nupG*, que presenta una mutación en el gen *recA* responsable de la recombinación, permitiendo insertar múltiples copias repetidas dentro del vector deseado.

Una vez construidos los vectores con los sitios de unión (Fig. S1 y S2) y con las proteínas purificadas, se procedió con los ensayos preliminares de interacción, iniciando con el ensayo de dispersión dinámica de luz, el cual nos muestra el peso estimado de las proteínas CAP y SdrP, el cual correspondió con el peso teórico de 45 y 44 kDa respectivamente. Sin embargo, el peso molecular teórico para el ADN (486kDa) y el determinado por el equipo ( $5.77 \times 10^4 \pm 3.07 \times 10^3$  kDa y  $2.18 \times 10^6 \pm 1.29 \times 10^5$  kDa) fueron significativamente distinto, entre 100-4000 veces mayor.

En el caso de dispersión dinámica de luz., el equipo estima a partir del movimiento browniano de las moléculas en solución un radio hidrodinámico considerando que las moléculas son globulares. Para las proteínas esta técnica es ampliamente usada, debido a que el peso molecular calculado por el equipo es muy cercano al peso teórico calculado, y permite determinar la dispersión de la molécula, es decir, si se encuentra mono o polidispersa<sup>43,44</sup>. Sin embargo, cuando se analiza el ADN, que es una molécula elongada no globular, el equipo interpreta su radio hidrodinámico como una molécula de mayor tamaño. Como el ADN en solución presenta fuerzas intermoleculares que permiten la interacción entre diferentes segmentos de distintas cadenas, formando hebras más largas, el equipo de DLS interpreta este conjunto de moléculas, como una sola hebra de mayor tamaño y por tanto estima un peso molecular mayor al teórico<sup>45,46</sup>. A pesar de lo anterior, este ensayo nos es de utilidad puesto que permite observar la dinámica de la interacción

entre las proteínas y el ADN al comparar la intensidad de la señal emitida por la proteína y por el complejo.

En el complejo SdrP-CONCAP (Fig. 10c) se observa claramente como a mayor concentración de ADN la intensidad de la proteína disminuye y se va incrementando la intensidad de los picos correspondientes al ADN, lo que nos indica que la proteína reconoce su sitio de interacción en el ADN y queda reclutado por este último, razón por la cual, el pico de la proteína va perdiendo intensidad.

En el caso de los complejos CAP-LacNat y CAP-CONCAP, cuando la proteína se encuentra inactiva, el cambio de intensidad de CAP en complejo con LacNat corresponde a una disminución del 47% de la intensidad, mientras que en CONCAP esta disminución es del 100% en la concentración más alta probada. Esta interacción “inespecífica” que observamos de la proteína inactiva (sin AMPc) con el ADN, se puede explicar analizando la estructura cristalográfica del complejo CAP/ADN (1O3Q). En la estructura se observa que los residuos Lys22, Lys26, Lys44, Lys166, His199 y Lys201 presentan interacciones electrostáticas con el ADN y se han propuesto como responsables del segundo dobles en el ADN provocado por CAP o “secondary kink”<sup>13</sup>. Por lo tanto, estos residuos son los responsables de las interacciones inespecíficas observadas, ya que esta interacción electrostática no depende del cambio alostérico desencadenado por la unión del AMPc, en el cual se reposiciona la hélice F. De igual forma, al existir un mayor número de moléculas en solución debido al incremento de la concentración del ADN, las interacciones inespecíficas entre la proteína y el ADN incrementan.

Es probable que, esta interacción no haya sido observada anteriormente, debido a que la concentración máxima empleada para estos análisis fue de 100 nM de CAP<sup>16</sup>, mientras que en los experimentos de DLS empleamos una concentración mayor de proteína (0.8 mg/mL), equivalente a 16,920 nM de CAP.

Respecto a los complejos CAP-LacNat y CAP-CONCAP activados con AMPc, se observa como la concentración de ADN requerida para tener un efecto significativo en la disminución de la intensidad de CAP se ven reducidas ~4 veces para el caso de LacNat y ~10 veces para CONCAP, lo que indica una afinidad mucho mayor por la secuencia CONCAP que por LacNat, lo cual concuerda con lo reportado por Ebright y col. en 1989<sup>16</sup>, sin embargo, no es posible determinar si la

afinidad es similar a la reportada, puesto que en este ensayo se emplean concentraciones muy altas tanto de proteína como de ADN.

El EMSA nos permitió comparar el comportamiento de SdrP al interactuar con LacNat y CONCAP, donde se observó que los complejos siguen la misma tendencia a la observada en los complejos CAP-LacNat y CAP-CONCAP en DLS. Así mismo, el ajuste con la ecuación de Boltzmann realizado con los datos obtenidos, nos permitió calcular constantes de disociación preliminares ( $K_d$ ) para ambas secuencias, sin embargo, para conseguir una mejor estimación de la misma, se necesitan probar concentraciones intermedias de la proteína para ampliar los puntos dentro de la curva de transición y obtener un mejor ajuste a la ecuación. Se sabe que la proteína CAP tiene una  $K_d$  por CONCAP de 0.02 nM y por LacNat de 10.75 nM, por lo que la estimación preliminar de la  $K_d$  de SdrP-CONCAP es de 0.23 nM, es 10 veces menor a lo reportado para CAP-CONCAP. Por el contrario, para SdrP-LacNat la  $K_d$  calculada es de 0.74 nM, siendo una afinidad mayor a la reportada. Anteriormente, ha sido reportado que CAP es 450 veces más afín por la secuencia CONCAP que por LacNat<sup>16</sup>, si bien, SdrP tiende a ser más afín por la secuencia CONCAP, la diferencia de afinidad entre ambas secuencias es apenas de 4 veces. Esto puede ser debido a que la secuencia LacNat no es la secuencia apropiada para este estudio, debido a que no es un sitio de unión nativo para SdrP, los cuales ya han sido reportados con anterioridad<sup>33</sup> o a que residuos particulares presentes en la interfase de interacción de SdrP con el ADN propician una mayor afinidad hacia este. Sin embargo, para poder determinar con exactitud el motivo de estas diferencias se requieren de más estudios. Cabe destacar que no existen reportes de la constante de disociación para la proteína SdrP, por lo que este ensayo preliminar sirve como referencia para ensayos posteriores que permitan evaluar nuevamente este resultado.

Los resultados obtenidos durante los análisis de interacción proteína-ADN con DLS y EMSA, muestran que las proteína CAP y SdrP son capaces de interactuar con el ADN CONCAP así como con LacNat, lo que indica que tanto las proteínas purificadas como el ADN construido son funcionales. Por tanto, es posible comenzar a preparar los estudios de la interacción de estas proteínas con el ADN construido en herramientas de molécula única, como son las pinzas ópticas.

De las tres proteínas implicadas en este estudio, Clp es la única que no fue posible expresar y purificar, esto debido a que el rendimiento obtenido previamente en nuestro laboratorio es muy bajo<sup>30,31</sup>, y no permite realizar los ensayos preliminares de interacción. Una manera de solucionar este problema es el uso de proteínas ortólogas que permitan llevar a cabo este proceso de forma más eficiente, para ello se llevó a cabo un análisis bioinformático con el fin de encontrar secuencias homologas con características más favorables para la expresión, purificación y cristalización que la secuencia de referencia<sup>47</sup>. Con los resultados preliminares obtenidos se puede sugerir la selección de las secuencias *X. albilineans* GPE PC73, *X. vesicatoria* ATCC 35937 y *X. cannabis* pv. *cannabis*. Esto debido a que *X. albilineans* GPE PC73 la predicción en Crysalis I y II fue superior a la de la secuencia de referencia y presenta un índice de agregación similar, mientras que en el caso de las secuencias *X. vesicatoria* ATCC 35937 y *X. cannabis* pv. *cannabis* las predicciones en Crysalis I y II fue similar a la referencia, sin embargo, los índices de agregación para ambas secuencias son mucho menores. Sin embargo, se requieren de otros análisis, como la dinámica molecular con los modelos construidos por homología de estas secuencias, para confirmar que la selección fue la adecuada, y probar estas proteínas a nivel laboratorio.

## REFERENCIAS

1. Acosta-viana KY, Zavala-castro JE (1996) Proteínas de unión a DNA . Rev. Biomed 7:163–172.
2. Raval A, Ray A Introduction to Biological Networks. In: CRC Press; 2013. pp. 12–20.
3. Stormo GD, Zhao Y (2010) Determining the specificity of protein–DNA interactions. Nat. Rev. Genet. [Internet] 11:751–760. Available from: <http://www.nature.com/doi/10.1038/nrg2845>
4. Le S, Chen H, Cong P, Lin J, Dro P, Yan J (2013) Mechanosensing of DNA bending in a single specific protein-DNA complex.
5. Goodman SD, Nash HA (1989) Functional replacement of a protein-induced bend in a DNA recombination site. Nature 341:251–254.
6. Ishihama A (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. 34:628–645.
7. Seshasayee A, Sivaraman K, Luscombe N (2011) An overview of prokaryotic transcription factors : a summary of function and occurrence in bacterial genomes. Subcell Biochem. [Internet] 52:50–52. Available from: <http://www.springerlink.com/index/10.1007/978-90-481-9069-0>
8. Rivera-Gómez N, Segovia L, Pérez-Rueda E (2011) Diversity and distribution of transcription factors: Their partner domains play an important role in regulatory plasticity in bacteria. Microbiology 157:2308–2318.
9. Shaw DJ, Rice DW, Guest JR (1983) Homology between CAP and Fnr, a regulator of anaerobic respiration in Escherichia coli. J. Mol. Biol. [Internet] 166:241–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6343617>
10. Zhou A, Chen YI, Zane GM, He Z, Hemme CL, Joachimiak MP, Baumohl JK, He Q, Fields MW, Arkin AP, et al. (2012) Functional characterization of Crp/Fnr-type global transcriptional regulators in Desulfovibrio vulgaris hildenborough. Appl. Environ. Microbiol. 78:1168–1177.
11. Körner H, Sofia HJ, Zumft WG (2003) Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: Exploiting the metabolic spectrum by controlling alternative gene programs. FEMS Microbiol. Rev. 27:559–592.
12. Matsui M, Tomita M, Kanai A (2013) Comprehensive computational analysis of bacterial CRP/FNR super family and its target motifs reveals stepwise evolution of transcriptional networks. Genome Biol. Evol. 5:267–282.
13. Lawson CL, Swigon D, Murakami KS, Darst SA, Berman HM, Ebricht RH (2004) Catabolite activator protein: DNA binding and transcription activation. Curr. Opin. Struct. Biol. 14:10–20.
14. Ishihama A, Shimada T, Yamazaki Y (2016) Transcription profile of Escherichia coli: Genomic SELEX search for regulatory targets of transcription factors. Nucleic Acids Res. 44:2058–2074.
15. Botsford JL, Harman JG (1992) Cyclic AMP in prokaryotes. Microbiol Rev [Internet] 56:100–122. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1315922><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC372856><http://www.ncbi.nlm.nih.gov/pubmed/1315922><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC372856><http://www.ncbi.nlm.nih.gov/pubmed/1315922><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC372856>
16. Ebricht RH, Ebricht YW, Gunasekera A (1989) Consensus DNA site for the Escherichia coli catabolite gene activator protein (CAP): CAP exhibits a 450-fold higher affinity for the consensus DNA site than for the E. coli lac DNA site. Nucleic Acids Res. 17:10295–10305.
17. Popovych N, Tzeng S-R, Tonelli M, Ebricht RH, Kalodimos CG (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. Proc. Natl. Acad. Sci. U. S. A. 106:6927–6932.
18. Parkinson G, Wilson C, Gunasekera A, Ebricht YW, Ebricht RH, Berman HM (1996)

Structure of the CAP-DNA complex at 2.5 Å resolution: A complete picture of the protein-DNA Interface. *J. Mol. Biol.* 260:395–408.

19. Fried MG, Liu G (1994) Molecular sequestration stabilizes CAP-DNA complexes during polyacrylamide gel electrophoresis. *Nucleic Acids Res.* 22:5054–5059.

20. Crothers DM, Haran TE, Nadeau JG (1990) Intrinsically bent DNA. *J. Biol. Chem.* 265:7093–7096.

21. He YW, Boon C, Zhou L, Zhang LH (2009) Co-regulation of *Xanthomonas campestris* virulence by quorum sensing and a novel two-component regulatory system RavS/RavR. *Mol. Microbiol.* 71:1464–1476.

22. Hsiao YM, Liao HY, Lee MC, Yang TC, Tseng YH (2005) Clp upregulates transcription of *engA* gene encoding a virulence factor in *Xanthomonas campestris* by direct binding to the upstream tandem Clp sites. *FEBS Lett.* 579:3525–3533.

23. Leduc JL, Roberts GP (2009) Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis* pv. *citri*. *J. Bacteriol.* 191:7121–7122.

24. Chin KH, Lee YC, Tu Z Le, Chen CH, Tseng YH, Yang JM, Ryan RP, McCarthy Y, Dow JM, Wang AHJ, et al. (2010) The cAMP Receptor-Like Protein CLP Is a Novel c-di-GMP Receptor Linking Cell-Cell Signaling to Virulence Gene Expression in *Xanthomonas campestris*. *J. Mol. Biol.* 396:646–662.

25. He Y-W, Ng AY-J, Xu M, Lin K, Wang L-H, Dong Y-H, Zhang L-H (2007) *Xanthomonas campestris* cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Mol. Microbiol.* [Internet] 64:281–292. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17378922>

26. deCrecy-Lagard V, Glaser P, Lejeune P, Sismeiro O, Barber CE, Daniels MJ, Danchin A (1990) A *Xanthomonas campestris* pv. *campestris* Protein Similar to Catabolite Activation Factor Is Involved in Regulation of Phytopathogenicity. *J. Bacteriol.* 172:5877–5883.

27. Tao F, He Y-W, Wu D-H, Swarup S, Zhang L-H (2010) The Cyclic Nucleotide Monophosphate Domain of *Xanthomonas campestris* Global Regulator Clp Defines a New Class of Cyclic Di-GMP Effectors. *J. Bacteriol.* [Internet] 192:1020–1029. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20008070>

28. Zhang L-H (2010) A novel C-di-GMP effector linking intracellular virulence regulon to quorum sensing and hypoxia sensing. *Virulence* [Internet] 1:391–394. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21178475>

29. Dong Q, Ebricht RH (1992) DNA binding specificity and sequence of *Xanthomonas campestris* catabolite gene activator protein-like protein. *J. Bacteriol.* 174:5457–5461.

30. Martínez-García E, Lara-González S (2015) Análisis funcional y estructural de *Xanthomonas campestris* pv. *campestris*.

31. Pérez-Márquez A, Lara-González S (2016) Análisis estructural y funcional de la proteína Clp de *Xanthomonas axonopodis*.

32. Agari Y, Kashiwara A, Yokoyama S, Kuramitsu S, Shinkai A (2008) Global gene expression mediated by *Thermus thermophilus* SdrP, a CRP/FNR family transcriptional regulator. *Mol. Microbiol.* 70:60–75.

33. Agari Y, Kuramitsu S, Shinkai A (2010) Identification of novel genes regulated by the oxidative stress-responsive transcriptional activator SdrP in *Thermus thermophilus* HB8. *FEMS Microbiol. Lett.* 313:127–134.

34. Agari Y, Kuramitsu S, Shinkai A (2012) 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 Struct. Funct. Bioinforma.* 80:1490–1494.

35. Gamboa-Rodríguez KE, Lara-González S (2015) Estudio estructura función del activador transcripcional SdrP de *Thermus thermophilus*.

36. Min W, English BP, Luo G, Cherayil BJ, Kou SC, Xie XS (2005) Fluctuating enzymes: Lessons from single-molecule studies. *Acc. Chem. Res.* 38:923–931.

37. Gebhardt JCM, Suter DM, Roy R, Zhao ZW, Chapman AR, Basu S, Maniatis T, Xie XS (2013) Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* [Internet] 10:421–426. Available from: <http://www.nature.com/doi/10.1038/nmeth.2411>
38. Zhang Z, English BP, Grimm JB, Kazane SA, Hu W, Tsai A, Inouye C, You C, Piehler J, Schultz PG, et al. (2016) Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis. *Genes Dev.* 30:2106–2118.
39. Nalefski EA, Nebelitsky E, Lloyd JA, Gullans SR (2006) Single-molecule detection of transcription factor binding to DNA in real time: Specificity, equilibrium, and kinetic parameters. *Biochemistry* 45:13794–13806.
40. Bustamante C, Smith SB, Liphardt J, Smith D (2000) Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.* [Internet] 10:279–285. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10851197>
41. Ericsson UB, Hallberg BM, DeTitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* [Internet] 357:289–298. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0003269706005355>
42. Bzymek M, Lovett ST (2001) Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* 98:8319–8325.
43. Tanner RE, Herpigny B, Chen S -H., Rha CK (1982) Conformational change of protein sodium dodecylsulfate complexes in solution: A study of dynamic light scattering. *J. Chem. Phys.* [Internet] 76:3866–3872. Available from: <http://aip.scitation.org/doi/10.1063/1.443528>
44. Jachimska B, Wasilewska M, Adamczyk Z (2008) Characterization of Globular Protein Solutions by Dynamic Light Scattering, Electrophoretic Mobility, and Viscosity Measurements. *Langmuir* [Internet] 24:6866–6872. Available from: <http://pubs.acs.org/doi/abs/10.1021/la800548p>
45. Sorlie SS, Pecora R (1990) A dynamic light scattering study of four DNA restriction fragments. *Macromolecules* [Internet] 23:487–497. Available from: <http://pubs.acs.org/doi/abs/10.1021/ma00204a022>
46. Pecora R (1991) DNA: A Model Compound for Solution Studies of Macromolecules. *Science* (80- ). 251:893–898.
47. Savchenko A, Yee A, Khachatryan A, Skarina T, Evdokimova E, Pavlova M, Semesi A, Northey J, Beasley S, Lan N, et al. (2003) Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. *Proteins Struct. Funct. Bioinforma.* [Internet] 50:392–399. Available from: <http://doi.wiley.com/10.1002/prot.10282>

**Tabla 1.** Oligonucleótidos empleados en la construcción de los vectores.

<b>Nombre</b>	<b>Secuencia</b>	<b>Tm</b>	<b>pb</b>	<b>%GC</b>
CAP38Fext	5'GGGAATTCCATATGCATGTCCGGTTTTCAACAAAC C3'	62	36	54
910_IntRev	5'GTAAAGCCTGGGGTGCCTAAAATGTGATCTAGAT CACATTTATTGCGTTGCGCTCACTG3'	72	59	46
910_IntFor	5'CAGTGAGCGCAACGCAATAAATGTGATCTAGATC ACATTTTAGGCACCCCAGGCTTTAC3'	72	59	46
CAP38Rext	5'CGCGGATCCAAGCTTAACGGCGGATTGACCGTAA TGGG3'	63	38	58
NEWB	5'CGCGGATCCAAGCTTCCACAGATGAAACGCCGA GTTAACGCC3'	73	42	57
HindIII_4sF	5'CTCCAAGCTTTCGATGCTGGTTGCCAACGATCAG ATGG3'	67	38	55
XhoI_4sR	5'CACGCTCGAGTGCATCTGCCAGTTTGAGGGGAC GACGAC3'	68	39	62
EcoRI_4sF	5'CCGGAATTCGCGATGCTGGTTGCCAACGATCAGA TGG3'	67	37	57
HindIII_4sR	5'CTCCAAGCTTTGCATCTGCCAGTTTGAGGGGACG ACGAC3'	68	39	56

**Tabla 2.** Concentración total de proteína empleada en EMSA

	<b>CP</b>	<b>CD</b>	<b>C1</b>	<b>C2</b>	<b>C3</b>	<b>C4</b>	<b>C5</b>	<b>C6</b>	<b>C7</b>
<b>Proteína</b>	1	-	0.375	0.75	1.158	3.16	6.32	9.48	12.64
<b>ADN</b>	-	160	160	160	160	160	160	160	160
	<b>C8</b>	<b>C9</b>	<b>C10</b>	<b>C11</b>	<b>C12</b>	<b>C13</b>	<b>C14</b>	<b>C15</b>	<b>C16</b>
<b>Proteína</b>	15.8	23.7	31.62	63.24	95	126.48	158.1	189.72	221.34
<b>ADN</b>	160	160	160	160	160	160	160	160	160
	<b>C17</b>	<b>C18</b>	<b>C19</b>	<b>C20</b>	<b>C21</b>				
<b>Proteína</b>	252.96	284.58	316.2	347.8	379.4				
<b>ADN</b>	160	160	160	160	160				

\*La concentración de ADN y proteína esta expresada en ng.

**Tabla 3.** Análisis de las secuencias seleccionadas.

<b>Organismo</b>	<b># aa</b>	<b>%i d</b>	<b>Crysalis I</b>	<b>Crysalis II</b>	<b>Solubilidad</b>	<b>Agregación</b>
<i>X. campestris</i>	230	-	0.578	0.613	Soluble	-4.7
<i>X. campestris</i> (AAA27598.1)	230	99	0.592	0.631	Soluble	-4.7
<i>X. cannabis pv. cannabis</i> (KHL55608.1)	230	99	0.577	0.615	Soluble	-5.5
<i>X. vesicatoria</i> ATCC 35937 (APP76330.1)	230	99	0.577	0.615	Soluble	-5.5
<i>X. oryzae pv. oryzae</i> (OLK94257.1)	230	98	0.578	0.619	Soluble	-4.9
<i>X. albilineans</i> GPE PC73 (87CBA14830.1)	229	87	0.607	0.664	Soluble	-4.2
<i>X. hyacinthi</i> DSM 19077 (KLD77317.1)	229	87	0.592	0.653	Soluble	-3.6
<i>X. translucens pv. translucens</i> (OAX61035.1)	229	87	0.596	0.659	Soluble	-4
<i>S. rhizophila</i> (AHY57369.1)	229	86	0.558	0.541	Soluble	-4.5
<i>X. albilineans</i> (WP_045757591.1)	229	86	0.604	0.671	Soluble	-3.5
<i>S. maltophilia</i> (KIP84382.1)	229	84	0.555	0.559	Soluble	-3

\*La secuencia marcada en gris corresponde a la secuencia de referencia para la búsqueda.

En Crysalis I y II, valores superiores a .5 se consideran exitosos

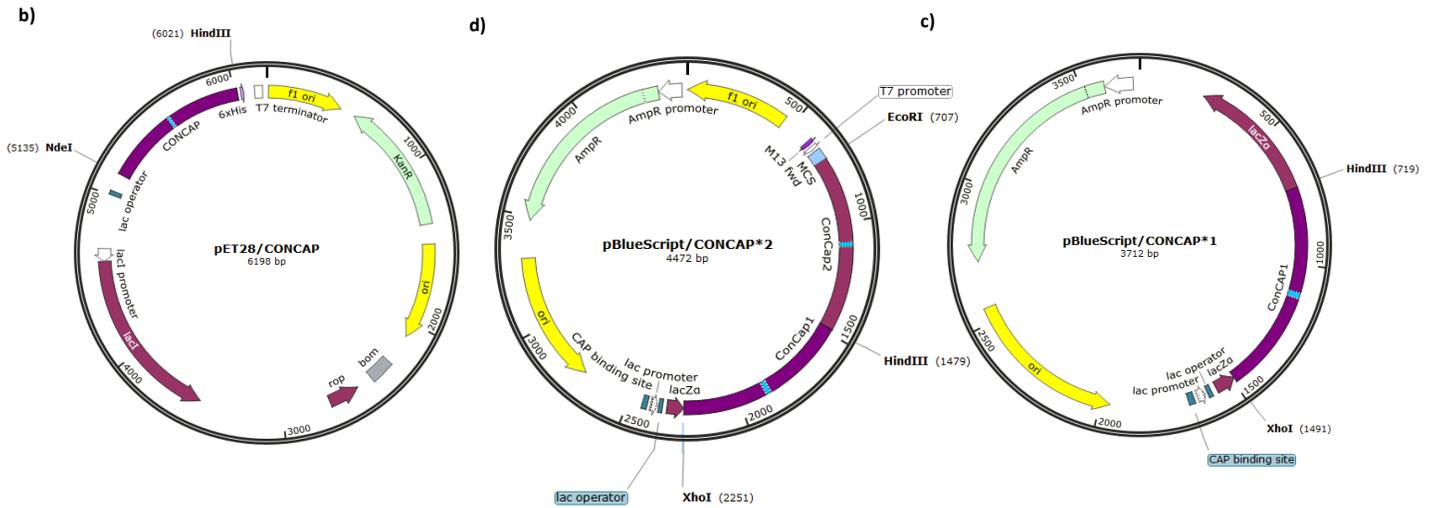
En Aggrescan, entre más negativo sea el valor arrojado el índice de agregación es menor

a) **gBlock G1**

5' CCATATGCATGTCCGGTTTTCAACAACCATGCAAATGCTGAATGAGGGCATCGTTCCTCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATT  
 CCGAGTCCGGGCTGCGCGTTGGTGGCGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCGTCAACCACCATCAAACAGGATTTTCGGCTGCT  
 GGGGCAAACAGCGTGGACCGCTTCTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACGTTGAAAAGAAAAACCCCTGGCGCCCAATACG  
 CAAACCGCCTCTCCCCGCGGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAGCGGGCAGTGAGCGCAACGCAATAAATGTGATCTAGATCACATTT  
 TAGGCACCCAGGCTTTAC3'

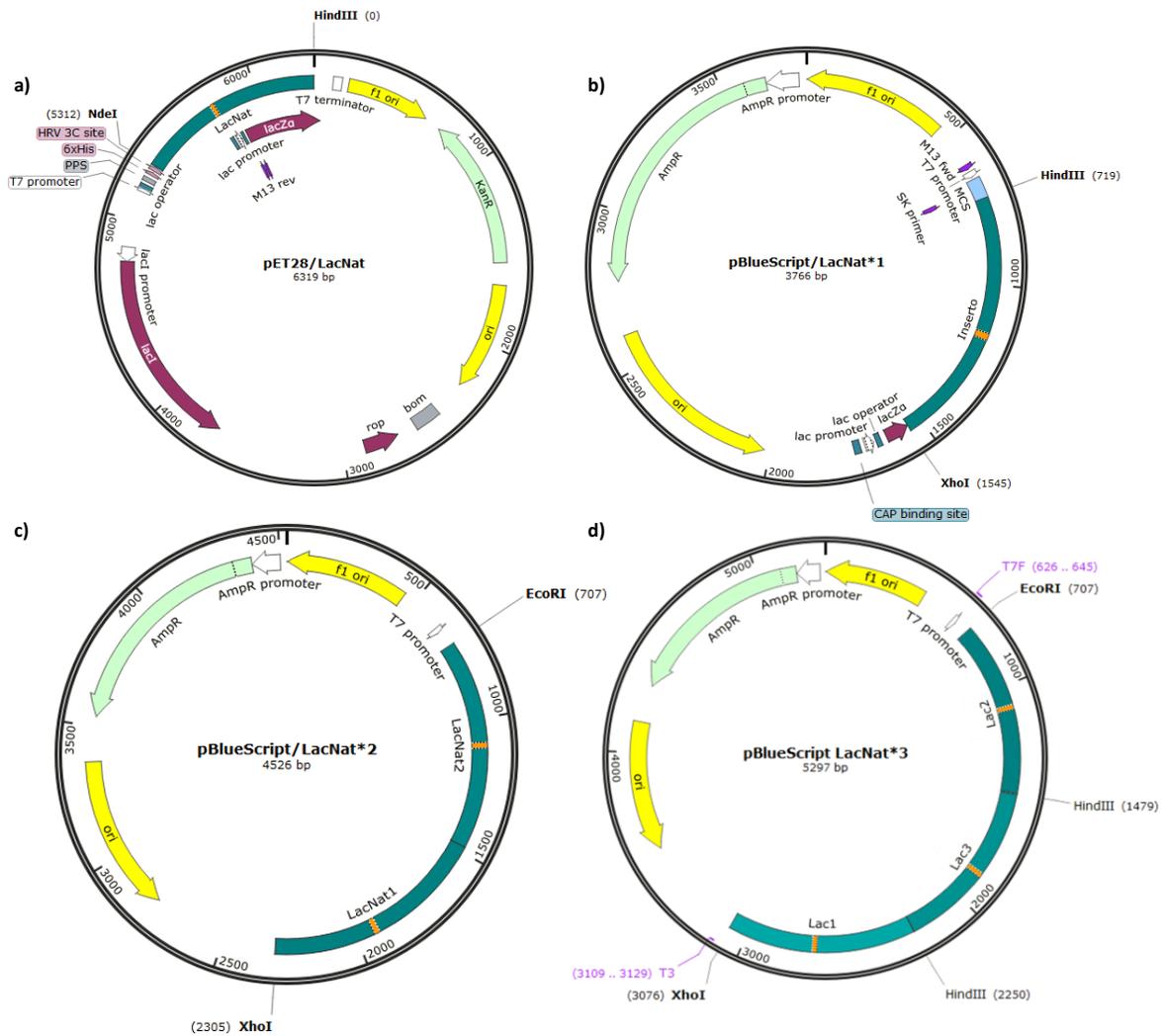
**gBlock G2**

5' CAGTGAGCGCAACGCAATAAATGTGATCTAGATCACATTTTAGGCACCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTC  
 ACACAGGAAACAGCTATGACCATGATTACGGATTACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCCTGGCGTTACCCAACCTAATCGCCTTGACGACATCCCCCTT  
 TCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTCGCTGGTTTCCGGCACCGAAGCGGTGCCGGA  
 AAGCTGGCTGGAGTGGCATCTTCTGAGGCCGATACTGTCTGCTGCCCTCAAACCTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTAACCTATCCCATTTACGGTC  
 AATCCGCGGTTAAGCTTGGCGCCGC3'



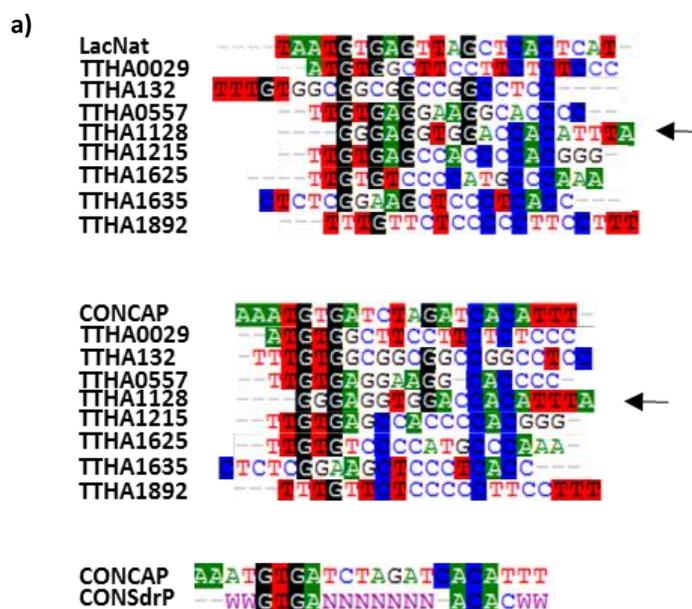
**Figura S1.** Vectores construidos con el inserto ConCap

a) Secuencia de los gBlocks G1 y G2 empleados para construir el inserto del vector pET28/CONCAP mediante PCR de fusión, cada gBlock comprende 455pb  
 b) Esquema del vector pET28/CONCAP, donde el inserto CONCAP fue clonado entre los sitios NdeI-HindIII del vector pET28.  
 c) Esquema del vector pBlueScript/CONCAP\*1, donde el inserto CONCAP1 fue subclonado entre los sitios XhoI-HindIII del vector pBlueScript.  
 d) Esquema del vector pBlueScript/CONCAP\*2 donde el inserto CONCAP2 fue subclonado entre los sitios HindIII-EcoRI del vector pBlueScript/CONCAP\*1.



**Figura S2.** Vectores construidos con el sitio LacNat

a) Esquema del vector pET28/LacNat, donde el inserto LacNat (1036 pb) se encuentra clonado entre los sitios NdeI-HindIII del vector pET28. b) Esquema del vector pBlueScript/LacNat\*1, donde el inserto LacNat1 fue subclonado entre los sitios XhoI-HindIII del vector pBlueScript. c) Esquema del vector pBlueScript/LacNat\*2, donde el inserto LacNat2 fue subclonado entre los sitios HindIII-EcoRI del vector pBlueScript/LacNat\*1. d) Esquema del vector pBlueScript/LacNat\*3, donde el inserto LacNat3 fue subclonado en el sitio HindIII del vector pBlueScript/LacNat\*2.



b)

Gen	Dominio conservado	e-value	Anotación de producto	Posible rol celular
TTHA0029	-	-	Proteína hipotética	Regulador transcripcional
TTHA0557	COD0605	8e-73	Superóxido dismutasa	Control redox
TTHA1128	COG1506	2e-21	Peptidasa	Control de calidad de proteínas, suplemento de nutrientes
TTHA1215	COG0492	2e-30	Tiorredoxina reductasa	Control redox
TTHA1625	TIGR03562	7e-54	Proteína inducible osmóticamente Osmc	Control redox
TTHA1635	TOGR00049	8e-41	Cluster hierro-azufre	Control redox
TTH1892	COG0556	1e-180	UvrB	Reparación del DNA
TTHB132	PRK00058	2e-72	Reductasa A metionin sulfoxido (MseA)	Reparación de proteínas

**Figura S3.** Análisis de los sitios de unión de SdrP

a) Alineamiento nucleotídico de los sitios de unión putativos de SdrP de la secuencia upstream de los genes TTHA0557, TTHA0029, TTHA1128, TTHA1215, TTHA1625, TTHA1635, TTHA1892 y TTHB132 contra las secuencias CONCAP y LacNat, así como el alineamiento entre CONCAP y la secuencia consenso de SdrP (CONSdrP). Las flechas marcan la secuencia que es más similar en ambos alineamientos, es decir, TTHA1128 b) Funciones de los genes putativos regulados por SdrP identificados por Agari y col.

## **ANEXO I**

### **Protocolo de PCR a partir de producto de ligación**

1. Digerir 360ng del inserto de interés y el vector a emplear con 10U de las enzimas apropiadas en un volumen final de 20  $\mu$ L.
2. Ligar 20ng de vector en una relación 1:6 M con el inserto empleando 2.5U de ligasa T4 en un volumen final de 10 $\mu$ L.
3. Incubar la ligación de acuerdo a las recomendaciones del proveedor.
4. Diluir 1 $\mu$ L de ligación en H<sub>2</sub>O MiliQ DEPC estéril en una relación 1:10 y tomar 1  $\mu$ L como templado para la PCR.
5. Preparar el mix de PCR con: 1X de amortiguador GM2, 0.5  $\mu$ M de cada oligo, 200 $\mu$ M de cada dNTP, 3mM de MgCl<sub>2</sub> y 0.5U de Taq polimerasa.
6. Correr la PCR por 30 ciclos con la T<sub>m</sub> óptima para los oligonucleótidos empleados.
  7. Verificar la amplificación de la banda en el tamaño esperado por electroforesis en gel de agarosa al 1%.