



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

**POSGRADO EN CIENCIAS EN BIOLOGIA MOLECULAR**

**Análisis de estabilidad estructural de la proteína  
Clp de *Xanthomonas axonopodis***

Tesis que presenta  
**José Iván Soria Rodríguez**

Para obtener el grado de  
**Maestro en Ciencias en Biología Molecular**

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

San Luis Potosí, S.L.P., junio de 2025



## Constancia de aprobación de la tesis

La tesis “**Análisis de estabilidad estructural de la proteína Clp de *Xanthomonas axonopodis***” presentada para obtener el Grado de Maestro(a) en Ciencias en Biología Molecular fue elaborada por **José Iván Soria Rodríguez** y aprobada el **23 junio de 2025** 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

---

**Dra. Daphne Edith González Juárez**  
Miembro del Comité Tutorial

---

**Dr. Francisco Elihú Bautista Redonda**  
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 1241158) 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) por los servicios otorgados y por el acceso a sus instalaciones

## **Dedicatorias**

Dedico el presente trabajo a mis padres, por su inquebrantable e incondicional amor a lo largo de mi vida y de esta experiencia. Por siempre mostrarme la importancia de perseverar en nuestras metas.

A mis compañeros de laboratorio, por cada momento de colaboración, apoyo, corrección y ánimo, por las pláticas amenas y enriquecedoras que hicieron del laboratorio un lugar cálido y acogedor.

A mi novia, Samantha, por siempre tener las palabras adecuadas y llenas de cariño para impulsarme y enfrentar cualquier cosa.

Al Dr. Samuel Lara, por su incansable paciencia y eterna disposición a ayudarme en todo momento, por su guía en este trabajo y todo el aprendizaje que he obtenido en este proyecto.

## **Agradecimientos**

Al Consejo de Ciencia y Tecnología (CONACYT).

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

Al Dr. Samuel Lara González por su guía y dirección de este trabajo.

A la Dra. Daphne Edith González Juárez y al Dr. Francisco Elihú Bautista

Redonda, por sus aportaciones como miembros del comité tutorial.

A la Tec. Rosa Espinosa Luna por todo el apoyo brindado dentro y fuera del laboratorio.

# Contenido

Constancia de aprobación de la tesis	ii
Créditos institucionales	¡Error! Marcador no definido.
Acta del examen	¡Error! Marcador no definido.
Dedicatorias	iv
Agradecimientos	v
Resumen	vii
Abstract	viii
1. Introducción	1
2. Materiales y métodos	9
2.1 Expresión heteróloga de Clp	9
2.2 Purificación de Clp mediante cromatografía por afinidad a níquel	9
2.3 Análisis de agregación por dispersión dinámica de luz (DLS)	10
2.4 Detección de Clp por Western Blot	11
2.5 Ensayos de cristalización	12
2.6 Análisis y comparación <i>in silico</i> de XaClp y EcCRP	12
3. Resultados y discusión	13
3.1 Expresión heteróloga de Clp	13
3.2 Purificación de Clp mediante cromatografía por afinidad a níquel	14
3.3 Análisis de agregación por dispersión dinámica de luz	15
3.4 Detección de Clp por Western Blot	21
3.5 Ensayos de cristalización	22
3.6 Análisis y comparación <i>in silico</i> de XaClp y EcCRP	22
Conclusión	31
Referencias	33
Figuras suplementarias, tablas y anexos	41

# Resumen

## **Análisis de estabilidad estructural de la proteína Clp de *Xanthomonas axonopodis***

El género *Xanthomonas* es un amplio grupo de bacterias fitopatógenas que representa una amenaza para múltiples cultivos de relevancia agrícola y económica, provocando la pérdida de grandes porcentajes de producción. Los factores de virulencia de estas bacterias como la producción de exopolisacáridos, proteasas, pectinasas y endoglucanasas, además de la transición desde un estado unicelular móvil a un estado multicelular unido a superficies y con la capacidad de formar biopelículas son regulados por el factor de transcripción Clp y su interacción con el segundo mensajero diguanilato monofosfato cíclico (c-di-GMP). Como el resto de las integrantes de la superfamilia de proteínas CRP/FNR a la que pertenece, Clp posee un dominio de unión a dinucleótido cíclico (CBD) y un dominio de unión de ADN (DBD). Clp tiene dos conformaciones: la activa que es capaz de unirse a su secuencia blanco en el ADN y la inactiva, que no une ADN y que se obtiene cuando se une a su regulador alostérico, el c-di-GMP. Por ende, el estudio detallado de la conformación activa de Clp y de su transición a la conformación inactiva representa un área de investigación relevante para la comprensión de los procesos regulados por Clp en las bacterias del género *Xanthomonas* y de otros grupos taxonómicos con proteínas homólogas. En un trabajo previo de nuestro grupo, se obtuvo la estructura cristalográfica de la conformación activa de Clp de *Xanthomonas axonopodis* (XaClp). Sin embargo, su alta propensión a agregarse ha impedido cristalizar la conformación inactiva. Por ello, en este trabajo evaluamos el papel de diferentes aditivos sobre la agregación de Clp mediante Dispersión Dinámica de Luz (DLS). Además, realizamos un análisis bioinformático para identificar los sitios de agregación presentes en Clp y diseñamos una construcción que potencialmente elimina este problema.

Palabras clave: Factor de transcripción, CRP, Clp, Agregación, DLS

## Abstract

### **Structural stability analysis of the Clp protein from *Xanthomonas axonopodis***

The genus *Xanthomonas* comprises a broad group of phytopathogenic bacteria that threaten multiple crops of agricultural and economic importance, leading to significant production losses. Virulence factors in these bacteria—such as the production of exopolysaccharides, proteases, pectinases, and endoglucanases, as well as the transition from a motile single-celled state to a surface-attached multicellular state capable of forming biofilms—are regulated by the transcription factor Clp and its interaction with the second messenger cyclic diguanylate monophosphate (c-di-GMP). Like other members of the CRP/FNR protein superfamily to which it belongs, Clp possesses a cyclic nucleotide-binding domain (CBD) and a DNA-binding domain (DBD). Clp exhibits two conformations: an active form that binds to its target DNA sequence and an inactive form, which does not bind DNA and is obtained when it binds to its allosteric regulator, c-di-GMP. Therefore, a detailed study of Clp's active conformation and its transition to the inactive state represents a key research area for understanding Clp-regulated processes in *Xanthomonas* and other taxonomic groups containing homologous proteins. In a previous study by our group, we determined the crystallographic structure of the active conformation of *Xanthomonas axonopodis* Clp (XaClp). However, its high aggregation propensity has prevented the crystallization of the inactive conformation. Thus, in this work, we assessed the effect of different additives on Clp aggregation using Dynamic Light Scattering (DLS). Additionally, we performed a bioinformatic analysis to identify aggregation-prone sites in Clp and designed a construct that may potentially resolve this issue.

Key words: Transcription factor, CRP, Clp, Aggregation, DLS

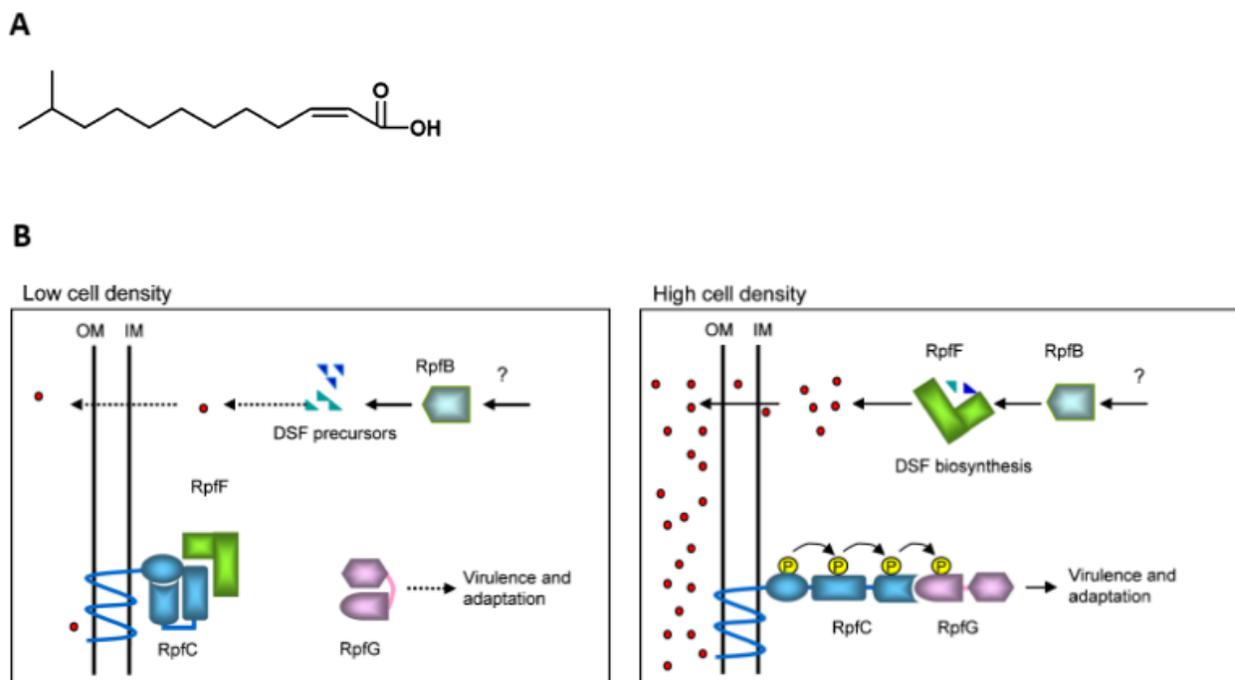
## 1. Introducción

La proteína Clp, cuya secuencia consta de 230 residuos de aminoácidos, es un factor de transcripción encontrado en bacterias fitopatógenas de alto impacto agrícola incluidas dentro del género *Xanthomonas* (Graham et. al, 2004; Hayward, 1993). Este grupo está compuesto por bacterias gram-negativas pertenecientes a la clase *Gammaproteobacteria* dentro del phylum *Proteobacteria*, con una morfología característica en forma de bastón, con un solo flagelo, y cuyas colonias adoptan un distintivo color amarillo debido a la producción del pigmento xanthomonadina (Bradbury, 1984; Starr & Stephens, 1964).

El proceso de patogénesis producido por las bacterias de este grupo ocurre tras su ingreso a la planta y genera lesiones características de acuerdo a la vía de entrada del patógeno: la infección a través de los hidatodos, en los márgenes de las hojas, genera lesiones en forma de V, mientras que la infección de los estomas produce lesiones redondeadas. Además, las lesiones previas en la planta también son sitios susceptibles al ingreso del patógeno. Una vez dentro de la planta, las bacterias pueden multiplicarse de manera local o colonizar el xilema para distribuirse de manera sistémica (Ryan et. al, 2011).

A nivel molecular, el proceso de patogénesis está regulado mediante un sistema de quorum sensing dependiente de la síntesis de un ácido graso *cis* insaturado denominado factor de señal difusible (DSF, por sus siglas en inglés). (Figura 1). Este sistema de regulación incluye a las enzimas RpfB (una ácido graso de cadena larga-CoA ligasa), RpfC (una cinasa sensora de DSF), RpfF (una enoil-CoA hidratasa) y RpfG (una fosfodiesterasa de c-di-GMP). El progreso de la infección depende de la concentración de DSF en el medio. En condiciones de baja población celular, la concentración de DSF se encuentra en un nivel basal pero insuficiente para activar el sistema de regulación compuesto por las enzimas Rpf, el cual requiere la activación de RpfC. Por otro lado, cuando la población celular aumenta, el nivel de DSF supera el umbral de activación de RpfC, cuya región extracelular de

su dominio transmembranal interactúa con DSF, produciendo su autofosforilación y un cambio conformacional consecuente que libera a RpfF. Una vez liberada, la síntesis de DSF aumenta significativamente. Este aumento en la síntesis de DSF produce la transferencia de un grupo fosforilo desde RpfC, mediante su dominio histidina fosfotransferasa en su región C-terminal, a RpfG. En su estado fosforilado, RpfG se encarga de la hidrólisis del enlace fosfodiéster de c-di-GMP. Como resultado final, la concentración intracelular de c-di-GMP es reducida, promoviendo una mayor actividad de Clp, con la consecuente activación de genes relacionados a patogénesis y virulencia, promoviendo la síntesis de exopolisacáridos, proteasas, pectinasas y endoglucanasas, así como la transición desde un estado unicelular móvil a un estado multicelular unido a superficies con la capacidad de formar biopelículas (An et. al, 2020; Ryan et. al, 2015; He et. al, 2007).



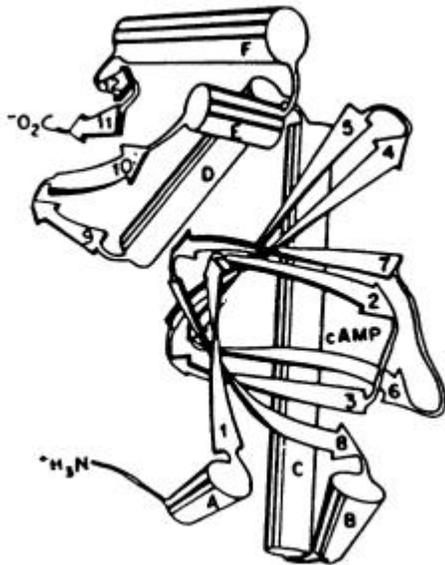
**Figura 1. Sistema de regulación por *Quorum sensing* de *Xanthomonas* mediante DSF (modificada de He et. al, 2008).**

El proceso de fitopatogénesis provocado por las bacterias del género *Xanthomonas* se encuentra regulado por un mecanismo de *Quorum sensing* dependiente de la concentración extracelular del factor de señal difusible —DSF por sus siglas en inglés— (A). A bajos niveles de densidad celular, la síntesis de DSF es reducida

debido a que la enzima RpfF, responsable del paso final de su síntesis, se encuentra inhibida por su unión a la proteína RpfC. En cambio, conforme la densidad celular aumenta y, con ello, el nivel de DSF extracelular es mayor, la región extracelular de la proteína RpfC cambia su conformación, autofosforilándose y liberando a la enzima RpfF, la cual adquiere una conformación catalíticamente activa que permite la síntesis de más DSF. Al mismo tiempo, la proteína RpfC transfiere un grupo fosforilo a la proteína RpfG, encargada de la regulación de procesos de virulencia y adaptación (B).

Clp pertenece a la superfamilia CRP/FNR, cuyos integrantes presentan una enorme variedad de funciones metabólicas que incluyen la regulación de factores de virulencia, fijación de nitrógeno, fotosíntesis, degradación de compuestos aromáticos, entre otras. Los elementos estructurales representativos de esta superfamilia son la presencia de un motivo hélice-alfa-hélice (HTH) en su región C-terminal (DBD), que se une al surco mayor del ADN y un dominio de unión a nucleótido cíclico en su región N-Terminal (CBD) (Körner et. al).

El primer miembro descrito de esta superfamilia es la proteína CRP de *E.coli*, un regulador transcripcional de ~50 kDa constituido por un homodímero de 209 residuos por subunidad (Lawson et. al, 2004). El sitio consenso de unión a ADN de *EcCRP* está constituido por la secuencia palindrómica 5'-AAATGTGATCTAGATCACATTT-3'. Cada subunidad del homodímero de CRP está conformada a su vez por 6 hélices alfa y 11 hebras beta (Figura 2) (Kim et. al, 1992).

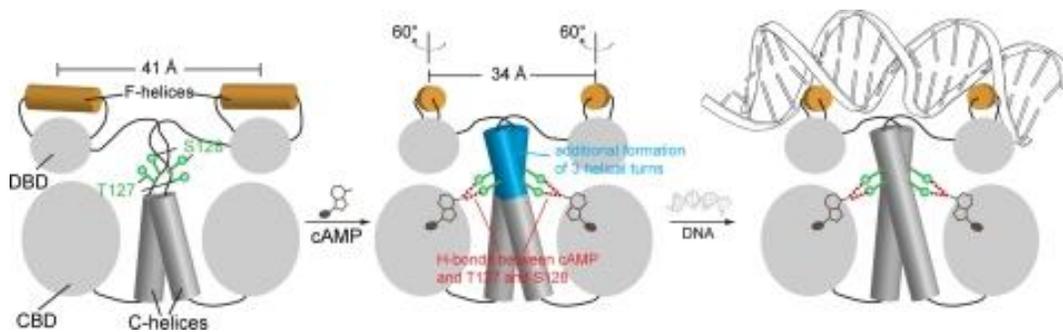


**Figura 2. Esquema estructural de un monómero de *EcCRP* (tomada de Kim et. al, 1992).**

Las seis hélices alfa presentes en la estructura de *EcCRP* se encuentran nombradas de la A a la F, mientras que sus 11 hebras beta se muestran numeradas. Entre las hélices C y D se encuentra una región de bisagra compuesta por los residuos 135-138, mientras que las hélices E y F constituyen el motivo hélice-giro-hélice presente en el dominio DBD, siendo la hélice F la que media el contacto entre la proteína y el ADN.

CRP posee un mecanismo de regulación alostérico, teniendo una conformación inactiva cuyo dominio DBD, conformado por los residuos 1-136, se une al ADN de manera inespecífica y con poca afinidad. En esta conformación, la hélice C, que sirve como interfaz de dimerización para los dominios CBD de cada subunidad, se encuentra como una hélice corta con tres vueltas menos. Esta disposición estructural conlleva a un posicionamiento de las hélices-F, localizadas en el motivo HTH, poco propicio para el reconocimiento correcto del ADN. En cambio, cuando el dominio CBD, conformado por los residuos 139-209, interactúa con el adenosín monofosfato-3',5' Cíclico (AMPc), ocurre un cambio conformacional que permite que la proteína adopte una conformación activa con una alta afinidad por su sitio de

unión en el ADN. En este cambio, la hélice-C experimenta una transición a una hélice extendida con tres vueltas adicionales, que conlleva a una rotación de  $60^\circ$  y un acortamiento en la distancia entre las hélices-F, lo que permite su posicionamiento óptimo para unirse al ADN (Figura 3) (Benoff et. al, 2002; Popovych et. al, 2009).



**Figura 3. Cambios conformacionales en *EcCRP* tras su unión a AMPc (tomada de Popovych et. al, 2009).**

En ausencia de AMPc, la proteína *EcCRP* muestra una configuración en la que las hélices F de la región DBD se encuentran a una distancia de 41 Å que no permite la unión al surco mayor del ADN, necesaria para su actividad de regulador transcripcional. En cambio, tras la unión de AMPc dentro del sitio de unión a AMPc, las hélices-C contiguas a éste experimentan una transición a una configuración extendida con tres vueltas adicionales, lo cual genera una rotación horizontal de  $60^\circ$  de las hélices F y el consecuente acortamiento de 7 Å en la distancia entre ellas, situándolas a una distancia final de 34 Å, propicia para su unión sobre surcos mayores sucesivos del ADN.

Si bien la mayoría de integrantes de la superfamilia CRP/FNR requieren la unión a nucleótido cíclico específico para permitir su unión al ADN, Clp no presenta este requerimiento. En cambio, la unión a su efector específico, el segundo mensajero c-di-GMP, inhibe de manera alostérica su unión al ADN (Leduc & Roberts, 2009). Mediante análisis bioinformáticos se ha encontrado que Clp tiene una identidad de

secuencia del 45% con la proteína CRP (de Crecy-Lagard, 1990). Una característica interesante dentro de esta identidad compartida es la conservación de 4 de los 6 residuos de aminoácidos (G87, E88, R98 y T148) que unen el AMPc en CRP (Dong et. al, 1992). Sin embargo, tanto la enzima adenilato ciclasa como el AMPc se encuentran ausentes en los miembros del género *Xanthomonas* (Chen et. al, 2010). En cambio, Clp ha adoptado un rol protagónico de la regulación génica dentro de este grupo, gracias a la adquisición de mutaciones particulares que le han conferido la versatilidad de modular la actividad de por lo menos 299 genes, entre los que destacan algunos involucrados en la modulación del comportamiento entre estados planctónico-sésil, la formación de biopelículas, y la producción de factores de virulencia (He et. al, 2007; Chen et. al, 2010). Estudios de cepas de diferentes especies del género *Xanthomonas* mutantes del gen *clp* o de genes involucrados en la regulación de éste, han permitido corroborar lo anterior (Egli et. al, 1990; Slater et. al, 2000; Tao et. al, 2010; Romeo et. al, 2012; Xiuhong et. al, 2012; Römling et. al, 2013; Jenal et. al, 2017).

La región N-terminal de Clp está conformada por los residuos 1-127, mientras que los residuos 158-230 conforman su región C-terminal, con una larga hélice alfa (hélice-C) intermedia compuesta por los residuos 128-157 que sirve como interfaz de dimerización. La región N-terminal está compuesta por un barril beta formada a su vez por 8 hebras beta antiparalelas, flanqueado por 2 hélices alfa (hélice-A y hélice-B), mientras que la región C-terminal se compone de 4 hebras beta antiparalelas y 3 hélices alfa (hélice-D, hélice-E y hélice-F), de las cuales la hélice-E y hélice-F, así como el bucle que conecta a ambas, constituyen el dominio HTH con capacidad de unión al ADN. Estudios cristalográficos muestran que el homodímero de Clp, con un peso molecular de ~52 kDa, posee una estructura simétrica, en la que cada subunidad presenta el acercamiento entre sus hélices-F propicio para su unión activa al ADN. Este acercamiento está dado por la interacción entre el residuo D162 y el residuo D159 mediante dos puentes de hidrógeno. Así mismo, el residuo V165 posee una cadena lateral con un mayor volumen respecto

al residuo A145 correspondiente en CRP. Este par de características estructurales propician la conformación activa de Clp con unión al ADN (Chin et. al, 2010).

En esta conformación activa, destaca un par de elementos particulares adicionales. En primer lugar, al comparar los sitios de unión de CRP y Clp a su respectivo efector, la mayoría de residuos presentan una disposición similar a excepción de los residuos S84 y E130 de CRP, correspondientes a los residuos E99 y R150 en Clp. Como consecuencia de esta última interacción se forma un puente salino en el que la cadena lateral del residuo E99 se dispone hacia fuera del sitio de unión a ligando, manteniéndolo cerrado e impidiendo su interacción con AMPc. (Chin et. al, 2010). Estos dos rasgos explican la baja afinidad de Clp por AMPc observada en investigaciones previas (Leduc & Roberts, 2009). Sin embargo, esta misma interacción está acompañada por una molécula de agua que forma dos puentes de hidrógeno entre un átomo de nitrógeno de la cadena lateral del residuo E99 y un átomo de oxígeno en la cadena lateral del residuo R150 que en última instancia propicia un efecto en Clp similar al efecto correspondiente que la unión de AMPc ejerce al unirse a CRP (Chin et. al, 2010). El segundo elemento clave consiste en el residuo R195 de Clp (correspondiente al residuo Q175 de CRP), el cual propicia un cambio en la disposición del residuo E192 que da lugar a la formación de un puente salino entre ambos residuos y, con ello, la formación de un puente de hidrógeno entre los residuos E192 y Y79. Dado que estos residuos se encuentran localizados en la región C-terminal y N-terminal respectivamente, su interacción ocasiona un acercamiento entre ambas regiones, reduciendo la distancia entre ellas en alrededor de 6 Å. En esta disposición, el dímero de Clp adopta una conformación simétrica en la que cada monómero presenta la forma cerrada adecuada para su unión al ADN (Chin et. al, 2010).

Por otro lado, uno de los obstáculos que impide un rendimiento óptimo en la expresión y purificación de una proteína en particular consiste en la formación de agregados que además complican su estudio estructural y funcional (de Marco, 2008). El fenómeno de agregación es común en la producción de proteínas recombinantes, en la que existe una amplia documentación sobre su acontecimiento

en múltiples sistemas de expresión (Gasser et. al, 2008). En trabajos previos de nuestro laboratorio se han estandarizado algunas metodologías en la expresión y purificación de la proteína Clp de *Xanthomonas campestris* y *Xanthomonas axonopodis* (XaClp), en las que se han hecho evidentes sus particularidades de estabilidad y rendimiento. En el caso de XaClp, se han obtenido cristales con una resolución de 2.6 Å gracias al trabajo de Karina Rosales (2019). Sin embargo, en el presente trabajo se observó que la proteína tiene una alta tendencia a agregarse, por lo cual se realizó un análisis bioinformático y de las condiciones que podrían estar implicadas en ello, con la finalidad de proponer una metodología que permita un estudio estructural y funcional óptimo para lograr profundizar en su caracterización. De esta manera, se lograron determinar algunos aditivos que muestran un efecto de reducción en la agregación de XaClp mediante su análisis por dispersión dinámica de luz. Además, tras el análisis *in silico* de las regiones de la proteína con mayor contribución a su agregación (denominadas como hotspots) se proponen mutaciones que permiten evitar su acontecimiento al identificar al hotspot de mayor relevancia en este fenómeno, así como una mutación puntual que potencialmente permite eliminar su formación.

## **2. Materiales y métodos**

### **2.1 Expresión heteróloga de Clp**

Se descongelaron 50  $\mu$ L de células BL21 (DE3 Star) competentes durante 10 min, y se añadieron 2  $\mu$ L del vector con la construcción pET28pps\_6XHis-Clp. Tras 20 min de incubación en hielo, las células fueron sometidas a choque térmico a 42°C durante 45 s, para posteriormente mantenerlas en incubación en hielo durante 2 min. Se añadieron 500  $\mu$ L de medio LB, dejando las células en incubación a 37°C con agitación constante a 190 rpm durante 1 hora. Posteriormente, las células fueron centrifugadas a 6,000 rpm, y se tomaron 500  $\mu$ L del sobrenadante para su desecho. El precipitado celular se resuspendió en 50  $\mu$ L de medio LB y sembrado en placas de LB agar adicionado con kanamicina [50  $\mu$ g/mL], dejando en incubación a 37°C durante 16 h. Tras el tiempo de incubación, se prepara un preinóculo en 25 mL de medio LB adicionado con kanamicina [50  $\mu$ g/mL] utilizando una de las colonias obtenidas, manteniendo en incubación a 37°C durante 12 h. Una vez transcurrido este tiempo, se añade el preinóculo a 500 mL de medio LB adicionado con kanamicina [50  $\mu$ g/mL], el cual es incubado a 37°C en agitación constante a 190 rpm hasta alcanzar una OD 600 de 0,6. La inducción fue realizada mediante la adición de IPTG a una concentración final de 0.5 mM al cultivo celular, para posteriormente mantenerlo en incubación a 28°C en agitación constante por 16 h. Se tomaron muestras de 1 mL del cultivo tanto antes de la adición de IPTG como después del tiempo de expresión para realizar la lectura de densidad óptica a 600 nm ( $OD_{600nm}$ ) de cada una, para su análisis por SDS-PAGE (Figura 4). Finalmente, el cultivo celular se centrifuga a 10,000 rpm en rondas de 10 min, desechando el sobrenadante y conservando la masa celular en refrigeración a -20°C hasta su uso posterior.

### **2.2 Purificación de Clp mediante cromatografía por afinidad a níquel**

La masa celular de 1 L de cultivo (2.42 gramos de peso húmedo) se resuspende en 35 mL de amortiguador de lisis (Tris-HCl pH 7.5 25 mM, glicerol 15%, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 1 mM) y se lisa por sonicación por pulsos 15s ON/30s OFF a 50% de amplitud durante 7:30 min. El lisado celular se centrifuga a 12,000

rpm durante 15 min. Posteriormente, se añade polietilenimina (PEI) a una concentración final de 0.5% para precipitar los ácidos nucleicos presentes en el lisado celular y se centrifuga nuevamente a 12,000 rpm durante 15 min. La fracción soluble obtenida es transferida a una columna de agarosa Ni-NTA de Qiagen previamente equilibrada con 5 volúmenes de columna (CV) de amortiguador de lisis. La elusión de la proteína se realiza mediante una serie de lavados en un gradiente de imidazol de la siguiente manera: 10 CV de amortiguador de lisis adicionado con imidazol 10 mM, 6 CV con amortiguador de lisis adicionado con imidazol 50 mM, 6 CV con amortiguador de lisis adicionado con imidazol 100 mM y 6 CV con amortiguador de lisis adicionado con imidazol 200 mM.

Se tomaron muestras del lisado celular, la fracción soluble, el frente y las eluciones para su análisis por SDS-PAGE. De igual manera, tras el análisis por SDS-PAGE se tomaron las muestras que contenían la proteína purificada para su cuantificación por Bradford.

### **2.3 Análisis de agregación por dispersión dinámica de luz (DLS)**

Se tomaron muestras de 200  $\mu$ L de la primera fracción eluida con imidazol 200 mM a las que se añadieron individualmente un aditivo distinto. Por otra parte, se tomó una muestra de 200  $\mu$ L de la misma fracción a la que no se añadió ningún aditivo. Los aditivos utilizados y sus concentraciones finales fueron: una mezcla equimolar de arginina y glutamato 50 mM, CHAPS 0.1%, Tween-20 0.05%, EDTA 1 mM y  $\beta$ -mercaptoetanol 15 mM.

A continuación, las muestras fueron analizadas por DLS para evaluar los distintos aditivos. Las muestras tomadas previamente fueron centrifugadas a 10,000 rpm durante 4.5 h previo a realizar sus análisis en el equipo ZetaSizer APS (Lorber et al, 2012). Se ejecutaron una serie de lavados automatizados del equipo hasta mantener la intensidad de kilocuentas por s (kcps) en un rango menor a 120 kcps. Se tomaron 100  $\mu$ L de cada muestra y se realizaron lecturas de cada una, en intervalos de 10 min durante 4 h a 25°C. De igual manera, se realizaron lecturas con las mismas condiciones 72 h después para evaluar el efecto de cada aditivo sobre la agregación de la proteína de interés.

## 2.4 Detección de Clp por Western Blot

Se corrió un gel espejo por SDS-PAGE tomando 10  $\mu$ L de las muestras antes ( $t_0$ ) y después de inducir la expresión de Clp ( $t_1$ ), de la muestra del sonicado  $t_1$  y una muestra de la primera fracción eluida con imidazol 200 mM. Una mitad del gel fue teñida con Coomassie R-250 y la otra mitad fue utilizada para evaluar la detección de la proteína de interés mediante Western Blot. De acuerdo con el protocolo (Bocanegra-Jimenez et al, 2021), el gel y la membrana de nitrocelulosa fueron sumergidos en amortiguador de transferencia (Trisma base 25 mM, glicina 192 mM, isopropanol 7.5% (v/v)) por 15 min en agitación suave en recipientes separados. Se utilizó la cámara Semidry Thermo Owl para realizar la transferencia a 15 voltios durante 90 min. Una vez transcurrido este período, la membrana fue incubada en solución de bloqueo (leche descremada tipo Svelty® al 5% disuelta en TBS) por 2 h a temperatura ambiente y en agitación suave, colocando hacia arriba la parte de la membrana que estuvo en contacto con el gel. Se realizaron tres lavados con 30 mL de TBS (Trisma base 20 mM, NaCl 0.5 M pH 7.5) por 5 min cada uno en agitación. Posteriormente, la membrana fue incubada con el anticuerpo primario 6X-His Epitope Tag de ThermoFisher #MA1-21315 en TTBS (Tris base 20 mM, NaCl 0.5 M pH 7.5, Tween-20 0.5% (v/v)) por 12 h a 4°C en agitación y 1 hora a temperatura ambiente. Después de ello, se realizó una serie de 6 lavados con 30 mL de TBS por 5 min cada uno. La incubación con el anticuerpo secundario se realizó utilizando el anticuerpo Goat Anti-mouse IgG Alkaline Phosphatase de Abcam #Ab97020 durante 2 h a temperatura ambiente en agitación. A continuación, se realizó una serie de 6 lavados con 30 mL de TTBS durante 5 min cada uno. Se incubó la membrana en 25 mL de amortiguador de fosfatasa alcalina (Trisma base 0.1 M pH 9.5, MgCl<sub>2</sub> 5 mM, NaCl 0.1 M) a temperatura ambiente y en agitación suave por 5 min.

Para el proceso de revelado, se incubaron 20 mL de amortiguador de fosfatasa alcalina a 37°C por 10 min y se adicionaron 200  $\mu$ L de Reactivo A y 200  $\mu$ L de Reactivo B del kit Alkaline Phosphatase Conjugate Substrate de BioRad #170-6432. La membrana fue sumergida en esta solución y se mantuvo en agitación a temperatura ambiente hasta que las bandas fueron visibles.

## **2.5 Ensayos de cristalización**

Posteriormente a la purificación por IMAC, las muestras con la proteína purificada fueron transferidas a un concentrador VivaSpin® Turbo 15. Para su concentración, se realizó la centrifugación de la muestra a 3,500 g y 20°C. El volumen de la muestra fue concentrado a 1 mL, y se diluyó nuevamente con 9 mL de amortiguador optimizado para eliminar el contenido de imidazol, repitiendo este proceso 3 veces más. La muestra concentrada fue transferida a un concentrador Vivaspin® 500 y se realizó su concentración a 10,000 g y 20°C hasta alcanzar un volumen final de 150 µL, con una concentración de proteína de 6.72 mg/mL.

Para el ensayo de cristalización, se realizó mediante la técnica de difusión de vapor en gota sentada, utilizando las 96 condiciones del kit de cristalización JCSG Plus™ (Anexo 1). Se depositaron 50 µL de precipitante en el pocillo mayor y gotas de la muestra de proteína y precipitante en una proporción 2:1 (1.2 µL de proteína:0.6 µL de precipitante) en el pocillo menor. Así mismo, como control negativo se depositaron gotas del amortiguador optimizado y precipitante en la misma proporción sobre el pocillo menor aledaño al pocillo sobre el que se depositaron la muestra de proteína y precipitante de cada condición. Tras sellar la placa de cristalización y mantenerla en resguardo en refrigeración a 14°C, se realizaron observaciones periódicas bajo el microscopio estereoscópico para dar seguimiento a la formación de cristales.

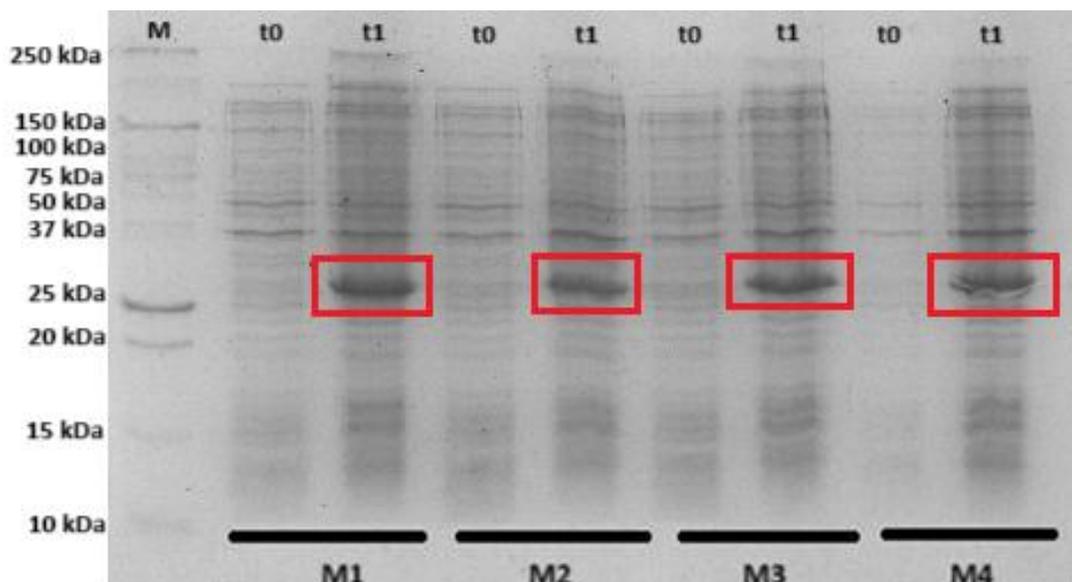
## **2.6 Análisis y comparación *in silico* de XaClp y EcCRP**

Se consultó la base de datos UniProt para descargar las secuencias en formato FASTA de la proteína de interés (código de entrada: Q8PQ45) así como de la proteína CRP de *E. coli* (código de entrada: P0ACJ8), las cuales fueron depositadas en el servidor AGGRESCAN (<http://bioinf.uab.es/aggrescan/>) para determinar los hotspots de agregación de cada una, con la intención de realizar un análisis comparativo entre éstos. Adicionalmente, se realizó la comparación estructural y de la disposición espacial de estos hotspots mediante el software UCSF Chimera (versión 1.17.3) utilizando los archivos pdb correspondientes a EcCRP (PDB ID: 1ILB) y a la estructura de XaClp resuelta por Karina Elisa Rosales Pérez durante su tesis de maestría (aún no depositada en el PDB).

### 3. Resultados y discusión

#### 3.1 Expresión heteróloga de Clp

La inducción por IPTG 0.5 mM, por 16 h a 28°C de cultivos celulares de células BL21 (DE3) Star transformadas con el vector que contiene la construcción pET28pps\_6XHis-Clp dio lugar a la expresión exitosa de nuestra proteína de interés, tal como puede observarse en el análisis de las muestras tomadas previa y posteriormente a la adición de IPTG mediante SDS-PAGE en geles al 15% de poliacrilamida (Figura 4).



**Figura 4. Expresión heteróloga de Clp**

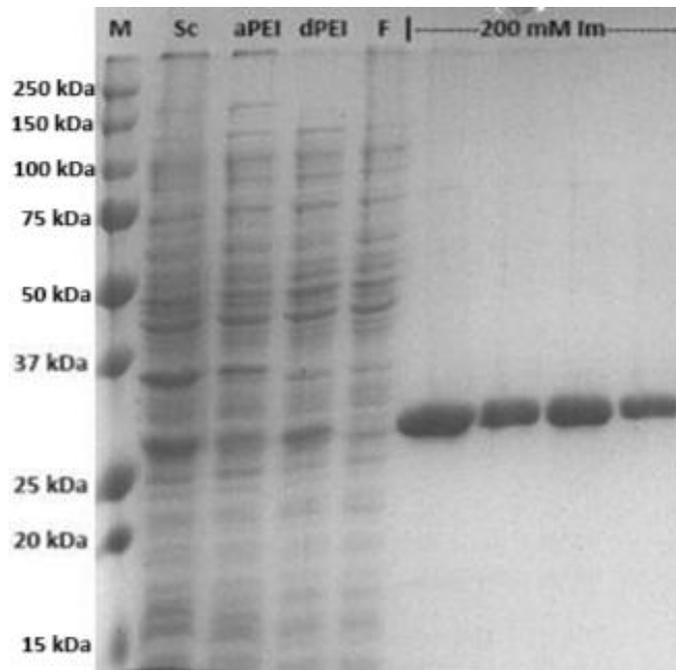
Gel de poliacrilamida al 15% en el que se muestra el electroferograma de un volumen de 10  $\mu$ L de 4 cultivos celulares (M1, M2, M3 y M4), donde se puede observar el perfil de expresión previo (t0) y posterior (t1) a la inducción por IPTG 0.5 mM. En este último, puede apreciarse una banda diferencial correspondiente al peso molecular de la proteína de interés (enmarcada en rojo) tomando como referencia el marcador de peso molecular (M) Precision Plus Protein™ Dual Color Standards de Bio-Rad.

### 3.2 Purificación de Clp mediante cromatografía por afinidad a níquel

Inicialmente, la purificación de la proteína de interés mediante IMAC se llevó a cabo con el amortiguador A (Tris-HCl 25 mM, pH 8.2, glicerol 15%, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 1 mM), obteniendo un rendimiento de 6.0 mg/L de acuerdo al método de Bradford. Además, se observó que las muestras colectadas mostraron signos notorios de precipitación y, al analizarse por DLS, una alta tendencia a agregarse. Sin embargo, al apreciar que dicho amortiguador tiene un pH cercano al punto isoeléctrico teórico de Clp (8.9, de acuerdo con el servidor [https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)), y lo cual promueve la tendencia de agregación, se optó por emplear un amortiguador optimizado con aquellos aditivos que mostraron mejores propiedades para reducir este inconveniente tras su evaluación por DLS, ajustando su pH final a un valor de 7.5.

Tras el análisis por SDS-PAGE de una muestra purificada con este amortiguador optimizado (Tris-HCl 25 mM, glicerol 15 %, KCL 500 mM. LiCl 25 mM,  $\beta$ -mercaptoetanol 15 mM, CHAPS 0.1%, EDTA 1 mM, pH 7.5) se encontró que la proteína pura eluye principalmente en las fracciones lavadas con imidazol 200 mM (Figura 5). De acuerdo con la cuantificación de estas fracciones por el método de Bradford se obtuvieron 14.6 mg/L de cultivo.

Por su parte, la inducción de la expresión heteróloga en dichos trabajos iniciales se realizó con una concentración final de IPTG de 0.2 mM, mientras que para los experimentos posteriores se realizó con una concentración final de 0.5 mM, de acuerdo a un trabajo previo de Chin y colaboradores en el que se reportó una estructura cristalográfica de la proteína Clp de *Xanthomonas campestris* (Chin et. al 2010).



**Figura 5. Purificación de Clp mediante cromatografía por afinidad a níquel**

Gel de poliacrilamida al 15% en el que se muestra el electroferograma de la muestra tras su sonicación (Sc), previo y posterior a la adición de PEI (aPEI, dPEI, respectivamente), el frente de la fracción soluble después de ser transferida a la columna Ni-NTA de Qiagen (F) y las fracciones eluidas con el amortiguador de lisis adicionado con imidazol 200 mM (200 mM Im), donde puede apreciarse en estas últimas la presencia de una banda con un peso molecular 26 kDa, correspondiente a la proteína de interés tomando como referencia el marcador de peso molecular (M) Precision Plus Protein™ Dual Color Standards de Bio-Rad.

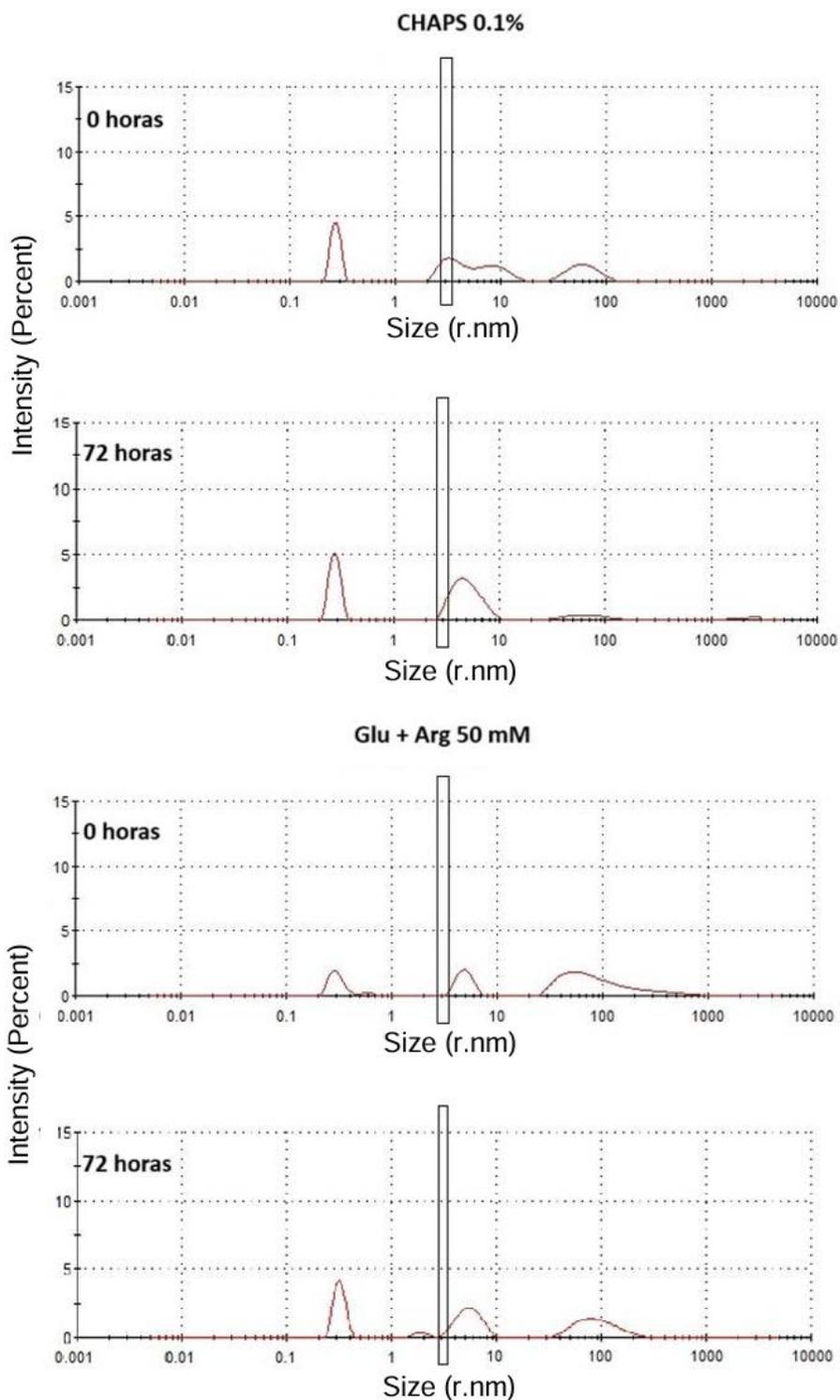
### 3.3 Análisis de agregación por dispersión dinámica de luz

Debido a la recurrente agregación de la proteína de interés, se decidió analizar el efecto de diferentes aditivos empleados con frecuencia para evitar este inconveniente (Arakawa & Timasheff, 1985; Golovanov et. al, 2004; Bondos & Bicknell, 2003).

Teniendo en cuenta que durante la purificación de proteínas mediante IMAC suele ocurrir la difusión de iones de níquel desde la columna cromatográfica hacia la muestra eluida, los cuales pueden promover la agregación de proteínas purificadas por esta técnica (Swaim et. al, 2020), el EDTA fue seleccionado como un aditivo a

evaluar por sus propiedades quelantes de diversos metales, entre ellos el níquel (Sharpe, 2014). De manera similar, el uso de detergentes no desnaturizantes como CHAPS y Tween-20 está documentado como una medida para prevenir la agregación proteica, debido a su capacidad de solubilizar agregados que se encuentran asociados por sus regiones hidrofóbicas expuestas al solvente (Qin et. al, 2010; Zhang et. al, 2021). Por otra parte, se ha encontrado que la adición de una mezcla equimolar de arginina y glutamato 50 mM es efectiva en reducir la tasa de agregación al interactuar con regiones cargadas e hidrofóbicas en la superficie de las proteínas, impidiendo así que interactúen entre ellas (Golovanov et. al, 2004). Por último, el  $\beta$ -mercaptoetanol es un agente reductor que forma aductos con cisteínas expuestas en la superficie de algunas proteínas, las cuales tienen la tendencia a formar residuos de cistina que promueven la agregación proteica (Sharpe, 2014). De esta manera, los aditivos que se tomaron en consideración fueron EDTA 1 mM, CHAPS 0.1%, Tween-20 0.05%, una mezcla equimolar de Arginina y Glutamato 50 mM, así como  $\beta$ -mercaptoetanol 15 mM.

Se realizaron lecturas durante 4 h a 25 °C en intervalos de 10 min por DLS en el equipo ZetaSizer APS de las diferentes muestras a las que se les añadió un aditivo particular a evaluar. Estas lecturas se realizaron tanto al momento de ser colectadas tras la purificación por IMAC como 72 h después. Las muestras colectadas tras la purificación por IMAC muestran la presencia de distintas poblaciones de partículas. En el caso de la muestra procesada con el amortiguador A (Tris-HCl 25 mM, pH 8.2, glicerol 15%, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 1 mM), destaca la presencia de partículas con un radio hidrodinámico en el rango de 100 nm, además de la presencia de una población con un radio hidrodinámico de alrededor de 500 nm tras 72 h. Si bien el resto de muestras a las que se le añadió un aditivo particular mostraron la presencia de una población de partículas con un radio hidrodinámico de 100 nm, pudo observarse que en algunas de ellas existe una señal menos intensa tras 72 h. En cambio, otras condiciones, como la adición de la mezcla equimolar de arginina y glutamato 50 mM, no consiguieron un efecto significativo en este aspecto (Figura 6).

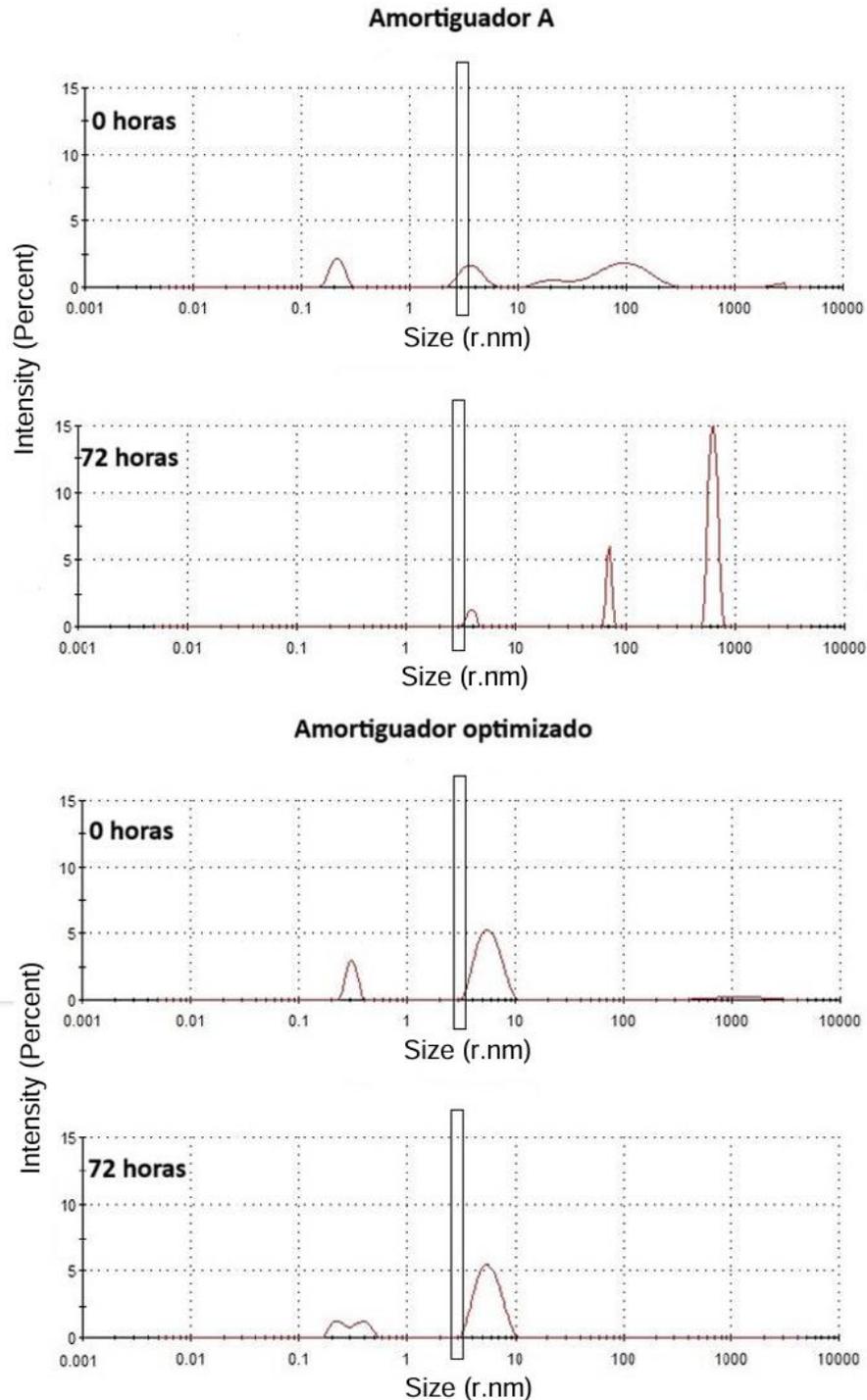


**Figura 6. Análisis de aditivos en el proceso de agregación de Clp.**

El uso de CHAPS 0.1% como aditivo en el amortiguador A (Tris-HCl 25 mM, pH 8.2, glicerol 15%, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 1 mM) no solo produce una estabilización en la señal de la población de partículas con un radio

hidrodinámico de 5 nm, sino que también puede observarse una reducción en la señal de la población de agregados con un radio hidrodinámico correspondiente a 100 nm tras 72 h. En cambio, si bien la mezcla equimolar de glutamato y arginina 50 mM muestra una señal positiva para una población con radio hidrodinámico de 5 nm, no logra una reducción significativa en la señal asociada a la población de agregados con radio hidrodinámico de 100 nm.

Al considerar el resultado positivo en la reducción de la señal de las poblaciones con un radio hidrodinámico de 100 nm en las muestras con las condiciones  $\beta$ -mercaptoetanol 15 mM, EDTA 1 mM y pH 7.5 (Figura S1) y CHAPS 0.1%, se evaluó su efecto en conjunto y, de manera concordante, este amortiguador optimizado (Tris-HCl 25 mM, glicerol 15 %, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 15 mM, CHAPS 0.1%, EDTA 1 mM, pH 7.5) mostró un efecto de mayor estabilidad para la proteína comparado al amortiguador sin modificar y cada condición por separado, en el cual no se observa la formación de poblaciones de grandes agregados, además de que puede apreciarse una señal más intensa para la población de partículas con un radio hidrodinámico de alrededor de 5 nm (Figura 7).



**Figura 7. El procesamiento de XaClp con el amortiguador optimizado reduce su tendencia a la agregación.**

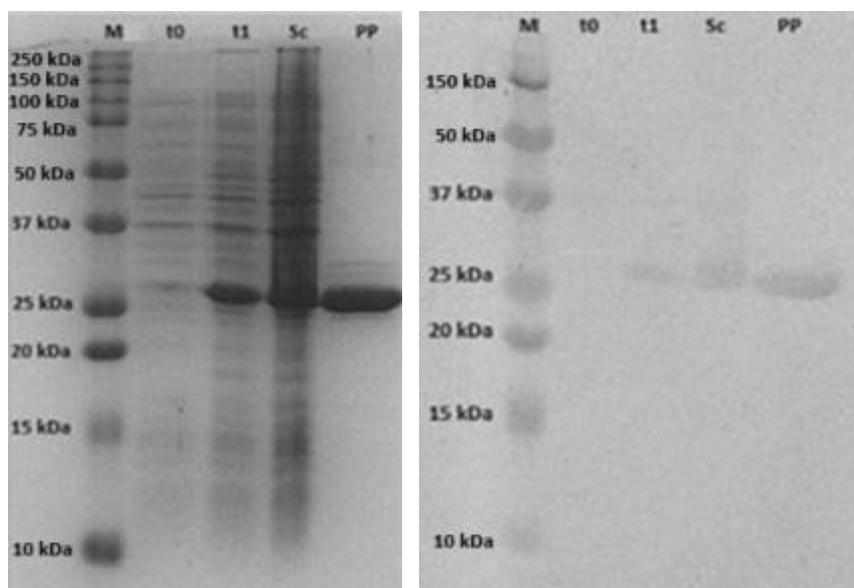
Inicialmente, las muestras procesadas con el amortiguador A (Tris-HCl 25 mM, pH 8.2, glicerol 15%, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 1 mM) muestran la presencia de una población de partículas con un radio hidrodinámico de 100 nm,

mientras que, tras 72 h, puede observarse la formación de agregados con un radio hidrodinámico de 1000 nm. En contraste, para las muestras procesadas con el amortiguador optimizado (Tris-HCl 25 mM, glicerol 15 %, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 15 mM, CHAPS 0.1%, EDTA 1 mM, pH 7.5) no se observa la formación de dichos agregados.

De acuerdo al peso molecular de nuestra proteína de interés (26 kDa) el radio hidrodinámico del homodímero de Clp, ronda los 3 nm. Por su parte, las poblaciones con un radio hidrodinámico de alrededor de 5 nm observadas en nuestras muestras corresponden a partículas con un peso molecular de 145 kDa. Teniendo en cuenta la capa de solvatación alrededor de la proteína, así como la presencia de la etiqueta 6XHis en ella, dichas poblaciones sugieren estar conformadas por tetrámeros de XaClp. De esta manera, podemos concluir que el uso del amortiguador optimizado evita la formación de los grandes agregados de nuestra proteína observados inicialmente con el amortiguador A, al mismo tiempo que mantiene una señal estable para las poblaciones de partículas correspondientes a complejos tetraméricos de XaClp.

### 3.4 Detección de Clp por Western Blot

Para poder verificar que la proteína purificada mediante IMAC correspondía efectivamente a nuestra proteína de interés, se realizó un Western Blot siguiendo los pasos del protocolo detallados previamente (Bocanegra-Jimenez et al, 2021). La membrana de nitrocelulosa a la cual se transfirió la muestra a analizar indicó positivamente que ésta corresponde a la proteína Clp, de acuerdo a la presencia de una banda en la membrana de nitrocelulosa que coincide específicamente con la banda de nuestra proteína purificada obtenida por SDS-PAGE, al mismo tiempo que pudo observarse la presencia de la proteína tanto en la muestra t1 así como en la muestra del sonicado, mientras que no hubo presencia de ninguna señal positiva en la muestra correspondiente a t0 (Figura 8).



**Figura 8. Detección de Clp mediante Western Blot**

**Izquierda:** gel de poliacrilamida al 15% que muestra el electroferograma de una muestra de cultivo celular previo y posterior a la adición de IPTG (t0, t1, respectivamente), así como de una muestra del cultivo celular t1 posterior a ser procesada por sonicación (Sc) y una muestra de la purificación mediante IMAC correspondiente a la fracción eluida con el amortiguador de lavado adicionado con imidazol 200 mM (PP). **Derecha:** membrana de nitrocelulosa a la cual fueron

transferidas las muestras t0, t1, Sc y PP para la detección de la proteína de interés mediante Western Blot. En ambas imágenes se muestra el marcador Precision Plus Protein™ Dual Color Standards de Bio-Rad (M) como referencia de peso molecular.

### **3.5 Ensayos de cristalización**

Tras la observación periódica de la placa de cristalización bajo el microscopio estereoscópico, no se logró obtener la formación de cristales de la proteína de interés bajo las condiciones evaluadas. En cambio, se observó la presencia de precipitados amorfos en el 44% (43 de 96) de las gotas con la muestra de proteína. Esto último podría ser consecuencia de que la concentración de proteína utilizada en estos ensayos (6.72 mg/mL) fue menor que la concentración utilizada en los trabajos previos de Karina Rosales (10 mg/mL) que permitió la cristalización de XaClp.

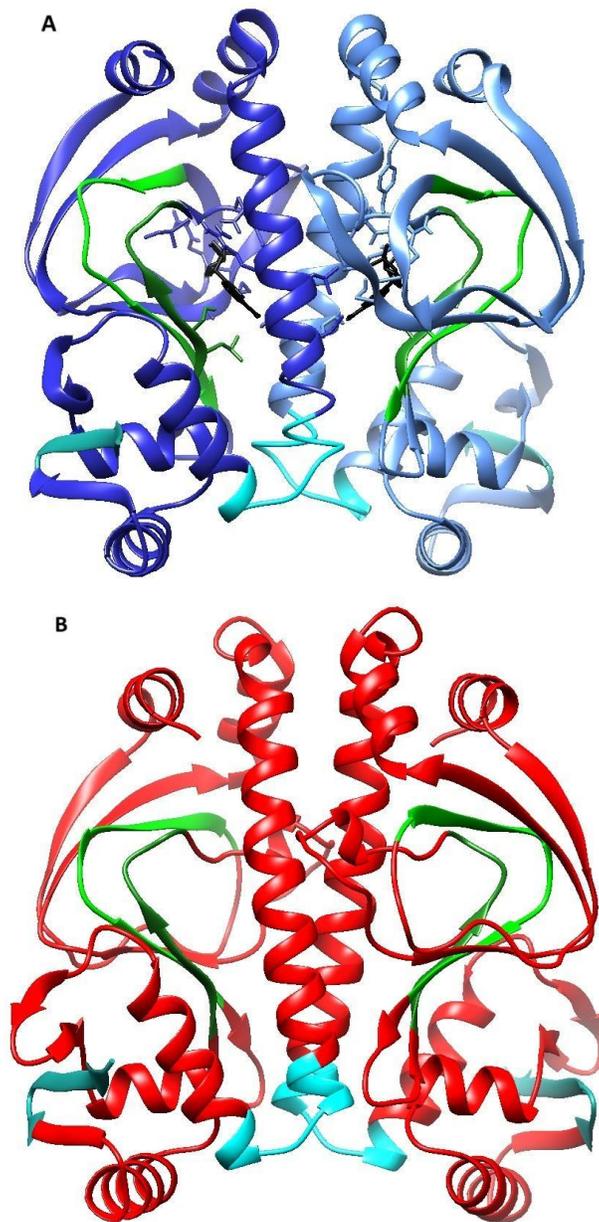
### **3.6 Análisis y comparación *in silico* de XaClp y EcCRP**

Debido a la tendencia de agregación de la proteína de interés observada durante la realización de este trabajo, se consideró analizar su secuencia en búsqueda de regiones implicadas en este fenómeno con la finalidad de determinar sugerencias a implementar en trabajos futuros para evitar la agregación de la proteína y obtener mayores avances en su caracterización bioquímica y estructural. De igual forma, se analizó la secuencia de EcCRP con la finalidad de tener un marco de referencia que nos permitiera localizar diferencias en XaClp con una posible implicación en su tendencia a la agregación.

Tras depositar las secuencias en formato FASTA correspondientes a XaClp y EcCRP, el servidor AGGRESCAN determinó que existen 7 y 8 regiones, respectivamente, propensas a la agregación de la proteína. 4 de las 7 regiones presentes en XaClp tienen un hotspot correspondiente en EcCRP, mientras que una región reportada en XaClp como hotspot no está determinada como tal en la región correspondiente en EcCRP (Figura 9 y 10). Por otra parte, el hotspot 4 reportado en XaClp tiene un hotspot parcialmente correspondiente en EcCRP, dada que la extensión del primero es más larga que la de este último (Tabla 1).

**Tabla 1. Sitios propensos a la agregación en XaClp y EcCRP.** Comparación de las secuencias de los hotspots en XaClp reportados por el servidor AGGRESCAN respecto a las secuencias correspondientes en EcCRP. Los “-“ se refieren a la ausencia de una secuencia correspondiente en EcCRP (hotspot 1 en XaClp) o bien a que dicha secuencia en EcCRP no está reportada como un sitio propenso a la agregación de la proteína (hotspot 4 en XaClp). La secuencia resaltada en azul del hotspot 5 en XaClp se refiere a los residuos correspondientes que conforman el hotspot equivalente en EcCRP.

Hotspot en XaClp	Secuencia	Hotspot correspondiente en EcCRP
1	5-NTTVVTTTV-13	-
2	53-GTLYYVISGSVSII-66	39-TLYYIVKGSVAVLI-52
3	76-VLGYFGSGE-84	60-MILSYLNQG-68
4	87-GEMGLFI-93	-
5	132-KILYAIGVQLSK-143	113-ILMRLS-118
6	154-RLAFLDVT-162	133-GNLAFLDVT-141
7	220-RGKTVVLYGTR-230	200-HGKTIVVYGTR-210



**Figura 9. Hotspots de agregación de *EcCRP-AMPc* y *XaClp* de acuerdo al servidor Aggrescan.**

Modelo de *EcCRP-AMPc* (PDB ID: 1ILB) (A) y de *XaClp* (B) (PDB aún no depositado) sobre los que se resaltan las regiones correspondientes a los hotspots de agregación comunes para ambas estructuras: hotspot 2 (verde claro), hotspot 3 (verde oscuro), hotspot 6 (cian) y hotspot 7 (verde mar).

<i>XaClp</i>	MSPG <sup>HS1</sup> <u>NTTVVTTTT</u> RNATPSLALDAGTIERFLAHSRRRYPTRTDVF <sup>HS1</sup> RPGDPA <u>GTLYYVIS</u>	54
<i>EcCRP</i>	-----MVLGKPQDPT <u>LEWFLSHCH</u> LHKYPSKSTLIHQGEKAE <u>TLYYIVK</u>	39
		HS1
<i>XaClp</i>	<sup>HS2</sup> <u>GSVSII</u> AEEDDDREL <sup>HS3</sup> <u>VLGYFGSGE</u> FV <sup>HS4</sup> <u>GEMGLFI</u> ESDTREVILRTRTQCELAEISYERLQQ	108
<i>EcCRP</i>	<u>GSVAVLI</u> <sup>HS2</sup> KDEEGKE <sup>HS3</sup> <u>MILSYLNQC</u> DFIGELGLFEEQERSAWVRAKTACEVAEISYK <u>KFRQ</u> <sup>HS4</sup>	93
<i>XaClp</i>	LFQTSLSPDAP <sup>HS5</sup> <u>KILYAIGVQLSK</u> RLLDTRKAS <sup>HS6</sup> <u>RLAFLDVTDR</u> IVRTLHDLAKEPEAMSH	162
<i>EcCRP</i>	<u>LIQV</u> -----NPD <sup>HS5</sup> <u>ILMRLS</u> QAQMARRLQVTSEKV <sup>HS6</sup> <u>GNLAFDVTGR</u> <sup>HS7</sup> <u>IAQTL</u> INLAKQPDAMTH	147
<i>XaClp</i>	PQGTQLRVSRQELARLVGCSREAGRVLKKLQADGLLHA <sup>HS7</sup> <u>RGKTVVLYGTR</u>	230
<i>EcCRP</i>	PDGMQIKITRQEIGQIVGCSRETVGRILKMLEDQNLISA <sup>HS8</sup> <u>HGKTIVVYGTR</u>	210
		HS8

**Figura 10. Disposición de los hotspots presentes en la secuencia de *XaClp* y *EcCRP*.**

De acuerdo al alineamiento entre las secuencias de *XaClp* y *EcCRP* puede observarse que algunos de los hotspots predichos por AGRRESCAN en la secuencia de cada una presentan algunas coincidencias, como su disposición en la secuencia. Sin embargo, algunos de los hotspots de *EcCRP* (enmarcados en negro) no tienen un hotspot correspondiente en *XaClp* (cuyos hotspots aparecen enmarcados en rojo), y viceversa.

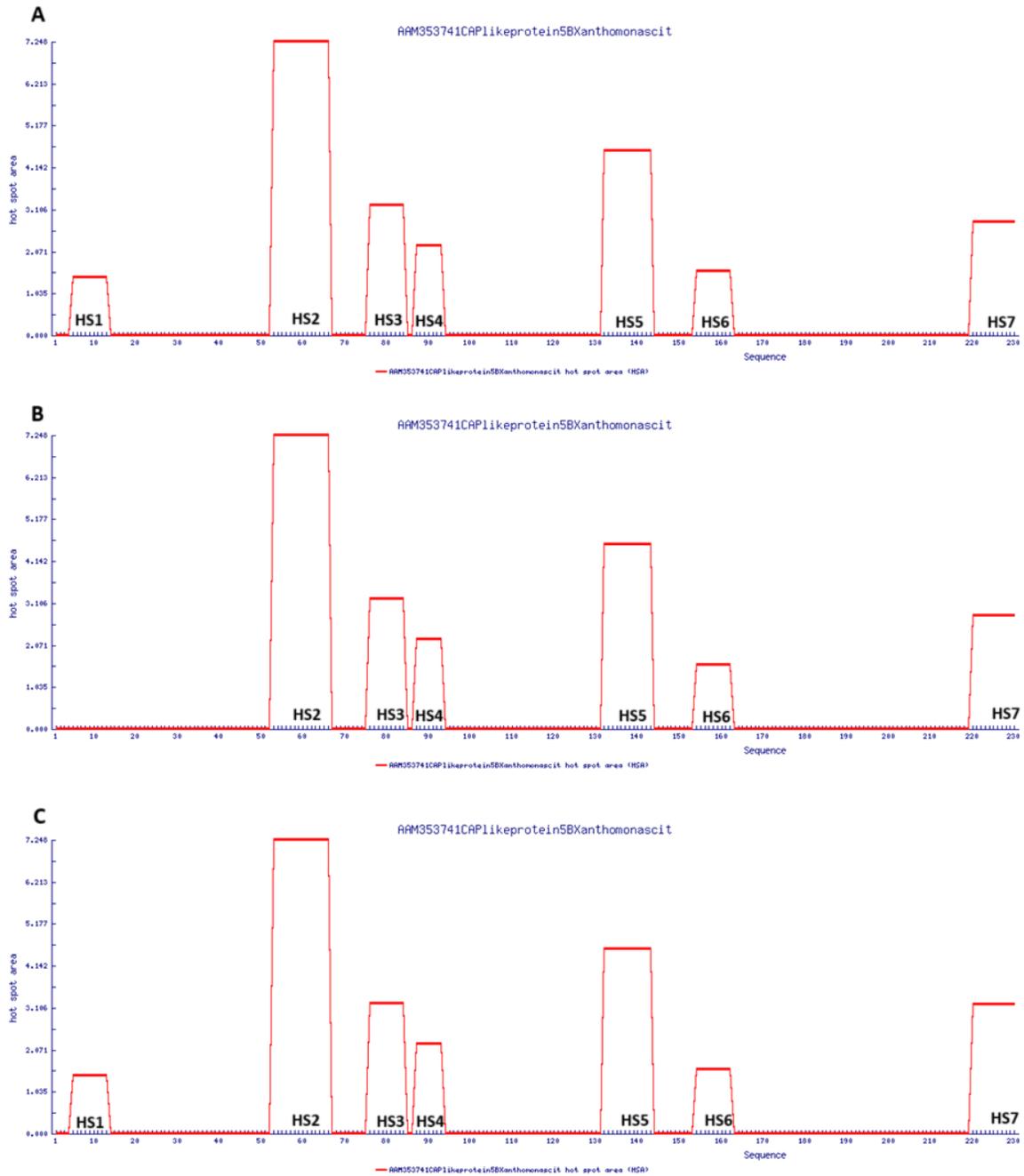
Es importante mencionar que el hotspot 1 de *XaClp* reportado por el servidor no tiene un hotspot correspondiente en *EcCRP*. Teniendo esto en cuenta, esta región podría ser relevante en el fenómeno de agregación particular de *XaClp*.

Por ello, se decidió analizar los puntajes de agregación asignados a cada residuo dentro de la secuencia del hotspot 1, en la cual se observa que los residuos N5, V9, T10, T11 y T12 presentan los puntajes más altos. En consecuencia, se realizó el análisis *in silico* del efecto de diferentes mutaciones sobre la secuencia de este hotspot, sustituyendo dichos residuos por residuos de alanina para encontrar qué modificaciones podrían reducir la tendencia de agregación de la proteína.

Como resultado, se encontró que las mutaciones N5A, T10A, T11A, T12A no reducen el puntaje de agregación para los residuos dentro del hotspot 1, y en algunos casos incluso lo incrementan (Tabla 2). En cambio, la mutación V9A no solo muestra una reducción en el puntaje de agregación de todos los residuos dentro del hotspot 1, si no que ya no es considerado como tal por el servidor (Figura 11). De acuerdo con esto, podríamos sugerir que una construcción de *XaClp* con dicha mutación permitiría eliminar los problemas de agregación observados en la proteína silvestre.

**Tabla 2. Efecto de las mutaciones puntuales sobre los residuos con mayor puntaje de agregación en la secuencia del hotspot 1 de XaClp de acuerdo al servidor AGGRESCAN.** Tras la sustitución puntual de los residuos con mayor contribución a la agregación en la secuencia del primer hotspot de XaClp (5-NTTVVTTTTV-13) predicho por el servidor AGGRESCAN por residuos de alanina, se puede observar que las mutaciones N5A, T10A, T11A y T12A no logran una reducción en el puntaje de agregación global del hotspot, sino que incluso aumenta. En cambio, la mutación V9A produce una reducción significativa en los puntajes de todos los residuos de la secuencia, al grado de que el servidor ya no la designa como un hotspot de agregación.

Construcción Residuo	XaClp WT	N5A	V9A	T10A	T11A	T12A
5 – Asparagina	0.139	0.217	0.000	0.151	0.149	0.148
6 - Treonina	0.139	0.217	0.000	0.151	0.149	0.148
7 – Treonina	0.139	0.217	0.000	0.151	0.149	0.148
8 – Valina	0.139	0.217	0.000	0.151	0.149	0.148
9 – Valina	0.139	0.217	0.000	0.151	0.149	0.148
10 – Treonina	0.139	0.217	0.000	0.151	0.149	0.148
11 – Treonina	0.139	0.217	0.000	0.151	0.149	0.148
12 – Treonina	0.139	0.217	0.000	0.151	0.149	0.148
13 – Valina	0.139	0.217	0.000	0.151	0.149	0.148



**Fig. 11. La mutación *in silico* V9A elimina el hotspot 1 en XaClp.**

El análisis con el servidor AGGRESCAN de la secuencia silvestre de XaClp muestra la presencia de 7 sitios de agregación en ella (A). Al sustituir la valina en la posición 9 por un residuo de alanina, puede apreciarse que el servidor solo predice 6 sitios de agregación, entre los cuales ya no se encuentra el hotspot 1 (B). También se muestra como referencia adicional el resultado de la mutación G228A del hotspot

con mayor exposición al solvente -hotspot 7-, en el que se puede observar que no produce ningún efecto significativo (C).

Adicionalmente, se utilizó la herramienta PDBePISA para el análisis de los 6 hotspots restantes de XaClp, depositando el archivo correspondiente a la estructura resuelta por Karina Rosales. Esta herramienta permite, entre otras cosas, llevar a cabo un análisis sobre el área accesible al solvente (ASA, por sus siglas en inglés) de una proteína (Tabla S1). Determinar esta característica es relevante para conocer si alguno de los hotspots presenta una superficie expuesta al exterior de la proteína y de esta manera nos permite proponer de una manera racional cuál de ellos podría tener una contribución relevante en la agregación de la proteína de interés.

Para ello, es necesario tener en cuenta el área de accesibilidad relativa al solvente (RSA, por sus siglas en inglés) de cada residuo que conforma a cada hotspot. Dicho valor es resultado de la relación entre el ASA de un residuo particular y su área máxima accesible al solvente (MaxASA) (Tabla S2). De manera convencional, un residuo particular puede considerarse como expuesto al solvente si su RSA es mayor al 20% (Chen & Zhou, 2005; Tien et. al, 2013).

De acuerdo a los resultados, puede apreciarse que los hotspots 2, 3, 6 y 7 destacan por presentar una gran proporción de residuos clasificados como accesibles al solvente. Por su parte, los hotspots 4 y 5 tienen una menor proporción de residuos accesibles al solvente, de los cuales ninguno de ellos supera un RSA del 20%. Entre los hotspots 2, 3, 6 y 7, existe una variación entre los residuos que superan el umbral del 20% para considerarlos expuestos al solvente: para el hotspot 2, son los residuos G53 y T54 junto con los residuos I59, S60 y G61; para el hotspot 3 son los residuos Y79 y los residuos G81, S82 y G83; para el hotspot 6 son los residuos T161 y D162; y, por último, para el hotspot 7 son los residuos R220, G221, K222, T223 y G228 (Tabla S3).

A continuación, entre estos residuos se consideraron aquellos de naturaleza hidrofóbica para evaluar el efecto de su mutación *in silico* por residuos de alanina. De esta manera, los residuos seleccionados fueron I59 y G61 para el hotspot 2;

Y79, G81 y G83 para el hotspot 3; y G221 y G228 para el hotspot 7. Sin embargo, la mutación *in silico* de estos residuos no generó una reducción relevante en sus puntajes de agregación (Tabla S3).

# Conclusión

El trabajo experimental realizado durante la investigación permitió ahondar más en algunos de los aspectos bioquímicos relevantes para la estabilidad estructural de XaClp. Se observó una mayor estabilidad en la señal de poblaciones de partículas con un radio hidrodinámico de ~5 nm en las muestras procesadas con el amortiguador optimizado (Tris-HCl 25 mM, glicerol 15 %, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 15 mM, CHAPS 0.1%, EDTA 1 mM, pH 7.5) respecto a aquellas procesadas con el amortiguador no optimizado (Tris-HCl 25 mM, glicerol 15%, KCl 500 mM, LiCl 25 mM,  $\beta$ -mercaptoetanol 1 mM, pH 8.9). De acuerdo al radio hidrodinámico de dichas poblaciones, su peso molecular equivale a 145 kDa. El peso molecular de una subunidad de la proteína Clp de *Xanthomonas axonopodis* es de 26 kDa, aunque es importante recalcar que no se eliminó la etiqueta 6XHis de la proteína recombinante, la cual presenta un peso molecular de 0.8 kDa, además de que la capa de solvatación alrededor de la proteína también conforma una fracción de su radio hidrodinámico. Teniendo esto en cuenta, dichas poblaciones sugieren estar formadas por complejos tetrámicos de la proteína de interés, lo que coincide con los hallazgos previos de Karina Rosales en los que se pudo apreciar la formación de cristales de XaClp con este tipo de complejos.

Resulta importante tener en cuenta que la proteína recombinante aún presenta la etiqueta 6XHis con sitio de corte para la enzima PreScission (PPS). Al incubar la proteína durante los trabajos iniciales con la enzima PPS, la escisión no resultó uniforme, probablemente debido al estado de agregación de XaClp. Si bien los efectos sobre el estado de agregación de proteínas recombinantes con la etiqueta 6XHis son variables (Carlson et. al, 2007; Majorek et. al, 2014), sería interesante realizar la incubación de la proteína recombinante en las condiciones del amortiguador optimizado en este trabajo, analizar la efectividad de la escisión enzimática en dicha condición y evaluar el estado de agregación de Clp sin la etiqueta 6XHis con sitio de corte para PPS para poder determinar si éstas contribuyen al fenómeno de agregación.

Por su parte, en cuanto a los ensayos de cristalización, no se obtuvo la formación de cristales bajo las condiciones realizadas en este trabajo. Teniendo en cuenta el alto porcentaje de gotas con precipitados amorfos en ellas, es necesario alcanzar una concentración de proteína acorde a la utilizada en los trabajos de cristalización de XaClp realizados anteriormente por Karina Rosales (2019) para permitir la formación de cristales.

Finalmente, de acuerdo al análisis bioinformático de los distintos hotspots de agregación de la proteína de interés, el hotspot 1 parece ser el de mayor contribución a este fenómeno, al observar que dicho hotspot deja de ser predicho como tal por la herramienta AGGREGSCAN tras sustituir los residuos de valina (de naturaleza hidrofóbica y de mayor puntaje de agregación en este hotspot) por residuos de alanina, a diferencia de lo observado con el resto de los hotspots de la proteína. Es importante añadir que la contribución de este hotspot en particular puede estar asociada a su aparente flexibilidad estructural, dado a que los trabajos previos en los que se han determinado la estructura cristalográfica de la proteína Clp de *Xanthomonas campestris* y *Xanthomonas axonopodis* no contienen la difracción correspondiente a esta región. De acuerdo con esto, se propone la generación y expresión de mutantes con la sustitución V9A como una metodología a analizar para intentar obtener una proteína con un perfil de menor agregación, y así facilitar el manejo y estabilidad necesarios para un estudio bioquímico y estructural en trabajos posteriores.

## Referencias

An SQ, Potnis N, Dow M, Vorhölter FJ, He YQ, Becker A, Teper D, Li Y, Wang N, Bleris Tang JL. Mechanistic insights into host adaptation, virulence and epidemiology of the phytopathogen *Xanthomonas*. *FEMS Microbiol Rev.* 2020 Jan 1;44(1):1-32. doi: 10.1093/femsre/fuz024. PMID: 31578554; PMCID: PMC8042644.

Arakawa T., Timasheff S. N. (1985) Theory of Protein Solubility. *Methods Enzymol.* 114, 49-77. doi:10.1016/0076-6879(85)14005-x

Auton M, Rösgen J, Sinev M, Holthauzen LM, Bolen DW. Osmolyte effects on protein stability and solubility: a balancing act between backbone and side-chains. *Biophys Chem.* 2011 Nov;159(1):90-9. doi: 10.1016/j.bpc.2011.05.012. Epub 2011 May 19. PMID: 21683504; PMCID: PMC3166983.

Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M., & Ebright, R. H. (2002). Structural basis of transcription activation: The Cap-actd-ADN complex. *Science*, 297(5586), 1562–1566. doi:10.1126/science.1076376

Bocanegra-Jiménez, F. Y., Montero-Morán, G.M., Lara-González, S. (2021) Purification and characterization of an FeII- and  $\alpha$ -ketoglutarate-dependent xanthine hydroxylase from *Aspergillus oryzae*, *Protein Expression and Purification*, Volume 183, 1046-5928, doi:10.1016/j.pep.2021.105862.

Bradbury, J.F. (1984) Genus II *Xanthomonas* Dowson 1939: Bergey's Manual of Systematic Bacteriology, 1 (Krieg, NR & Holt JG, eds), pp. 199-210. Williams and Wilkins, London.

Carson M, Johnson DH, McDonald H, Brouillette C, Delucas LJ. His-tag impact on structure. *Acta Cryst D*. 2007; 63:295-301 doi: 10.1107/S0907444906052024.

Chen CH, Lin NT, Hsiao YM, Yang CY, Tseng YH. Two non-consensus Clp binding sites are involved in upregulation of the gum operon involved in xanthan polysaccharide synthesis in *Xanthomonas AMPcestris* pv. *AMPcestris*. *Res Microbiol*. 2010 Sep;161(7):583-9. doi: 10.1016/j.resmic.2010.05.012. Epub 2010 Jun 15. PMID: 20558281.

Chen, H., Zhou, H.X. Prediction of solvent accessibility and site of deleterious mutations from protein sequence. *Nucleic Acids Res*. 2005 Jun 3;33(10):3192-9. doi: 10.1093/nar/gki633

Chin, K. H., Lee, Y. C., Tu, Z. L., Chen, C. H., Tseng, Y. H., Yang, J. M., Ryan, R. P., McCarthy, Y., Dow, J. M., Wang, A. H., and Chou, S. H. 2010. The AMPc receptor-like protein Clp is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in *Xanthomonas AMPcestris*. *J. Mol. Biol*. 396:646-662. doi:/10.1016/j.jmb.2009.11.076

de Crecy-Lagard V, Glaser P, Lejeune P, Sismeiro O, Barber CE, Daniels MJ, Danchin A. A *Xanthomonas AMPcestris* pv. *AMPcestris* protein similar to catabolite

activation factor is involved in regulation of phytopathogenicity. *J Bacteriol.* 1990 Oct;172(10):5877-83. doi: 10.1128/jb.172.10.5877-5883.1990. PMID: 2170330; PMCID: PMC526907.

de Marco A. Minimal information: an urgent need to assess the functional reliability of recombinant proteins used in biological experiments. *Microb Cell Fact.* 2008 Jul 23;7:20. doi: 10.1186/1475-2859-7-20

Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodríguez-Carmona E, Baumann K, Giuliani M, Parrilli E, Branduardi P, Lang C, Porro D, Ferrer P, Tutino ML, Mattanovich D, Villaverde A. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact.* 2008 Apr 4;7:11. doi: 10.1186/1475-2859-7-11. PMID: 18394160; PMCID: PMC2322954.

Graham, J. H., Gottwald, T. R., Cubero, J., & Achor, D. S. (2003). *xanthomonas axonopodis* pv. *citri*: Factors affecting successful eradication of citrus canker. *Molecular Plant Pathology*, 5(1), 1–15. doi: /10.1046/j.1364-3703.2004.00197.x

Golovanov AP, Hautbergue GM, Wilson SA, Lian LY. A simple method for improving protein solubility and long-term stability. *J Am Chem Soc.* 2004 Jul 28;126(29):8933-9. doi: 10.1021/ja049297h.

Harman, J. G., McKenney, K., & Peterkofsky, A. (1986). Structure-function analysis of three AMPC-independent forms of the AMPC receptor protein. *Journal of Biological Chemistry*, 261(35), 16332-16339. doi: 10.1016/s0021-9258(18)66570-6

Hayward, A.C. (1993). The hosts of *Xanthomonas* . In: Swings, J.G., Civerolo, E.L. (eds) *Xanthomonas*. Springer, Dordrecht. doi: 10.1007/978-94-011-1526-1\_1

He YW, Ng AY, Xu M, Lin K, Wang LH, Dong YH, Zhang LH. *Xanthomonas* AMPcestris cell-cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Mol Microbiol*. 2007 Apr;64(2):281-92. doi: 10.1111/j.1365-2958.2007.05670.x. Epub 2007 Mar 23. PMID: 17378922.

He, Y.W., & Zhang, L.H. (2008). Quorum sensing and virulence regulation in *xanthomonas campestris*. *FEMS Microbiology Reviews*, 32(5), 842–857. doi: 10.1111/j.1574-6976.2008.00120.x

Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. *Nat. Rev. Microbiol.* 15, 271–284. doi: 10.1038/nrmicro.2016.190

Kim, J., Adhya, S., & Garges, S. (1992). Allosteric changes in the AMPC receptor protein of *Escherichia coli*: hinge reorientation. *Proceedings of the National Academy of Sciences of the United States of America*, 89(20), 9700-9704. doi: 10.1073/pnas.89.20.9700

Körner, H., Sofia, H.J., & Zumft, W.G. (2003). Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs. *FEMS Microbiology Reviews* 27(5), 559-592. doi: 10.1016/s0168-6445(03)00066-4

Lawson CL, et. al (2004). Catabolite activator protein: ADN binding and transcription activation. *Curr Opin Struct Biol* 14:10-20. doi: 10.1016/j.sbi.2004.01.012

Leduc JL, Roberts GP. Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis* pv. *citri*. *J Bacteriol.* 2009 Nov;191(22):7121-2. doi: 10.1128/JB.00845-09. Epub 2009 Jul 24. PMID: 19633082; PMCID: PMC2772467.

Lorber B, Fischer F, Bailly M, Roy H, Kern D. Protein analysis by dynamic light scattering: methods and techniques for students. *Biochem Mol Biol Educ.* 2012 Nov-Dec;40(6):372-82. doi: 10.1002/bmb.20644. Epub 2012 Oct 10. PMID: 23166025.

Majorek KA, Kuhn ML, Chruszcz M, Anderson WF, Minor W. Double trouble - Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. *Protein Sci.* 2014; 23(10): 1359-68. doi: 10.1002/pro.2520

Popovych, N., Tzeng, S., Tonelli, M., Ebright, R. H., & Kalodimos, C. G. (2009). Structural basis for AMPC-mediated allosteric control of the catabolite activator protein. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17), 6927-6932. doi: 10.1073/pnas.0900595106

Qin, X., Liu, M., Yang, D., Zhang, X. (2010) Concentration-dependent aggregation of CHAPS investigated by NMR Spectroscopy. *The Journal of Physical Chemistry*. 114(11), 3863-3868. doi: 10.1021/jp911720w

Romeo, T., Vakulskas, C. A., & Babitzke, P. (2012). Post-transcriptional regulation on a global scale: Form and function of CSR/RSM systems. *Environmental Microbiology*, 15(2), 313-324. doi: 10.1111/j.1462-2920.2012.02794.x

Römling, U., Galperin, M. Y. & Gomelsky, M. (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* 77, 1–52

Rosales, K. & Lara-González, S. (2019). Caracterización bioquímica y estructural del factor de transcripción Clp y del complejo Clp/ADN. Instituto Potosino de Investigación Científica y Tecnológica.

Ryan RP, Vorhölter FJ, Potnis Net al... Pathogenomics of Xanthomonas: understanding bacterium-plant interactions. *Nat Rev Microbiol.* 2011;9:344–55

Sharma, H. K., Yu, S., Kong, J., Wang, J., & Steitz, T. A. (2009). Structure of APO-CAP reveals that large conformational changes are necessary for ADN binding. *Proceedings of the National Academy of Sciences of the United States of America*, 106(39), 16604-16609. doi: 10.1073/pnas.0908380106

Slater H, Alvarez-Morales A, Barber CE, Daniels MJ, Dow JM. A two-component system involving an HD-GYP domain protein links cell-cell signalling to pathogenicity gene expression in *Xanthomonas AMPcestris*. *Mol Microbiol*. 2000 Dec;38(5):986-1003. doi: 10.1046/j.1365-2958.2000.02196.x. PMID: 11123673.

Sharpe, T. (2014) Preventing protein aggregation. [https://www.biozentrum.unibas.ch/fileadmin/redaktion/05\\_Facilities/01\\_Technology\\_Platforms/BF/Protocols/Preventing\\_Protein\\_Aggregation.pdf](https://www.biozentrum.unibas.ch/fileadmin/redaktion/05_Facilities/01_Technology_Platforms/BF/Protocols/Preventing_Protein_Aggregation.pdf)

Starr, M. P. & Stephens, W. L. (1964) Pigmentation and taxonomy of the genus *Xanthomonas*. *J Bacteriol* 87: 293-302

Swaim CM, Brittain TJ, Marzolf DR, Kokhan O. Quantification of Metal Leaching in Immobilized Metal Affinity Chromatography. *J Vis Exp*. 2020 Jan 17;(155). doi: 10.3791/60690. PMID: 32009648.

Tao F, He YW, Wu DH, Swarup S, Zhang LH. The cyclic nucleotide monophosphate domain of *Xanthomonas AMPcestris* global regulator Clp defines a new class of cyclic di-GMP effectors. *J Bacteriol*. 2010 Feb;192(4):1020-9. doi: 10.1128/JB.01253-09. Epub 2009 Dec 11. PMID: 20008070; PMCID: PMC2812978.

Tien, M.Z., Meyer, A.G., Sydykova, D.K., Spierlman, S. J., Wilke, C.O. (2013) Maximum allowed solvent accessibilities of residues in proteins. *PLoS One*. Nov 21;8(11)e80635. doi: 10.1371/journal.pone.0080635.

Xiuhong, L., An, S., Tang, D., McCarthy, Y., Tang, J., Dow, J. M., & Ryan, R. P. (2012). RSMA regulates biofilm formation in *Xanthomonas ampcestris* through a regulatory network involving cyclic di-GMP and the CLP transcription factor. *PLOS ONE*, 7(12), e52646. doi: 10.1371/journal.pone.0052646

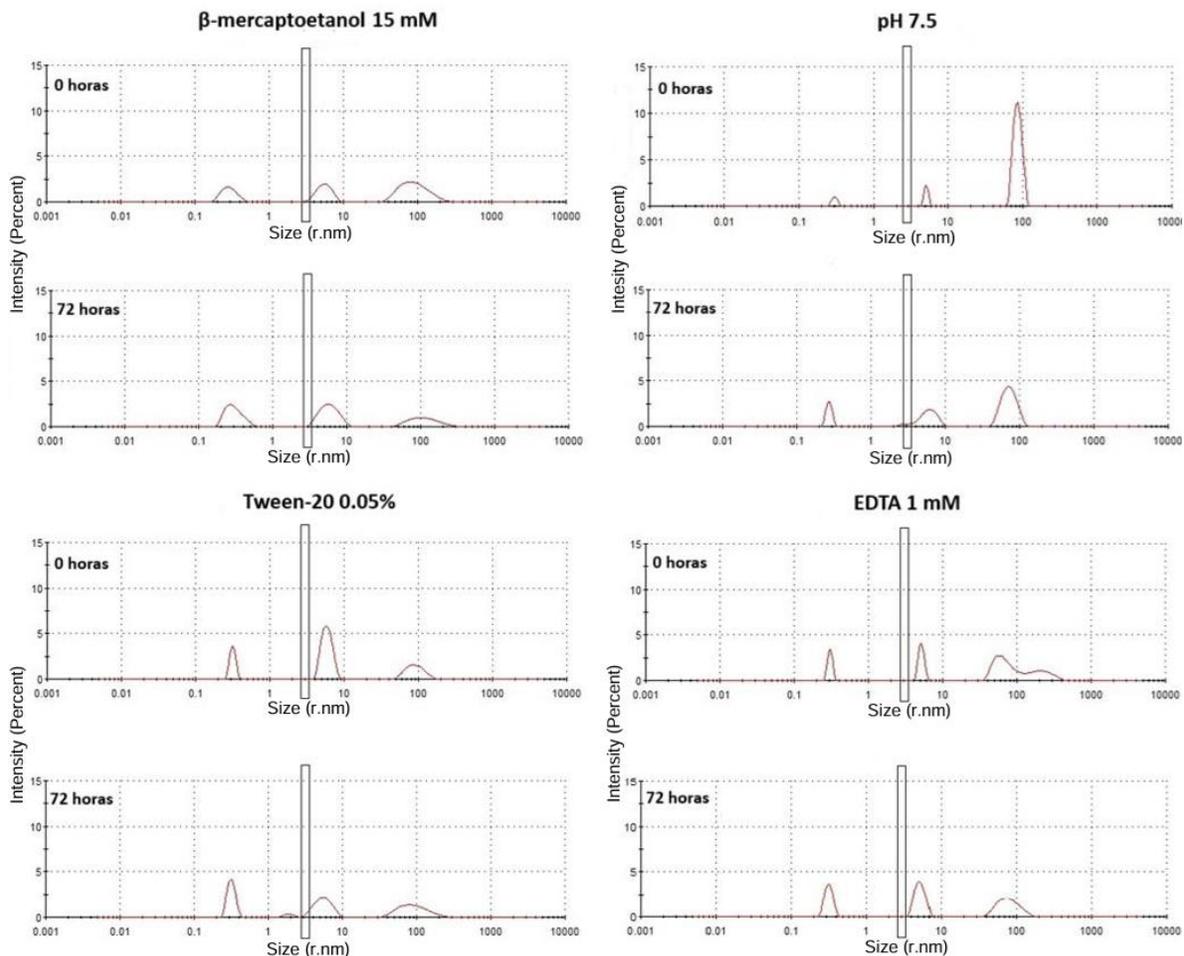
Zhang, S. B., Yang, D.Q., Jiang, Y. S., Ding, C.H. Competitive displacement of interfacial soy protein by Tween 20 and its effect of the physical stability of emulsions. *Food Hydrocolloids* 113, doi: 10.1016/j.foodhyd.2020.106515

# Figuras suplementarias, tablas y anexos

## Figura S1. Efecto de distintas condiciones sobre la agregación de XaClp

Si bien la presencia de grandes agregados de XaClp fue común en todas las muestras evaluadas, se pudo observar que condiciones concretas como la adición de  $\beta$ -mercaptoetanol 15 mM, el ajuste del pH a un valor de 7.5 y la adición de EDTA 1mM redujeron la señal de estas poblaciones tras 72 h. Adicionalmente, también mostraron una estabilización en la señal de poblaciones de complejos tetraméricos de Clp, correspondientes a un radio hidrodinámico de aproximadamente 5 nm. Por otro lado, se observa que la adición de Tween-20 0.05% no mostró un efecto significativo en estos aspectos.

La región correspondiente al radio hidrodinámico esperado del dímero de XaClp (3 nm) se muestra enmarcada en negro.



**Tabla S1. Residuos expuestos al solvente en los hotspots 2, 3, 4, 5, 6 y 7 de XaC1p.**

A. El análisis con la herramienta PDBePISA de la estructura de XaC1p obtenida por Karina Rosales muestra los residuos expuestos al solvente (azul claro) y los residuos inaccesibles al solvente (azul oscuro), junto a su respectiva área accesible al solvente (ASA). Además, se muestra el respectivo puntaje de agregación de cada residuo (aa4V) calculado por la herramienta AGGRESKAN.

B. El cálculo de la superficie relativa accesible al solvente (RSA) de cada residuo presente en los hotspots 2, 3, 6 y 7, producto de la relación entre el ASA de cada residuo calculado por PDBePISA y su respectivo MaxASA, muestra que hay una proporción particular de residuos que propiamente pueden considerarse accesibles al solvente (RSA > 20%) para cada uno de estos hotspots (marcados en negritas).

A.

Inaccessible residues						Solvent-accessible residues					
<b>Hotspot 2</b>						<b>Hotspot 3</b>					
aa	#	ASA	#	aa	aa4v	aa	#	ASA	#	aa	aa4v
A:GLY	53	33.35	53	G	0.029	A:VAL	76	6.54	76	V	0.249
A:THR	54	44.77	54	T	0.266	A:LEU	77	60.38	77	L	0.394
A:LEU	55	0.66	55	L	0.672	A:GLY	78	15.90	78	G	0.499
A:TYR	56	26.25	56	Y	0.677	A:TYR	79	85.16	79	Y	0.596
A:TYR	57	13.89	57	Y	0.621	A:PHE	80	23.65	80	F	0.286
A:VAL	58	0.00	58	V	0.648	A:GLY	81	28.19	81	G	0.304
A:ILE	59	42.16	59	I	0.843	A:SER	82	78.75	82	S	0.328
A:SER	60	49.31	60	S	0.657	A:GLY	83	24.89	83	G	0.328
A:GLY	61	28.34	61	G	0.730	A:GLU	84	42.69	84	E	0.042
A:SER	62	27.02	62	S	0.804						
A:VAL	63	0.00	63	V	0.623						
A:SER	64	7.95	64	S	0.264						
A:ILE	65	22.60	65	I	0.139						
A:ILE	66	1.67	66	I	-0.005						
<b>Hotspot 4</b>											
aa	#	ASA	#	aa	aa4v						
A:GLY	87	10.03	87	G	0.134						
A:GLU	88	18.14	88	E	0.389						
A:MET	89	73.21	89	M	0.748						
A:GLY	90	15.99	90	G	0.396						
A:LEU	91	20.79	91	L	0.186						
A:PHE	92	94.47	92	F	0.042						
A:ILE	93	97.64	93	I	0.181						

**Tabla S1. (Continuación)**

**Hotspot 5**

aa	#	ASA	#	aa	aa4v
A:ARG	132	38.10	132	K	0.095
A:ILE	133	0.50	133	I	0.334
A:LEU	134	75.61	134	L	0.479
A:TYR	135	164.22	135	Y	0.660
A:ALA	136	12.26	136	A	0.560
A:ILE	137	1.01	137	I	0.817
A:GLY	138	30.21	138	G	0.582
A:VAL	139	78.06	139	V	0.325
A:GLN	140	24.60	140	Q	0.059
A:LEU	141	64.11	141	L	0.216
A:SER	142	70.13	142	S	0.167
A:LYS	143	50.92	143	K	0.022

**Hotspot 6**

aa	#	ASA	#	aa	aa4v
A:ARG	154	45.43	154	R	0.082
A:LEU	155	71.21	155	L	0.016
A:ALA	156	70.56	156	A	0.296
A:PHE	157	173.32	157	F	0.283
A:LEU	158	35.56	158	L	0.111
A:ASP	159	82.93	159	D	0.111
A:VAL	160	29.36	160	V	0.160
A:THR	161	61.08	161	T	0.341
A:ASP	162	54.14	162	D	0.009

**Hotspot 7**

aa	#	ASA	#	aa	aa4v
A:ARG	220	62.07	220	R	0.047
A:GLY	221	68.08	221	G	0.070
A:LYS	222	59.32	222	K	0.070
A:THR	223	60.32	223	T	0.314
A:VAL	224	3.52	224	V	0.259
A:VAL	225	5.19	225	V	0.379
A:LEU	226	35.35	226	L	0.300
A:TYR	227	82.51	227	Y	0.283
A:GLY	228	102.77	228	G	0.283

**B.**

Hotspot (Secuencia)	Residuos con RSA > 20%
2 (53-GTLYYVISGSVSII-66)	G53 = 34.4% T54 = 27.5% I59 = 21.6% S60 = 34.5% G61 = 29.2%
3 (76-VLGYFGSGE-84)	Y79 = 32.3% G81 = 29.1% S82 = 55.1% G83 = 25.7%
6 (154-RLAFLDVTD-162)	T161 = 37.5% D162 = 29%
7 (220-RGKTVVLYGTR-230)	R220 = 23.4% G221 = 70.2% K222 = 25.8% T223 = 37% G228 = 98.8%

**Tabla S2. Área máxima accesible al solvente (MaxASA) de los 20 aminoácidos.**

En la tabla se muestra el valor del MaxASA (medido en Å<sup>2</sup>) de los 20 aminoácidos proteínogénicos de acuerdo a información empírica recopilada a partir de estructuras cristalográficas de diversas proteínas (Tien et. al, 2013)

<b>Aminoácido</b>	<b>Área máxima accesible al solvente (Å<sup>2</sup>)</b>
Alanina	121
Arginina	265
Asparagina	187
Aspartato	187
Cisteína	148
Glutamato	214
Glutamina	214
Glicina	97
Histidina	216
Isoleucina	195
Leucina	191
Lisina	230
Metionina	203
Fenilalanina	228
Prolina	154
Serina	143
Treonina	163
Triptófano	264
Tirosina	255
Valina	165

**Tabla S3. Efecto de mutaciones *in silico* de los residuos con RSA>20% y de naturaleza hidrofóbica en los hotspots 2, 3 y 7 en XaCip predichos por AGGRESCAN.**

La sustitución de los residuos con RSA>20% y de naturaleza hidrofóbica presentes en los hotspots 2, 3 y 7 (A#, B# y C#, respectivamente) muestran resultados variables, cuya característica compartida es que ninguna de las mutaciones elimina la predicción de la formación del hotspot por parte de AGGRESCAN. Entre estas mutaciones (enmarcadas en negro), algunas reducen el puntaje de agregación (aa4v) de algunos residuos en la secuencia del hotspot. Sin embargo, también existe el incremento de dicho valor de otros residuos respecto al puntaje de la secuencia WT.

A1			A2								
Cip I59A			Cip WT			Cip G61A			Cip WT		
#	aa	aa4v	#	aa	aa4v	#	aa	aa4v	#	aa	aa4v
53	G	0.029	53	G	0.029	53	G	0.029	53	G	0.029
54	I	0.266	54	I	0.266	54	I	0.266	54	I	0.266
55	L	0.466	55	L	0.672	55	L	0.672	55	L	0.672
56	Y	0.470	56	Y	0.677	56	Y	0.677	56	Y	0.677
57	Y	0.415	57	Y	0.621	57	Y	0.677	57	Y	0.621
58	V	0.447	58	V	0.648	58	V	0.703	58	V	0.648
59	A	0.636	59	I	0.843	59	I	0.898	59	I	0.843
60	S	0.450	60	S	0.657	60	S	0.712	60	S	0.657
61	G	0.524	61	G	0.730	61	A	0.786	61	G	0.730
62	S	0.598	62	S	0.804	62	S	0.860	62	S	0.804
63	V	0.417	63	V	0.623	63	V	0.678	63	V	0.623
64	S	0.264	64	S	0.264	64	S	0.319	64	S	0.264
65	I	0.139	65	I	0.139	65	I	0.195	65	I	0.139
66	I	-0.005	66	I	-0.005	66	I	-0.005	66	I	-0.005

A3			B1								
Cip I59A G61A			Cip WT			Cip Y79A			Cip WT		
#	aa	aa4v	#	aa	aa4v	#	aa	aa4v	#	aa	aa4v
53	G	0.029	53	G	0.029	76	V	0.117	76	V	0.249
54	I	0.266	54	I	0.266	77	L	0.261	77	L	0.394
55	L	0.466	55	L	0.672	78	G	0.366	78	G	0.499
56	Y	0.470	56	Y	0.677	79	A	0.464	79	Y	0.594
57	Y	0.470	57	Y	0.621	80	E	0.153	80	E	0.286
58	V	0.497	58	V	0.648	81	G	0.171	81	G	0.304
59	A	0.692	59	I	0.843	82	S	0.195	82	S	0.328
60	S	0.506	60	S	0.657	83	G	0.195	83	G	0.328
61	A	0.579	61	G	0.730	84	E	0.042	84	E	0.042
62	S	0.653	62	S	0.804						
63	V	0.472	63	V	0.623						
64	S	0.319	64	S	0.264						
65	I	0.195	65	I	0.139						
66	I	-0.005	66	I	-0.005						

Tabla S3. (Continuación)

B4			C1p WT			B5			C1p WT		
C1p Y79A G81A			C1p WT			C1p Y79A G83A			C1p WT		
#	aa	aa4v	#	aa	aa4v	#	aa	aa4v	#	aa	aa4v
76	V	0.117	76	V	0.249	76	V	0.117	76	V	0.249
77	L	0.317	77	L	0.394	77	L	0.261	77	L	0.394
78	G	0.423	78	G	0.499	78	G	0.366	78	G	0.499
79	A	0.519	79	Y	0.596	79	A	0.519	79	Y	0.596
80	F	0.209	80	F	0.286	80	F	0.209	80	F	0.286
81	A	0.227	81	G	0.304	81	G	0.227	81	G	0.304
82	S	0.250	82	S	0.328	82	S	0.250	82	S	0.328
83	G	0.250	83	G	0.328	83	A	0.250	83	G	0.328
84	E	0.098	84	E	0.042	84	E	0.098	84	E	0.042
85	F	0.004	85	F	-0.052	85	F	0.004	85	F	-0.052
						86	V	0.004	86	V	-0.052
						87	G	0.190	87	G	0.134
						88	E	0.389	88	E	0.389
						89	M	0.748	89	M	0.748
						90	G	0.396	90	G	0.396
						91	L	0.186	91	L	0.186
						92	F	0.042	92	F	0.042
						93	I	0.181	93	I	0.181

B6			C1p WT			B7			C1p WT		
C1p G81A G83A			C1p WT			C1p G81A G83A			C1p WT		
#	aa	aa4v	#	aa	aa4v	#	aa	aa4v	#	aa	aa4v
76	V	0.249	76	V	0.249	76	V	0.117	76	V	0.249
77	L	0.449	77	L	0.394	77	L	0.317	77	L	0.394
78	G	0.554	78	G	0.499	78	G	0.422	78	G	0.499
79	Y	0.707	79	Y	0.596	79	A	0.575	79	Y	0.596
80	F	0.397	80	F	0.286	80	F	0.264	80	F	0.286
81	A	0.415	81	G	0.304	81	A	0.287	81	G	0.304
82	S	0.439	82	S	0.328	82	S	0.306	82	S	0.328
83	A	0.439	83	G	0.328	83	A	0.306	83	G	0.328
84	E	0.153	84	E	0.042	84	E	0.153	84	E	0.042
85	F	0.059	85	F	-0.052	85	F	0.059	85	F	-0.052
86	V	0.004	86	V	-0.052	86	V	0.004	86	V	-0.052
87	G	0.190	87	G	0.134	87	G	0.190	87	G	0.134
88	E	0.389	88	E	0.389	88	E	0.389	88	E	0.389
89	M	0.748	89	M	0.748	89	M	0.748	89	M	0.748
90	G	0.396	90	G	0.396	90	G	0.396	90	G	0.396
91	L	0.186	91	L	0.186	91	L	0.186	91	L	0.186
92	F	0.042	92	F	0.042	92	F	0.042	92	F	0.042
93	I	0.181	93	I	0.181	93	I	0.181	93	I	0.181

C1			C1p WT			C2			C1p WT		
C1p G221A			C1p WT			C1p G228A			C1p WT		
#	aa	aa4v	#	aa	aa4v	#	aa	aa4v	#	aa	aa4v
220	R	0.102	220	R	0.047	220	R	0.047	220	R	0.047
221	A	0.126	221	G	0.070	221	G	0.070	221	G	0.070
222	K	0.126	222	K	0.070	222	K	0.070	222	K	0.070
223	T	0.369	223	T	0.314	223	T	0.314	223	T	0.314
224	V	0.314	224	V	0.259	224	V	0.314	224	V	0.259
225	V	0.434	225	V	0.379	225	V	0.434	225	V	0.379
226	L	0.300	226	L	0.300	226	L	0.356	226	L	0.300
227	Y	0.283	227	Y	0.283	227	Y	0.339	227	Y	0.283
228	G	0.283	228	G	0.283	228	A	0.339	228	G	0.283
229	T	0.283	229	T	0.283	229	I	0.339	229	I	0.283
230	R	0.283	230	R	0.283	230	R	0.339	230	R	0.283

Tabla S3. (Continuación)

**C3**

Cip G221A G228			Cip WT		
#	aa	aa4v	#	aa	aa4v
220	R	0.103	220	R	0.043
221	A	0.126	221	G	0.070
222	K	0.126	222	K	0.070
223	T	0.369	223	T	0.314
224	V	0.369	224	V	0.259
225	V	0.490	225	V	0.379
226	L	0.356	226	L	0.300
227	Y	0.339	227	Y	0.283
228	A	0.339	228	G	0.283
229	T	0.339	229	T	0.283
230	R	0.339	230	R	0.283

Tabla S3. (Continuación)

## Anexo 1. Condiciones del kit de cristalización JCSG Plus™.

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
A1	0.2 M	Lithium sulfate	0.1 M	Sodium acetate	4.5	50 % w/v	PEG 400
A2		None	0.1 M	Sodium citrate	5.5	20 % w/v	PEG 3000
A3	0.2 M	Ammonium citrate dibasic		None		20 % w/v	PEG 3350
A4	0.02 M	Calcium chloride dihydrate	0.1 M	Sodium acetate	4.6	30 % v/v	MPD
A5	0.2 M	Magnesium formate dihydrate		None		20 % w/v	PEG 3350
A6	0.2 M	Lithium sulfate	0.1 M	Phosphate/citrate	4.2	20 % w/v	PEG 1000
A7		None	0.1 M	CHES	9.5	20 % w/v	PEG 8000
A8	0.2 M	Ammonium formate		None		20 % w/v	PEG 3350
A9	0.2 M	Ammonium chloride		None		20 % w/v	PEG 3350
A10	0.2 M	Potassium formate		None		20 % w/v	PEG 3350
A11	0.2 M	Ammonium phosphate monobasic	0.1 M	Tris	8.5	50 % v/v	MPD
A12	0.2 M	Potassium nitrate		None		20 % w/v	PEG 3350
B1	0.8 M	Ammonium sulfate	0.1 M	Citrate	4.0		None
B2	0.2 M	Sodium thiocyanate		None		20 % w/v	PEG 3350
B3		None	0.1 M	BICINE	9.0	20 % w/v	PEG 6000
B4		None	0.1 M	HEPES	7.5	10 % w/v	PEG 8000
B5		None	0.1 M	Sodium cacodylate	6.5	40 % v/v	MPD
B6		None	0.1 M	Phosphate/citrate	4.2	40 % v/v	Ethanol
B7		None	0.1 M	Sodium acetate	4.6	8 % w/v	PEG 4000
B8	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	7.0	10 % w/v	PEG 8000
B9		None	0.1 M	Citrate	5.0	20 % w/v	PEG 6000
B10	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium cacodylate	6.5	50 % v/v	PEG 200
B11	1.6 M	Sodium citrate tribasic dihydrate pH 6.5		None			None
B12	0.2 M	Potassium citrate tribasic monohydrate		None		20 % w/v	PEG 3350
C1	0.2 M	Sodium chloride	0.1 M	Phosphate/citrate	4.2	20 % w/v	PEG 8000
C2	1.0 M	Lithium chloride	0.1 M	Citrate	4.0	20 % w/v	PEG 6000
C3	0.2 M	Ammonium nitrate		None		20 % w/v	PEG 3350
C4		None	0.1 M	HEPES	7.0	10 % w/v	PEG 6000
C5	0.8 M	Sodium phosphate monobasic monohydrate	0.1 M	Sodium HEPES	7.5		None
C6		0.8 M Potassium phosphate monobasic		None		40 % v/v	PEG 300
C7	0.2 M	Zinc acetate dihydrate	0.1 M	Sodium acetate	4.5	10 % w/v	PEG 3000
C8		None	0.1 M	Tris	8.5	20 % v/v	Ethanol
C9		None	0.1 M	Sodium/potassium phosphate	6.2	25 % v/v	1,2-Propanediol
C10		None	0.1 M	BICINE	9.0	10 % w/v	PEG 20,000
C11	2.0 M	Ammonium sulfate	0.1 M	Sodium acetate	4.6		None
C12		None		None		10 % w/v	PEG 1000
D1		None		None		10 % w/v	PEG 8000
D2		None		None		24 % w/v	PEG 1500
D2	0.2 M	Magnesium chloride hexahydrate	0.1 M	Sodium HEPES	7.5	30 % v/v	PEG 400
D3	0.2 M	Sodium chloride	0.1 M	Sodium/potassium phosphate	6.2	50 % v/v	PEG 200
D4	0.2 M	Lithium sulfate	0.1 M	Sodium acetate	4.5	30 % w/v	PEG 8000
D5		None	0.1 M	HEPES	7.5	70 % v/v	MPD
D6	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.5	20 % w/v	PEG 8000
D7	0.2 M	Lithium sulfate	0.1 M	Tris	8.5	40 % v/v	PEG 400
D8		None	0.1 M	Tris	8.0	40 % v/v	MPD
D9	0.17 M	Ammonium sulfate		None		25.5 % w/v	PEG 4000
D10		None		None		15 % v/v	Glycerol
D10	0.2 M	Calcium acetate hydrate	0.1 M	Sodium cacodylate	6.5	40 % v/v	PEG 300
D11	0.14 M	Calcium chloride dihydrate	0.07 M	Sodium acetate	4.6	14 % v/v	2-Propanol
D12		None		None		30 % v/v	Glycerol
D12	0.04 M	Potassium phosphate monobasic		None		16 % w/v	PEG 8000
D12		None		None		20 % v/v	Glycerol

Well #	Conc.	Salt	Conc.	Buffer	pH	Conc.	Precipitant
E1	1.0 M	Sodium citrate tribasic dihydrate	0.1 M	Sodium cacodylate	6.5		None
E2	2.0 M	Ammonium sulfate	0.1 M	Sodium cacodylate	6.5		None
E3	0.2 M	Sodium chloride	0.1 M	HEPES	7.5	10 % v/v	2-Propanol
E4	1.26 M	Ammonium sulfate	0.1 M	Tris	8.5		None
E5		None	0.1 M	CAPS	10.5	40 % v/v	MPD
E6	0.2 M	Zinc acetate dihydrate	0.1 M	Imidazole	8.0	20 % w/v	PEG 3000
E7	0.2 M	Zinc acetate dihydrate	0.1 M	Sodium cacodylate	6.5	10 % v/v	2-Propanol
E8	1.0 M	Ammonium phosphate dibasic	0.1 M	Sodium acetate	4.5		None
E9	1.6 M	Magnesium sulfate heptahydrate	0.1 M	MES	6.5		None
E10		None	0.1 M	BICINE	9.0	10 % w/v	PEG 6000
E11	0.16 M	Calcium acetate hydrate	0.08 M	Sodium cacodylate	6.5	14.4 % w/v	PEG 8000
E12		None	0.1 M	Imidazole	8.0	10 % w/v	PEG 8000
F1	0.05 M	Cesium chloride	0.1 M	MES	6.5	30 % v/v	Jeffamine® M-600
F2	3.2 M	Ammonium sulfate	0.1 M	Citrate	5.0		None
F3		None	0.1 M	Tris	8.0	20 % v/v	MPD
F4		None	0.1 M	HEPES	7.5	20 % v/v	Jeffamine® M-600
F5	0.2 M	Magnesium chloride hexahydrate	0.1 M	Tris	8.5	50 % v/v	Ethylene glycol
F6		None	0.1 M	BICINE	9.0	10 % v/v	MPD
F7	0.8 M	Succinic acid pH 7.0		None			None
F8	2.1 M	DL-Malic acid pH 7.0		None			None
F9	2.4 M	Sodium malonate dibasic monohydrate pH 7.0		None			None
F10	1.1 M	Sodium malonate dibasic monohydrate	0.1 M	HEPES	7.0	0.5 % v/v	Jeffamine® ED-2003
F11	1.0 M	Succinic acid	0.1 M	HEPES	7.0	1 % w/v	PEG 2000 MME
F12		None	0.1 M	HEPES	7.0	30 % v/v	Jeffamine® M-600
G1		None	0.1 M	HEPES	7.0	30 % v/v	Jeffamine® ED-2003
G2	0.02 M	Magnesium chloride hexahydrate	0.1 M	HEPES	7.5	22 % w/v	Poly(acrylic acid sodium salt) 5100
G3	0.01 M	Cobalt(II) chloride hexahydrate	0.1 M	Tris	8.5	20 % w/v	Polyvinylpyrrolidone
G4	0.2 M	TMAO	0.1 M	Tris	8.5	20 % w/v	PEG 2000 MME
G5	0.005 M	Cobalt(II) chloride hexahydrate	0.1 M	HEPES	7.5	12 % w/v	PEG 3350
	0.005 M	Cadmium chloride hemi(pentahydrate)					
	0.005 M	Magnesium chloride hexahydrate					
	0.005 M	Nickel(II) chloride hexahydrate					
G6	0.2 M	Sodium malonate dibasic monohydrate		None		20 % w/v	PEG 3350
G7	0.1 M	Succinic acid		None		15 % w/v	PEG 3350
G8	0.15 M	DL-Malic acid		None		20 % w/v	PEG 3350
G9	0.1 M	Potassium thiocyanate		None		30 % w/v	PEG 2000 MME
G10	0.15 M	Potassium bromide		None		30 % w/v	PEG 2000 MME
G11	2.0 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5		None
G12	3.0 M	Sodium chloride	0.1 M	BIS-Tris	5.5		None
H1	0.3 M	Magnesium formate dihydrate	0.1 M	BIS-Tris	5.5		None
H2	1.0 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5	1 % w/v	PEG 3350
H3		None	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H4	0.2 M	Calcium chloride dihydrate	0.1 M	BIS-Tris	5.5	45 % v/v	MPD
H5	0.2 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	45 % v/v	MPD
H6	0.1 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	17 % w/v	PEG 10,000
H7	0.2 M	Ammonium sulfate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H8	0.2 M	Sodium chloride	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H9	0.2 M	Lithium sulfate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H10	0.2 M	Ammonium acetate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H11	0.2 M	Magnesium chloride hexahydrate	0.1 M	BIS-Tris	5.5	25 % w/v	PEG 3350
H12	0.2 M	Ammonium acetate	0.1 M	HEPES	7.5	45 % v/v	MPD

Lista de las 96 condiciones que conforman el kit de cristalización JCSG Plus™ utilizados para el ensayo de cristalización de XaClp. Se depositaron 50 µL de cada una sobre el pocillo mayor de la placa de cristalización, y en un volumen de 0.6 µL en cada micropocillo en el que se depositaron la muestra de XaClp ultrafiltrada y el amortiguador optimizado como control negativo.