

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS

INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS

CENTRO DE INVESTIGACIÓN EN DINÁMICA CELULAR

Determinación de estructuras cristalográficas mediante el uso
de proteínas de fusión

TESIS PROFESIONAL PARA OBTENER EL GRADO DE:

DOCTORA EN CIENCIAS

PRESENTA:

María Cristina Cardona Echavarría

DIRECTOR: Enrique Rudiño Piñera

CUERNAVACA, MORELOS

MAYO, 2024

JURADO EVALUADOR

Presidente: Dra. Carmen Nina Pastor Colón

Secretario: Dr. Rodrigo Said Razo Hernández

Vocal: Dra. Adela Rodríguez Romero

Vocal: Dr. Cesar Millán Pacheco

Vocal: Dra. Lina Andrea Rivillas Acevedo

Suplente: Dr. Carlos Daniel Amero Tello

Suplente: Dr. Enrique Rudiño Piñera

PUBLICACIONES ASOCIADAS

Cardona-Echavarría MC, Santillán C, Miranda-Blancas R, Stojanoff V, Rudiño-Piñera E. Unveiling success determinants for AMB-assisted phase expansion of fusion proteins in ARP/wARP. *J Struct Biol.* 2024 Jun;216(2):108089. <https://doi.org/10.1016/j.jsb.2024.108089>

AGRADECIMIENTOS

Al Dr. Enrique Rudiño Piñera, por su guía, sus enseñanzas, su apoyo y su amistad, que me acompañaron siempre y me hicieron crecer profesional y personalmente; gracias por enseñarme, pero también por aprender juntos.

A los integrantes de mi comité tutorial, la Dra. Nina Pastor Colón y el Dr. Carlos Amero Tello, por guiarme en el mejor camino para culminar esta tesis.

Al comité evaluador, por sus valiosos aportes para obtener una mejor versión del presente escrito.

Al Instituto de Biotecnología de la UNAM, por recibirme y acogerme como un miembro más de su comunidad.

A mis compañeros del laboratorio de Bioquímica Estructural, por la compañía, retroalimentación y apoyo.

A la Dra. Vivian Stojanoff y la Mtra. Carmén Santillán, por sus valiosos aportes para la escritura del artículo que surgió a partir de este trabajo.

A mi familia, Irma, Ramiro y Ana Caro, por su amor incondicional.

A Ricardo, por ser mi familia, mi compañero, mi maestro, mi colega, mi apoyo incondicional.

A mis adorados amigos en Colombia y México, una luz en momentos de oscuridad.

A cada persona que me apoyó durante esta etapa.

Tabla de contenido

1. Resumen	6
2. Abstract.....	8
3. Acrónimos.....	10
4. Introducción	12
5. Marco teórico	13
5.1 Proteínas de fusión	13
5.2 Cristalografía de proteínas.....	16
5.3 Proteínas de fusión en cristalografía de proteínas	20
5.4 Proteínas reguladoras HiID y HiIE.....	25
5.5 γ S y β A1 Cristalina humana	27
6. Justificación	31
7. Hipótesis	33
8. Objetivos.....	34
8.1 Objetivo general.....	34
8.2 Objetivos específicos	34
9. Cristalización asistida por etiquetas en proteínas de difícil cristalización	35
9.1 Métodos y materiales	35
9.1.1 Construcciones de proteínas de fusión.....	35
9.1.2 Expresión de TRX-HiIE	37
9.1.3 Expresión de MBP-HiID	38
9.1.4 Expresión de SUMO- γ S Cristalina y SUMO- β A1 Cristalina humanas ...	38
9.1.5 Purificación de TRX-HiIE	39
9.1.6 Purificación de MBP-HiID.....	40
9.1.7 Purificación de SUMO- γ S cristalina humana	41
9.1.8 Purificación de SUMO- β A1 cristalina humana	42
9.1.9 Cristalización de las Proteínas de Fusión	44
9.2 Resultados	44
9.2.1 Expresión y purificación de TRX-HiIE	44
9.2.2 Pruebas de cristalización de TRX-HiIE	48

9.2.3	Expresión y purificación de MBP-HiID	49
9.2.4	Pruebas de cristalización de MBP-HiID	51
9.2.5	Expresión y purificación de SUMO- γ S cristalina	52
9.2.6	Pruebas de cristalización de SUMO- γ S cristalina	54
9.2.7	Expresión y purificación de SUMO- β A1 cristalina.....	55
9.2.8	Pruebas de cristalización de SUMO- β A1 cristalina.....	59
9.3	Discusión	59
10.	Determinación estructural mediante la expansión de fases en proteínas de fusión depositadas en el PDB	68
10.1	Métodos y materiales	68
10.1.1	Estructuras cristalográficas de PFs depositados en PDB	68
10.1.2	Determinación de la estructura de la proteína de interés.....	68
10.1.3	Análisis estadístico de los resultados obtenidos con ARP/wARP	72
10.2	Resultados	73
10.2.1	Estructuras cristalográficas de PFs depositados en PDB	73
10.2.2	Determinación de la estructura de la proteína de interés.....	73
10.2.3	Análisis estadístico de los resultados obtenidos con ARP/wARP	76
10.3	Discusión	83
11.	Conclusiones	87
12.	Bibliografía.....	89
13.	Glosario	103
14.	Material suplementario	106
14.1	Tabla suplementaria 1.....	106
14.2	Tabla suplementaria 2.....	120
14.3	Tabla suplementaria 3.....	124
14.4	Bibliografía suplementaria.....	125

1. Resumen

Las proteínas de fusión constituyen una herramienta biotecnológica, que suelen ser usadas como biosensores, para mejoramiento enzimático, fármacos, purificación de proteínas entre otras aplicaciones. En cristalografía de proteínas es común el uso de éstas para facilitar el proceso de expresión y purificación, al fusionar la proteína de interés con un péptido o proteína que funciona como una etiqueta. La práctica más común es retirar la etiqueta previo a las pruebas de cristalización, sin embargo, varios autores han resaltado que no retirarla puede ser beneficioso, ya que podría ayudar a disminuir los dos cuellos de botella en el proceso de cristalización de proteínas: la formación de cristales y el proceso de resolución de fases. En el presente trabajo se evaluó el uso de las proteínas etiqueta como herramienta en la cristalografía de proteínas desde los dos enfoques propuestos: 1) evaluando los beneficios de no retirar la etiqueta en el proceso de purificación y cristalización de cuatro proteínas que han mostrado ser de difícil cristalización; y 2) evaluando la capacidad de determinar estructuras completas a partir de las fases de la etiqueta mediante la expansión de fases a la proteína de interés. Lo anterior se hizo sin tener ninguna información estructural previa de la proteína de interés, mediante el uso de remplazo molecular de la etiqueta y con la construcción automática de todo el complejo con ARP/wARP. Además, se analizaron algunas características de las etiquetas, proteínas de interés, el cristal y los datos cristalográficos, utilizando diferentes aproximaciones estadísticas, con el fin de identificar qué factores pueden influir en la capacidad de extender las fases a partir de la etiqueta. En este trabajo fue posible expresar y purificar las proteínas HiID, HiIE, γ S cristalina y β A1 cristalina

en forma de proteínas de fusión con las etiquetas Tiorredoxina (TRX), Proteína de unión a maltosa (MBP) y la proteína SUMO (por sus siglas en inglés *small ubiquitin-like modifier*). Esto permitió obtener purezas y solubilidades mayores a las obtenidas sin la etiqueta, alcanzando las condiciones que permitieron poner cientos de pruebas de cristalización, algo que no había sido sencillo lograr sin la presencia de la etiqueta. Para el caso específico de estas proteínas de difícil cristalización, y con las etiquetas elegidas, no fue posible obtener cristales. Estos resultados lejos de indicar que la aproximación es incorrecta nos indican que, al igual que con las proteínas no fusionadas, es necesario probar diferentes condiciones que en el caso de las proteínas de fusión incluyen múltiples etiquetas y enlazadores. Además, fue posible identificar la presencia de 385 estructuras de proteínas de fusión depositadas en el PDB con cinco de las etiquetas más reportadas. De estas 385 estructuras, fueron elegidas 116 proteínas de fusión para evaluar la expansión de las fases de la etiqueta a la totalidad de la proteína. Usando estas etiquetas como generadores de fases fue posible reconstruir 36 de las 116 estructuras de proteínas de fusión. Los análisis estadísticos permitieron identificar que la resolución, el factor B de Wilson, el porcentaje de solvente, la completitud, la multiplicidad, el porcentaje de la proteína etiqueta en la proteína de fusión, y la longitud del enlazador desempeñan papeles fundamentales al utilizar nuestro enfoque. En los casos en los que no existe un homólogo estructural, este método merece ser incluido en el conjunto de herramientas de los cristalógrafos de proteínas.

2. Abstract

Fusion proteins are a powerful tool in biotechnology, used for biosensors, enzymatic improvement, drug design, and protein purification. These proteins are commonly used in protein crystallography to facilitate the expression and purification process, by fusing the target protein to an affinity tag, which can be a peptide or a protein. Nevertheless, the most common practice is to remove the tag before crystallization. However, some researchers have shown the utility of not removing them because they can decrease the two bottlenecks in protein crystallization: the crystal formation and the phases determination. The present study evaluated the usefulness of tag proteins in protein crystallography using two approaches: 1) Evaluation of the benefits of not removing the tag on four difficult crystallization proteins and 2) Expanding the phases from the tag to the target to determine the complete structure. The last was carried out without any structural information of target proteins, using the tag coordinates for molecular replacement followed by an automatic model building with ARP/wARP. Afterward, some characteristics of the tags and the target proteins, the crystal, and the data sets were analyzed with different statistical approximations to identify factors that may influence the ability to extend the phases from the tag. In this study, it was possible to express and purify the proteins HilD, HilE, γ S crystalline and β A1 crystalline fused with the thioredoxin, maltose binding protein and the small ubiquitin-like modifier. The tag presence enables hundreds of crystallization tests to be conducted, by increasing the purity and solubility. This would have been difficult to achieve without the tag. It was not possible to obtain protein crystals from those difficult crystallization proteins. These results do not

indicate a wrong approximation but rather show the need to evaluate multiple conditions, including different tags and linkers. This study also identified 385 fusion protein structures with five of the most used tags in the PDB. 116 fusion proteins deposited on PDB were chosen to test the phase extension starting from the tag. Using these tags as phase generator it was possible to construct 36 of the 116 fusion protein structures. Statistical analysis identified that resolution, Wilson B factor, solvent percentage, completeness, multiplicity, percentage of tag protein in the fusion protein, and the linker length play critical roles when using our approach. In cases where a structural homologue does not exist, this method deserves to be included in the toolkit of protein crystallographers.

3. Acrónimos

PF: Proteína de fusión

MBP: Proteína de unión a maltosa (*Maltose binding protein*)

TRX: Tiorredoxina

SUMO: *Small ubiquitin-related modifier*

GST: Glutación transferasa

GFP: Proteína verde fluorescente (*Green fluorescent protein*)

RM: Reemplazo molecular

MA: Métodos anómalos

MI: Métodos isomorfos

MIR: *Multiple isomorphous replacement*

SIR: *Single isomorphous replacement*

MAD: *Multiple anomalous dispersion*

SAD: *Single anomalous dispersion*

MIRAS: *Multiple isomorphous replacement with anomalous scattering*

SIRAS: *Single isomorphous replacement with anomalous scattering*

HisTag: Etiqueta de histidina

PDB: *Protein Data Bank*

RMSD: *Root mean square deviation*

LB: Medio Luria Bertani

IPTG: Isopropil- β -D-1-tiogalactopiranosido

BLAST: *Basic local alignment search tool*

PCA: Análisis de componentes principales (*Principal component analysis*)

PCAmix: Análisis de componentes principales de datos mixtos (*Principal component analysis of mixed data*)

4. Introducción

Las proteínas de fusión (PF) son un tipo de proteína en la que dos o más dominios se integran en una molécula [1,2]. Pueden ocurrir naturalmente, pero a menudo están diseñadas para aplicaciones biotecnológicas como biosensores, ingeniería enzimática, purificación de proteínas y tratamientos médicos [1,3–6]. Algunas proteínas de fusión están compuestas por una proteína de interés fusionada con una etiqueta o *tag*, que puede ser un péptido o una proteína; la adición de la etiqueta aumenta la expresión y la estabilidad, también mejora la solubilidad y puede proteger a la proteína de interés de proteólisis. Además, muchas de ellas pueden facilitar la purificación mediante el uso de cromatografía de afinidad. Este método puede dar como resultado, en unos pocos pasos de purificación, purezas de hasta el 90 % [7,8]. En la cristalografía de proteínas, obtener proteínas con alta pureza es de suma importancia, por lo que es común el uso de PFs antes de los ensayos de cristalización [9–11]; sin embargo, lo más común es eliminar las etiquetas, antes de las pruebas de cristalización [7,12]. Algunos investigadores afirman que es necesario eliminar la etiqueta, ya que puede afectar la estructura cristalográfica o incluso la capacidad de cristalización, mientras que otros, afirman que su influencia es nula e incluso puede ser beneficiosa [12].

En el presente trabajo se evaluaron los beneficios de usar PFs en la cristalografía, tanto en el proceso de formación de cristales como el de resolución de fases del cristal, con el fin de determinar su utilidad y qué características se requieren para poder utilizar las PFs como una herramienta en la cristalografía de proteínas.

5. Marco teórico

5.1 Proteínas de fusión

Las PFs, también conocidas como proteínas quiméricas, son un tipo de proteína en la que dos o más dominios proteicos de origen diferente se integran en una sola molécula [1,3,4]. Es posible encontrar ejemplos de este tipo de proteínas de forma natural, algunos asociados a enfermedades, como en el caso de la leucemia mieloide aguda, donde rearrreglos cromosomales generan la expresión de proteínas de fusión oncogénicas, que en general pueden interferir con el proceso de diferenciación mieloide [13,14]. Otras proteínas de fusión encontradas en forma natural se han asociado a eventos donde la cercanía de dos dominios proteicos con funciones independientes confiere una ventaja evolutiva, esto debido a que la mayor parte de las proteínas reportadas están relacionadas funcionalmente, y están implicadas en los mismos complejos proteicos y/o procesos bioquímicos [5,15–18].

Sin embargo, la mayoría de las proteínas de fusión existentes son el resultado de su diseño para aplicaciones biotecnológicas [1]. Algunos de los usos más comunes incluyen: 1) el diseño de proteínas quiméricas que permitan mejorar el desempeño de proteínas de interés y reducir sus costos de producción [19,20]; 2) en ingeniería enzimática, donde se busca aumentar las tasas de reacción consecutiva o generar enzimas multifuncionales [1]; 3) su uso como biosensores o reporteros, donde su principal representante son las PFs con la Proteína verde fluorescente [1,21]; 4) su

aplicación para tratamientos médicos, en los cuales medicamentos de origen proteico se fusionan a moléculas como anticuerpos o péptidos señal, que pueden llegar a conferir un mayor tiempo de vida, transporte eficaz a través de las barreras biológicas, mayor especificidad, entre otros beneficios [2,4,6]; y 5) su diseño con el fin de facilitar el proceso de purificación y/o cristalización de proteínas [7].

Para el uso de PFs en procesos como señalización, expresión y/o purificación, la proteína de interés se fusiona con una etiqueta, que puede ser desde un péptido hasta una proteína con estructura tridimensional definida [7]; la adición de la etiqueta confiere beneficios como aumentar la expresión y estabilidad, mejorar la solubilidad y en muchos casos, proteger a la proteína de interés de la proteólisis [22,23]. La presencia de la etiqueta facilita la purificación mediante técnicas cromatográficas, ya que muchos de ellos son específicos para purificar mediante cromatografía de afinidad (Tabla 5.1-1). Estos métodos pueden dar como resultado altos niveles de pureza en pocos pasos [7,8,24].

Tabla 5.1-1: Etiquetas de afinidad más utilizadas [7].

Etiqueta de afinidad	Residuos	Tamaño (kDa)
Poli-Arginina	5-6 (usualmente 5)	0.80
Poli-Histidina	2-10 (usualmente 6)	0.84
Flag	8	1.01
Strep-tag II	8	1.06
HAT- (<i>Histidine affinity tag</i>)	19	2.31
3x Flag	22	2.73
Péptido de unión a Calmodulina	26	2.96
Dominios de unión a celulosa	27-189	3.00-20.00
Péptido de unión a estreptavidina	38	4.03
Dominio de unión a quitina	51	5.59
Glutación transferasa	211	26.00
Proteína de unión a maltosa	396	40.00

En las PFs donde la etiqueta es una proteína, se pueden identificar tres secciones principales: la proteína de interés, la proteína etiqueta y una secuencia enlazadora o *linker* (Figura 5.1-1). La selección de la etiqueta y del enlazador es un proceso fundamental a la hora del diseño de la PF. La elección de la etiqueta depende del objetivo que se busca cumplir con la construcción de la PF, del sistema de expresión a usar y de las características de la proteína de interés [7]. La elección de los enlazadores también es decisiva, pues puede ayudar a mejorar la actividad biológica, aumentar o disminuir el movimiento y homogeneidad de la muestra, aumentar el rendimiento de expresión, lograr mejores perfiles farmacocinéticos y hasta ayudar a promover o impedir la formación de cristales de proteína [2,3].

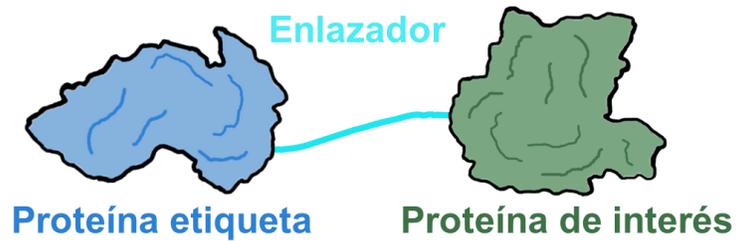


Figura 5.1-1: Estructura general de una PF.

5.2 Cristalografía de proteínas

La cristalografía de proteínas es una técnica que permite determinar la estructura tridimensional de proteínas, ácidos nucleicos e inclusive de virus o ribosomas [10]. Para esta técnica, un cristal de proteína es irradiado con una fuente de rayos X, y la radiación dispersada por el cristal es colectada [25]. Cuando los rayos X interactúan con un electrón que se encuentra dentro de los átomos que constituyen a una proteína, éste se convierte en fuente de radiación electromagnética secundaria de nuevos frentes de onda de rayos X; esta nueva radiación origina un patrón de difracción resultado de la interferencia constructiva y destructiva de la radiación dispersada [26]. Es así que, el patrón de difracción obtenido es una consecuencia directa de la estructura molecular de las proteínas que forman el cristal, la geometría (tamaño y forma) de la celda unitaria y la longitud de onda de los rayos X [27,28].

Debido a que el primer paso para realizar cristalografía de proteínas es la obtención de un cristal de buena calidad, es necesario obtener cantidades considerables de la proteína, ya sea directamente del organismo que la produce o usando proteínas

recombinantes [11]. Para la obtención de un cristal, es además deseable que la proteína se encuentre con una alta pureza (alrededor del 90%), y sea soluble y homogénea [25].

Estas características necesarias para obtener un cristal convierten el proceso de cristalogénesis en uno de los cuellos de botella en esta técnica. Esto se debe a que, en muchos sistemas, no es una tarea simple obtener grandes cantidades de proteína soluble y homogénea, que permitan llegar a un estado de sobresaturación que promueva la cercanía entre las moléculas presentes y permita que éstas interactúen [28]. Además, la formación de un cristal se da mediante interacciones específicas, altamente direccionales y organizadas (Figura 5.2-1), razón por la cual además de requerirse una proteína con alto grado de pureza, deben explorarse diferentes condiciones como cambios en el pH, concentraciones de sales y temperaturas, entre otras [9,10].

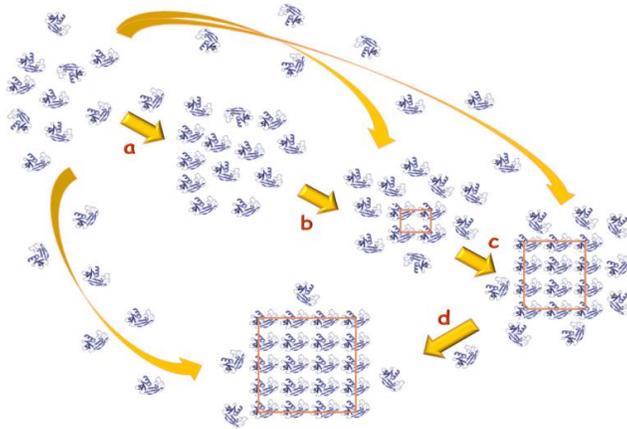


Figura 5.2-1: Modelo de formación de un cristal de proteína tomado de Krauss y colaboradores (2013) [9]. El aumento de la sobresaturación promueve la asociación entre las moléculas, que comienzan a organizarse en grandes agregados desordenados (a) las moléculas dentro de los núcleos de tales agregados se reorientan, redistribuyen y forman interacciones geoméricamente más rigurosas. (b) Estas últimas interacciones tienden a ordenar y estabilizar el núcleo agregado, que aumenta para producir un núcleo crítico. (c) Finalmente, se convierte en un verdadero cristal. (d) Las moléculas libres se adsorben en la superficie del cristal y éste aumenta su tamaño.

Una vez que se logra obtener un cristal adecuado de proteína, éste se irradia con una fuente de rayos X con el fin de obtener el patrón de difracción. Este experimento inicialmente permite obtener información sobre el arreglo cristalino [25]. Sin embargo, para llegar a obtener la densidad electrónica (medida de cuántos electrones hay por unidad de volumen en una región específica) que originó dicho patrón, y que se traduce en el modelo estructural de la proteína, es necesario conocer las fases dentro del cristal, las cuales no son posibles de obtener a partir de los datos experimentales. Este es el segundo cuello de botella en la cristalografía y es denominado el problema de las fases [27–29].

Existen diferentes métodos para determinar estas fases, y la elección del método depende de la información disponible sobre la proteína, su secuencia de aminoácidos y los requisitos experimentales específicos [28,29]. En general, las fases se pueden obtener de dos formas: 1) usando las coordenadas de una proteína con una secuencia similar, en un proceso llamado reemplazo molecular (RM), o 2) encontrando la posición de los átomos pesados dentro de la proteína a través del reemplazo isomorfo (*Single isomorphous replacement (SIR)*, *Multiple isomorphous replacement (MIR)*), dispersión anómala (*Single anomalous dispersion (SAD)*, *Multiple anomalous dispersion (MAD)*) o la combinación de ambos métodos (*Single isomorphous replacement with anomalous scattering (SIRAS)*, *Multiple isomorphous replacement with anomalous scattering (MIRAS)*) [30]. El lograr calcular las fases permite finalmente obtener un modelo de la estructura tridimensional de la proteína dentro del cristal [31].

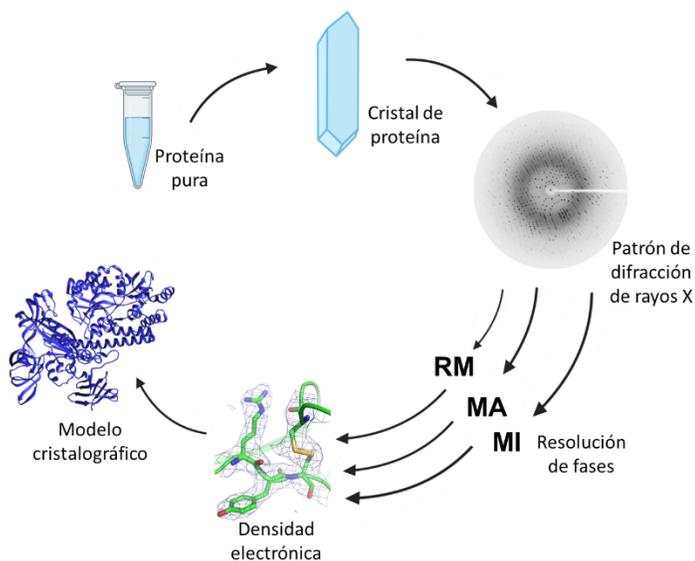


Figura 5.2-2: Esquema general del proceso de resolución de la estructura de proteínas por cristalografía. RM: reemplazo molecular, MA: métodos anómalos, MI: métodos isomorfos.

5.3 Proteínas de fusión en cristalografía de proteínas

La cristalografía de proteínas es un campo en el que el proceso de purificación de proteínas es crucial, por lo que el uso de PFs antes de los ensayos de cristalización es muy beneficioso [9–11]. Debido a que la cristalogénesis requiere grandes cantidades de proteína de alta pureza, como se resaltó en apartados anteriores, las PFs resultan una herramienta útil para aumentar la expresión y alcanzar los niveles necesarios de pureza en pocos pasos [12,24,32]. Sin embargo, en la cristalografía de proteínas, también es una práctica común eliminar la etiqueta después de la purificación y antes de continuar con el proceso de cristalización [7,24].

El impacto de dejar la etiqueta en la estructura de la proteína durante la cristalización es controvertido. Como se mencionó previamente, algunos investigadores afirman que puede afectar la estructura cristalográfica o incluso la capacidad de cristalización de la proteína de interés; mientras que, otros argumentan que la influencia es insignificante o nula [12]. Aquellos que afirman que afecta el proceso de cristalización, argumentan que la etiqueta no tiene una estructura tridimensional definida y puede impedir la cristalización de la PF al aumentar la heterogeneidad de la muestra [3,12]. La mayor parte de la evidencia experimental que respalda la última idea está relacionada con el efecto de la presencia de etiquetas de histidina (*HisTag*). La presencia y ubicación de esta etiqueta (C-terminal o N-terminal) puede afectar la función, la estabilidad de las proteínas, y el proceso de cristalización [33–36]; en algunos casos se ha demostrado que la eliminación o el cambio en la posición de *HisTag* promueve la formación de cristales, mejora la resolución y

disminuye la mosaicidad¹ [34,37,38]. Se ha descrito que el efecto generado por *HisTag* puede deberse a la interacción de la etiqueta con la proteína y el sitio activo de la misma; además se ha descrito un aumento en la movilidad, reflejada por mayores valores de B [33,34,39].

Sin embargo, existen varios argumentos que indican que no es necesario, e incluso puede ser beneficioso no eliminar la etiqueta. Una prueba de ello es la existencia de cientos de estructuras cristalográficas en el Protein Data Bank (PDB) [40] en las cuales el *HisTag* está presente en la construcción, pero no es visible en el mapa de densidad electrónica y donde el plegamiento de la proteína de interés no se ve seriamente afectado [39]. Otros autores han demostrado la utilidad de seleccionar etiquetas de mayor tamaño y no eliminarlas para la cristalización, principalmente en proteínas de difícil cristalización [41]. La presencia de la etiqueta aumenta la solubilidad, lo que permite obtener mayores concentraciones que puedan promover la formación del cristal, y en algunos casos se ha demostrado que sirven como matriz para promover la cristalización de proteínas que de otra manera no habían dado lugar a monocristales que permitan determinar su estructura cristalográfica (Figura 5.3-1) [3,12,42,43].

¹ Mosaicidad es una traducción libre del inglés *mosaicity*. Algunos cristalógrafos hispanohablantes utilizan el término *mosaicismo*. Sin embargo, este es un término más utilizado en genética. Corresponde a una medición de la dispersión de las orientaciones del plano cristalino.

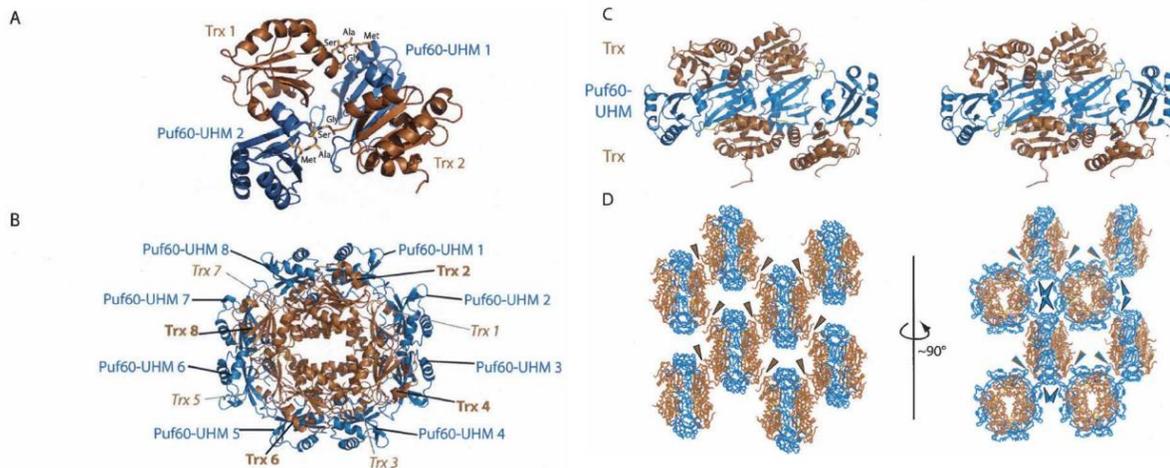


Figura 5.3-1: Estructura de la tiorredoxina fusionada al dominio de homología al factor U2AF perteneciente al factor de transcripción y *splicing Poly(U)* (TRX-Puf60(UHM)) y su empaquetamiento en la red cristalina (código PDB: 3DXB). Figura tomada de Corsini y colaboradores (2008) [32]. Esta estructura es un ejemplo de cómo la etiqueta puede servir como matriz para cristalización. Los dominios TRX se muestran en marrón; los dominios Puf60(UHM) se muestran en azul. El enlazador GSAM se muestra en naranja. A: Estructura de un dímero de TRX-GSAM-Puf60(UHM). B: La unidad asimétrica del cristal. C: Representación estereoscópica de la unidad asimétrica con un giro de 90° alrededor del eje x con respecto a la orientación que se muestra en B, se puede observar cómo los TRX en la capa superior no interactúan con los TRX en la capa inferior. D: Empaquetamiento del cristal desde dos perspectivas. En el panel izquierdo, los contactos TRX:TRX están resaltados flechas marrones. En el panel derecho, los contactos Puf60(UHM):Puf60(UHM) y TRX:Puf60(UHM) están resaltados con flechas azul y azul/marrón, respectivamente.

Adicionalmente, existe evidencia de que el cambio estructural generado por la presencia de etiquetas más grandes que *HisTag* es insignificante en la mayoría de los casos; esta observación se respalda en una comparación de 24 estructuras de PFs con y sin la etiqueta de proteína de unión a maltosa, que mostró que el cambio

en el RMSD (desviación cuadrática media) de los carbonos α entre las estructuras fusionadas y sin fusionar (calculado sobre los aminoácidos superpuestos) tenían un valor promedio de 1 Å (Tabla 5.3-1). El análisis de la superposición de estas estructuras indicó que, al menos para las estructuras analizadas, no existen cambios mayores en el plegamiento general de la proteína de interés, y que la mayor parte de las diferencias se dan en los extremos de las proteínas y con menos frecuencia en las asas presentes entre algunos elementos de la estructura secundaria [42].

Tabla 5.3-1: Superposiciones de carbonos α de la estructura de proteínas fusionadas y no fusionadas con MBP tomado y modificado de Waugh (2016) [42].

PF con MBP	Proteína no fusionada	RMSD	Aminoácidos superpuestos / aminoácidos totales	Porcentaje de aminoácidos superpuestos (%)
4O4B	4O4C	0.74	233/239	97.5
3OAI	1NEU	0.94	113/115	98.3
4B3N	2LM3	0.63	175/202	87.1
2OK2	3ZLJ	0.93	31/31	100
3WAI	3WAJ	0.52	294/368	79.9
2VCQ	4P4H	0.67	93/94	98.9
1NMU	1CK2	1.65	100/104	96.1
3G7V	2KB8	3.97*	25/37	67.6
2NVU	3GZN	0.65	338/429	78.8
3O3U	4LP4	1.05	209/210	99.5
4KV3	4KV2	0.55	90/90	100
4EDQ	3CX2	0.84	105/107	98.1
3H4Z	3UV1	0.95	187/190	98.4
4EXK	2LV4	1.43	103/103	100
4BL8	4KM9	1.30	354/371	95.7

4IFP	3KAT	0.71	82/84	97.6
3VD8	4O7Q	1.03	91/93	97.8
4KEG	4KGO	1.20	160/196	81.6
1R6Z	1VYN	1.71	112/116	96.5
4QVH	4QJK	0.68	224/227	98.6
4WMS	4WMR	1.17	145/149	97.3
4WMT	4WMR	0.82	148/149	99.9
4WMU	4WMR	0.92	148/149	99.9
4EXK	2LV4	1.43	103/103	100

* Ambos polipéptidos adoptan una conformación α -helicoidal, pero las hélices no se alinean.

También se ha propuesto el uso de PFs para facilitar el proceso de determinación de la estructura de las proteínas, principalmente en casos donde no se tiene información del posible plegamiento de la proteína de interés. Esto se logra al expandir las fases de la etiqueta, que posee una estructura conocida, a la proteína desconocida; este proceso consiste en construir a partir de la secuencia y la densidad electrónica, regiones de la sección desconocida y calculando las fases asociadas a las nuevas regiones, hasta completar la totalidad de la estructura [3,32,44]. Este enfoque se ha utilizado en múltiples PFs que fueron determinadas utilizando las fases parciales obtenidas de la etiqueta a través del reemplazo molecular, y expandiendo estas fases a la proteína objetivo usando enfoques manuales o automáticos [8,22].

De acuerdo a la literatura y a las estructuras de PFs depositadas en el PDB, las etiquetas proteicas más comúnmente usadas para cristalografía de proteínas corresponden la Proteína de unión a maltosa (MBP), la Tiorredoxina (TRX), la Glutación transferasa (GST), la Proteína verde fluorescente (GFP) y la Proteína

modificadora similar a la ubiquitina (SUMO) [12,45,46]. Estas proteínas suelen ser elegidas por su facilidad para el proceso de purificación, su alta solubilidad, su color característico y/o su alta estabilidad [42,45]. También existen algunos reportes del uso de proteínas como el Apocitocromo termoestabilizado b562RIL, Proteínas repetidas diseñadas Ankyrin (DARPin) y la Anexina A2 (ANXA2) [47–51].

En el presente trabajo, se eligieron las siguientes cuatro proteínas de interés para abordar el estudio de la utilidad de las proteínas de fusión en la determinación de estructuras cristalográficas, las cuales fueron previamente diseñadas con la presencia de etiquetas.

5.4 Proteínas reguladoras HilD y HilE

Las proteínas HilD y HilE, son reguladores fundamentales en el proceso de invasión de *Salmonella enterica* serovar Typhimurium, bacteria de importancia médica debido a su patogenicidad en humanos y animales, la cual puede causar desde infecciones intestinales localizadas hasta severas, infectando zonas como los huesos, las articulaciones, el tracto urinario y los pulmones [52].

HilD es un regulador transcripcional que induce la expresión del sistema de secreción tipo 3 (SST3-1) al unirse al promotor de *hilA* y contrarrestar la represión mediada por el regulador H-NS [53]. Además, forma un circuito de regulación con RtsA y HilC que amplifica la señal de activación de HilA [54]. Debido a la importancia de HilD para iniciar la infección de *Salmonella*, esta proteína está altamente

controlada. La transcripción se reprime con el regulador H-NS y se auto activa positivamente. HilD se regula positivamente por FliZ y Fur y negativamente por la proteína HilE (Figura 5.4-1) [55,56].

HilD es una proteína de 309 aminoácidos de la familia AraC/Xyls. Las proteínas de esta familia poseen funciones de regulación que se caracterizan estructuralmente por la presencia de un dominio de unión a ADN en el C-terminal de alrededor de 100 aminoácidos, y en general son proteínas de baja solubilidad [57]. Debido a estas características, la determinación de la estructura cristalográfica de las proteínas pertenecientes a esta familia no es fácil. En muchos casos, las únicas estructuras existentes corresponden al dominio de unión a ADN que representa la región más soluble de la proteína [57,58]. Estos rasgos tienen como consecuencia que de las al menos 830 proteínas de esta familia que han sido descritas [58], solo es posible encontrar algunas decenas de estructuras de éstas en el PDB.

HilE es un regulador transcripcional que se encuentra codificado en un gen localizado en una región similar a una isla de patogenicidad en el cromosoma de *Salmonella*, indicando que su origen podría ser el resultado de una transferencia horizontal [59–61]. La función principal que se ha descrito para esta proteína es el regular negativamente la expresión de los genes implicados en el SST3-1 [52]. Esta proteína se une a HilD evitando su unión a ADN y, consecuentemente que este regulador active la transcripción de SST3-1 [61–63]. En general, se sabe que su expresión es regulada positivamente por FimZ y LeuO y es reprimida por lsmR y MIC (Figura 5.4-1) [64,65].

HilE es una proteína de 148 aminoácidos, que comparte aproximadamente un 30% de identidad con algunas proteínas denominadas Hcp (*Haemolysin coregulated proteins*) de bacterias Gram negativas [61]. Las Hcp se caracterizan por ser componentes estructurales del sistema de secreción tipo 6. Actualmente no se cuenta con la estructura de HilE, ya que es una proteína con baja solubilidad, lo que dificulta su estudio [66].

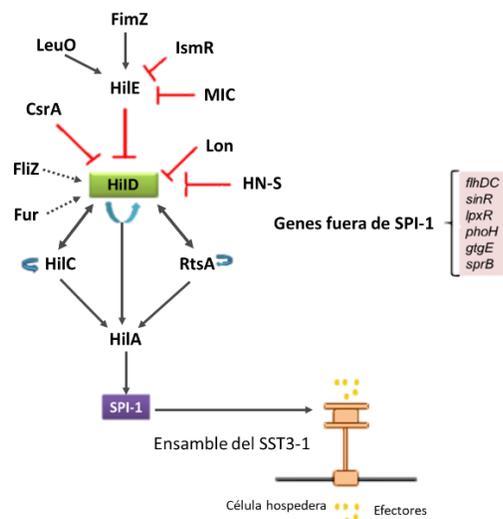


Figura 5.4-1: Cascada de regulación de SST3-1 en *S. enterica* Typhimurium. Figura tomada de Paredes (2019) [66]. Las líneas negras continuas representan la activación de la expresión de los genes, las líneas punteadas representan un mecanismo de activación desconocido aún y las líneas rojas indican represión.

5.5 γ S y β A1 Cristalina humana

Las cristalinas son las proteínas que se encuentran con mayor abundancia en el lente ocular [67]. Estas células se caracterizan por ser alargadas, y no tienen

recambio durante toda la vida del organismo; por lo que, para el correcto funcionamiento del ojo, las proteínas de estas células deben mantener su estructura y solubilidad, asegurando la transparencia y la organización molecular requerida para las propiedades refractivas del cristalino [68].

Estas proteínas se caracterizan por ser altamente solubles, para poder incorporarse a la lente en abundancia suficiente para contribuir a la refractividad requerida, ser extremadamente estables, y ser capaces de tener interacciones supramoleculares con otras cristalinas formando una matriz estable con un alto grado de orden, necesario para la transparencia [67,69,70].

Las cristalinas pueden clasificarse en tres tipos, alfa (α), beta (β) y gamma (γ) cristalinas y el contenido de éstas en el cristalino humano es de aproximadamente 28, 43 y 28 % respectivamente [71]. Las α cristalinas son chaperonas moleculares que previenen interacciones proteicas aberrantes; mientras que las β y γ cristalinas tienen funciones estructurales [68,70,72].

Con el tiempo, las proteínas del cristalino cambian de conformación y acumulan daños estructurales debido a eventos de oxidación, desamidación o escisión, lo que da como resultado interacciones proteína-proteína incorrectas y agregación, que finalmente da lugar a cataratas [70]. Además, existen mutaciones en las cristalinas que pueden provocar anomalías en la organización y la estructura de las fibras del cristalino, lo que genera opacidad o pérdida de transparencia del cristalino, acelerando el proceso de la formación de cataratas [67]. Las estructuras de las β y

γ y cristalinas han sido estudiadas ampliamente, y se sabe que, en general, las β cristalinas son una mezcla heterogénea de dímeros y oligómeros con masas moleculares entre 50 kDa y 200 kDa; mientras que la mayoría de las γ y cristalinas, son proteínas monoméricas con masas moleculares de alrededor de 20 kDa [67,70,71].

Existen seis β cristalinas homólogas con masas moleculares entre 22 y 28 kDa, tres de ellas con puntos isoeléctricos ligeramente más bajos conocidos como βA cristalinas (ácidas), y las tres restantes conocidas como las βB cristalinas (básicas) [67,71]. Dentro de este grupo, las βA3/A1 cristalinas destacan al ser el único caso donde dos polipéptidos se traducen a partir de un único ARNm utilizando sitios de inicio alternativos [73,74]. Para el gen que codifica esta proteína, se ha reportado una mutación denominada Nuc 1, la cual consiste en una inserción de 27 pares de bases en el exón 6 del gen que resulta en la pérdida de un residuo de glicina absolutamente conservado (G178 en βA3) y su reemplazo por 10 nuevos residuos de aminoácidos [67,75]. Esta mutación genera la enfermedad denominada catarata zonular con opacidades suturales [73]. La βA1 cristalina es una proteína constituida por 198 residuos y posee una masa molecular de 23 kDa, de la cual no se ha logrado determinar su estructura experimentalmente hasta la fecha.

Por otra parte, las cristalinas más abundantes en el cristalino humano corresponden a las γC, γD y γS [76]. Existen reportes de múltiples mutaciones en este grupo de cristalinas, las cuales afectan su empaquetamiento en el cristalino y comprometen la transparencia, dando lugar a la opacificación del cristalino. Tales mutaciones,

cuando se heredan, pueden provocar cataratas congénitas [76–78]. La γ S-cristalina es una de las principales cristalinas, encontrándose principalmente en la región externa del cristalino humano, donde su abundancia aumenta con la edad debido a la expresión posnatal [79]. La γ S cristalina es una proteína constituida por 177 residuos, se compone por dos dominios que se conectan por un enlazador de seis residuos y sus estructuras tridimensionales han sido resueltas por cristalografía de rayos X y por resonancia magnética nuclear [77,80,81]. Sin embargo, existe una discrepancia entre sí la forma funcional es el estado monomérico o dímérico, y el papel de los mismos en la formación de cataratas [79]; por lo tanto, cristales con nuevos arreglos cristalográficos y el análisis de las interacciones proteína-proteína generadas en un cristal con una etiqueta, pueden proveer más información que aporte a comprender la discrepancia observada entre las diferentes estructuras determinadas.

6. Justificación

La cristalografía de proteínas es una de las técnicas más exitosas para determinar la estructura tridimensional de proteínas [10]. Sin embargo, existen dos cuellos de botella que pueden dificultar el proceso; la formación de cristales adecuados y la determinación de las fases de las ondas electromagnéticas registradas en el detector [12,28]. Varios autores han resaltado el uso de PFs para superar estos dos cuellos de botella. La presencia de otra proteína en la construcción puede ayudar a promover la cristalogénesis al generar la matriz y los contactos necesarios para la formación del cristal [3,12,32]. Además, si esta proteína posee una estructura conocida, permite extender las fases a la proteína de interés sin necesidad de tener información estructural de la misma [12,42].

En el presente trabajo se evaluó el uso de PFs en cristalografía desde las dos aproximaciones. Para la primera, se determinó la capacidad de tres etiquetas para mejorar el proceso de purificación, aumentar la solubilidad y promover la cristalogénesis en cuatro proteínas de difícil cristalización. Para la segunda aproximación, se realizó la búsqueda de estructuras cristalográficas de PFs depositadas en el PDB con las etiquetas más utilizadas: Proteína de unión a maltosa (MBP), Tiorredoxina (TRX), Glutación transferasa (GST), Proteína verde fluorescente (GFP) y la Proteína modificadora similar a la ubiquitina (SUMO). Entre ellas, se eligieron 116 entradas para evaluar la eficacia del uso de RM parcial de la etiqueta, seguido de construcción automática como método para mejorar las fases y construir la proteína de interés. El objetivo final fue determinar estadísticamente

las características necesarias para expandir con éxito las fases desde la etiqueta hasta la proteína de interés en una PF utilizando un programa de construcción automática (ARP/wARP).

7. Hipótesis

- Las etiquetas TRX, MBP y SUMO fusionados a las proteínas de difícil cristalización H1E, H1D y dos cristalinas humanas (γ S y β A1), ayudarán a mejorar el proceso de la purificación y la solubilidad de estas proteínas, y permitirán generar la matriz cristalina necesaria para formar cristales de proteína.
- La presencia de etiquetas que correspondan a proteínas con estructura tridimensional definida dentro del cristal permitirá calcular las fases de la totalidad de una proteína de fusión sin necesidad de tener información estructural de la proteína fusionada a dichas etiquetas, ya que se pueden extender las fases utilizando programas de construcción automática como ARP/wARP.

8. Objetivos

8.1 Objetivo general

Evaluar el uso de etiquetas como estrategia para cristalizar y determinar estructuras de proteínas.

8.2 Objetivos específicos

- Purificar las proteínas HiID, HiIE, β A1 cristalina y γ S cristalina fusionadas a las etiquetas MBP (unida a HiID), TRX (unida a HiIE) y SUMO (unido a β A1 cristalina y γ S cristalina).
- Evaluar la capacidad de cristalización asistida por etiquetas en las proteínas HiID, HiIE, γ S cristalina y β A1 cristalina.
- Obtener una base de datos actualizada de las estructuras de proteínas de fusión depositadas en el PDB.
- Evaluar la capacidad de reconstruir la proteína de interés a partir de la proteína etiqueta utilizando ARP/wARP, con datos de difracción libres del PDB.
- Determinar las variables que tienen una influencia significativa en lograr o no, reconstruir la proteína de interés a partir de la proteína etiqueta.

9. Cristalización asistida por etiquetas en proteínas de difícil cristalización

9.1 Métodos y materiales

9.1.1 Construcciones de Proteínas de Fusión

Las cepas que expresan las PFs utilizadas en el presente trabajo fueron donadas para este estudio. En el caso de las cepas *Escherichia coli* BL21(DE3) transformada con el plásmido pET32-HilE y *E. coli* TOB10 β transformada con el plásmido PMAL-HilD fueron donadas por el Dr. Víctor Bustamante del IBt-UNAM. Las cepas de *E. coli* BL21(DE3) transformadas con el pE-SUMO CRYGS y pE-SUMO CRYGS fueron donadas por el Dr. Carlos Amero del CIQ-UAEM. Todas las construcciones fueron diseñadas previo a este estudio con la presencia de las etiquetas tiorredoxina (TRX), Proteína de unión a maltosa (MBP por sus siglas en inglés *maltose binding protein*) y la proteína SUMO (por sus siglas en inglés *small ubiquitin-like modifier*); ninguna de estas construcciones fueron diseñadas originalmente para cristalización asistida, por lo que la elección de la etiqueta se basó en sus características para mejorar la expresión, purificación y/o mejorar su solubilidad. Para el caso de las construcciones TRX-HilE y MBP-HilD, se conocía previamente que la presencia de la etiqueta no afecta la actividad [66]. En el caso de las SUMO- γ S y la SUMO- β A1 cristalinas, han sido estudiadas sin la presencia de la etiqueta. A continuación, se muestran las construcciones de las PFs. En letras azules se muestra la secuencia de la etiqueta respectiva, en letras verdes la secuencia de la proteína de interés, en

letras rojas la etiqueta de histidinas de estar presente y, en caso de presentarlo, subrayado en cian la secuencia del enlazador.

TRX-HiE (Tiorredoxina-HiE)

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTV
AKLNIDQNPQTAPKYGIRGIPTLLLFKNGEVAATKVGALSQGLKEFLDANLA **GSG**
SGMH**HHHHH****HSSGLVPRGSGMKETA****AAKFERQHMDSPDLGTDDDDKA****MDAIYL**
KLDGIEGESRIKGFENQIKLIAYNHNPTKRESGEARGTYIGGLTLTKPVDLATPGLY
EHYCNKKTVKEGVLTLCRRDKGAMLPFIYTLTNVRISRMSNHGDAEGSATETVD
LVYSHIRWDIPALASKSKTRRPLHRQALWR

MBP-HiD (*Maltose binding protein*-HiD)

MKIEEGKLVWINGDKGYNGLAEVGGKFEKDTGIKVTVEHPDKLEEKFPQVAATG
DGPDIIFWAHDFRGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAV
EALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGG
YAFKYENKDYDIKDVGVNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGE
TAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELA
KEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMP
NIPQMSAFYAVRTAVINAASGRQTVDEALKDAQT **NSSNNNNNNNNNNL****GIEGRIS**
EFGS**MENVTFVSN****SHQRPAADNLQKLK****SLLTNTRQQIKSQTQQVTIKNLYVSSFT**
LVCFRSGKLTISNNHDTIYCDEPGMLVLKKEQVVNVTLEEVNGHMDFDILEIPTQR
LGALYALIPNEQQTKMAVPTEKAQKIFYTPDFPARREVFEHLKTAFSCTKDTSKGC

SNCNNKSCIENEELIPYFLFLTAFLRLPESYEILSSAQITLKERVYNISSSSPSRQ
WKLTD

SUMO- γ S cristalina (*small ubiquitin-like modifier- γ S cristalina*)

MGHHHHHGHSLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPL
RRLMEAFAKRQGKEMDSLRFYDGIQADQAPEDLDMEDNDIEAHREQIGGSK
TGTKITFYEDKNFQGRRYDCDCDCADFHTYLSRCNSIKVEGGTWAVYERPNFAG
YMYILPQGEYPEYQRWMGLNDRLSSCRAVHLPSSGGQYKIQIFEKGFSGQMYET
TEDCPSIMEQFHMREIHSCKVLEGVWIFYELPNYRGRQYLLDKKEYRKPIDWGAA
SPAVQSFRRIVE

SUMO- β A1 cristalina (*small ubiquitin-like modifier- β A1 cristalina*)

MGHHHHHGHSLQDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPL
RRLMEAFAKRQGKEMDSLRFYDGIQADQAPEDLDMEDNDIEAHREQIGGAQ
TNPTPGSLGPWKITIYDQENFQGKRMEFTSSCPNVSERSFDNVRSLKVESGAWI
GYEHTSFCGQQFILERGEYPRWDAWSGSNAYHIERLMSFRPICSANHKESKMTIF
EKENFIGRQWEISDDYPSLQAMGWFNNEVGSMKIQSGAWVCYQYPGYRGYQYI
LECDHHGGDYKHWREWGSHAQTSQIQSIRRIQQ

9.1.2 Expresión de TRX-HilE

La expresión se realizó a partir de la cepa de *Escherichia coli* BL21(DE3) transformada con el plásmido pET32-HilE (*E. coli*- pET32-HilE), que expresa la PF TRX-HilE. Para esto se realizó un cultivo de esta cepa en medio Luria Bertani (LB) adicionado con ampicilina a una concentración final de 2 μ M. El cultivo se dejó

crecer a 37 °C hasta alcanzar una DO_{600nm} de 0.7. Posteriormente, se indujo la expresión de la proteína TRX-HilE añadiendo isopropil- β -D-1-tiogalactopiranosido (IPTG) a una concentración final de 200 μ M al medio de cultivo e incubándolo a 30 °C por 4 horas. La biomasa se colectó por centrifugación durante 25 minutos a 3,091 RCF a 4 °C.

9.1.3 Expresión de MBP-HilD

La expresión se realizó a partir de la cepa de *E. coli* TOB10 β transformada con el plásmido PMAL-HilD (*E. coli* PMAL-HilD) que expresa la PF MBP-HilD. Las bacterias transformadas se cultivaron en medio LB con ampicilina a una concentración final de 2 μ M y glucosa a una concentración final de 0.2 %, hasta alcanzar una DO_{600nm} de 0.7. Posteriormente, se indujo la expresión de la proteína MBP-HilD añadiendo IPTG a una concentración final de 200 μ M al medio de cultivo e incubando por 4 horas a 37 °C. La biomasa se colectó por centrifugación durante 25 minutos a 3,091 RCF a 4 °C. Este procedimiento fue descrito originalmente por Paredes [66] con modificaciones.

9.1.4 Expresión de SUMO- γ S Cristalina y SUMO- β A1 Cristalina humanas

La expresión se realizó a partir de la cepa de *E. coli* BL21(DE3) transformada con el pE-SUMO CRYGS y pE-SUMO CRYGS, que expresan las PFs SUMO- γ S Cristalina y SUMO- β A1 Cristalina respectivamente. Para esto se realizaron cultivos en medio LB adicionado con ampicilina a una concentración de 100 μ g/ml y

cloranfenicol a 30 µg/ml, y se dejaron crecer a 37 °C hasta alcanzar una DO_{600nm} de 0.7. Posteriormente, se indujo la expresión de las respectivas proteínas de fusión añadiendo IPTG a una concentración final de 200 µM al medio de cultivo e incubándolo a 37 °C por 4 horas. La biomasa se colectó por centrifugación durante 25 minutos a 3,091 RCF a 4 °C.

9.1.5 Purificación de TRX-HilE

Para la purificación de la proteína TRX-HilE, se lisaron las células producto de la expresión en la solución amortiguadora HEPES pH 7.0 50mM adicionado con 150 mM de NaCl y 5 mM de imidazol, junto con solución inhibidora de proteasas cOmplete™ (Roche). El lisado se realizó mediante sonicación por 3 minutos en ciclos de 15 segundos de sonicado y 30 segundos de descanso en una cama de hielo. El sonicado fue centrifugado por 25 min a 12,298 RCF para obtener el extracto crudo de proteínas. Después de evaluar la presencia de la proteína en la fracción soluble, mediante SDS-PAGE al 12 %, el extracto se sometió a una cromatografía de afinidad a níquel. Las proteínas retenidas en esta resina se eluyeron con una solución amortiguadora de HEPES pH 7.0 a una concentración de 50 mM adicionado con 150 mM de NaCl y 500 mM de imidazol. Se colectaron 40 mL en un gradiente de 0 a 100% de la solución con 500 mM de imidazol y 20 mL más con 100% de esta solución con 500 mM imidazol, en fracciones de 2 mL.

Se realizó una electroforesis SDS-PAGE al 12 % para evidenciar la presencia de la proteína en las fracciones colectadas. Las fracciones fueron reunidas y concentradas en una unidad para filtración con centrifuga (Amicon Ultra) con

tamaño de filtro de 10 kDa, hasta obtener un volumen final de 2 mL, para posteriormente eliminar el exceso de imidazol mediante una columna de desalación SP-10 (Sephadex 25) haciendo cambios escalonados de 100 en 100 desde 400 mM hasta 0 mM de imidazol, esto con el fin de evitar la precipitación de la proteína. En cada paso se colectaron fracciones de 0.5 mL y aquellos que presentaban absorción UV 280 nm medida en un Nanodrop, fueron colectadas y llevadas a un volumen de 2 mL y sometidas nuevamente a la columna Sephadex 25 hasta llegar a la concentración de 0 mM de imidazol. Las fracciones finales que presentaban señal de absorción UV 280 nm se reunieron y se confirmó la presencia y pureza de la proteína mediante un SDS-PAGE 12 %. La solución con la proteína se concentró en una unidad para filtración con centrifuga (Amicon Ultra) con tamaño de filtro de 10 kDa y la concentración final fue estimada mediante el método de Bradford.

9.1.6 Purificación de MBP-HiID

Para purificar la proteína MBP-HiID, se lisaron las células en la solución amortiguadora Tris pH 7.5 200 mM, 200 mM de NaCl, 1 mM de ácido etilendiaminotetraacético (EDTA) y 0.7 mg/mL de β mercaptoetanol, durante 4 minutos de sonicación, con 15 segundos de sonificado y 30 segundos de descanso en una cama de hielo. El sonificado fue centrifugado por 25 min a 12,298 RCF para obtener el extracto crudo de proteínas. Después de evaluar la presencia de la proteína en la fracción soluble, el extracto se sometió a cromatografía de afinidad a amilosa. Las proteínas retenidas en esta resina se eluyeron con la misma solución amortiguadora descrita previamente, adicionada con maltosa a una concentración

final de 10mM y se colectaron 40 mL en un gradiente de 0 a 100% de la solución con maltosa, en fracciones de 2 mL cada una.

Se realizó un SDS-PAGE al 12 % para evidenciar la presencia de la proteína en las fracciones colectadas. Estas se reunieron y concentraron utilizando una unidad para filtración con centrifuga (Amicon Ultra) con tamaño de filtro de 30 kDa. La proteína concentrada se sometió a una cromatografía de exclusión molecular en una columna Superdex 200 prep. grade 26/60, colectando 80 fracciones de 5 mL utilizando como solución amortiguadora Tris pH 7.5 a una concentración de 200 mM, adicionado con 200 mM de NaCl y 1 mM de EDTA. Las fracciones eluidas se analizaron mediante SDS-PAGE al 12 % y las fracciones correspondientes a una masa molecular de 78 kDa, fueron reunidas y concentradas utilizando una unidad para filtración con centrifuga (Amicon Ultra) con tamaño de filtro de 30 kDa. La concentración final fue estimada mediante la absorción UV 280 nm en un equipo NanoDrop.

9.1.7 Purificación de SUMO- γ S cristalina humana

Para la purificación de esta proteína de fusión, se lisaron las células en una solución amortiguadora Tris 50 mM, imidazol 10 mM, pH 8.0, durante 5 minutos de sonicación, que constaban de 30 segundos de sonicatedo y 30 segundos de descanso en una cama de hielo, por 10 minutos. El sonicatedo se centrifugó por 25 min a 12,298 RCF para obtener el extracto total de proteínas. Después de evaluar la presencia de la proteína en la fracción soluble mediante SDS-PAGE 12%, el extracto se sometió a una cromatografía de afinidad a níquel. Las proteínas retenidas en esta

resina se eluyeron con la solución amortiguadora Tris 50 mM, pH 8.0 adicionado con 500 mM de imidazol. Se colectaron 90 mL en un gradiente de 0 a 100% de la solución con 500 mM de imidazol en fracciones de 3 mL.

Se realizó un SDS-PAGE al 12 % para evidenciar la presencia de la proteína en las fracciones colectadas. Estas se reunieron y concentraron utilizando una unidad para filtración con centrifuga (Amicon Ultra) con tamaño de filtro de 10 kDa. La proteína concentrada se sometió a una diálisis en buffer Tris 50 mM pH 8.0, adicionado con 150 mM de NaCl. Posteriormente se realizó una cromatografía de exclusión molecular en una columna Superdex 75 prep. grade 26/60, colectando 50 fracciones de 3 mL y utilizando como solución amortiguadora Tris 50 mM pH 8.0, adicionado con 150 mM de NaCl. Las fracciones eluidas se analizaron mediante SDS-PAGE 12 % y las fracciones con masa molecular de 43 kDa, correspondientes a la SUMO- γ S cristalina humana, se reunieron y concentraron en una unidad para filtración con centrifuga (Amicon Ultra) con tamaño de filtro de 10 kDa. La concentración final de la proteína se estimó mediante la absorción UV 280 nm en un equipo NanoDrop.

9.1.8 Purificación de SUMO- β A1 cristalina humana

Para la purificación de la PF SUMO- β A1 cristalina humana, se lisaron las células en una solución amortiguadora Tris 50 mM, imidazol 10 mM, pH 8.0, durante 5 minutos de sonicación, que constaban de 30 segundos de sonificado y 30 segundos de descanso en una cama de hielo, por un tiempo total de 10 minutos. El sonificado se

centrifugó por 25 min a 12,298 RCF para obtener el extracto total de proteínas. Posteriormente, se evaluó la presencia de la proteína en la fracción soluble mediante SDS-PAGE 12 %, y el extracto se sometió a una cromatografía de afinidad a níquel. Las proteínas retenidas en esta resina se eluyeron con la solución amortiguadora Tris 50 mM, pH 8.0 adicionado con 500 mM de imidazol. Se colectaron 90 mL en un gradiente de 0 a 100% de la solución con 500 mM de imidazol en fracciones de 3 mL. Se realizó un SDS-PAGE al 12 % para evidenciar la presencia de la proteína en las fracciones colectadas.

Las fracciones donde se encontraba la proteína fueron reunidas con el fin de realizar una cromatografía de interacción hidrofóbica y obtener una muestra con mayor pureza. Para esto, inicialmente se agregó a la muestra de proteína $(\text{NH}_4)_2\text{SO}_4$ (sulfato de amonio) de forma gradual, en cama de hielo y agitación hasta llegar a una concentración de 1.5 M. Esta muestra se aplicó a una columna con resina butil sefarosa y la proteína fue eluida con un gradiente descendiente de $(\text{NH}_4)_2\text{SO}_4$ de 1.5 a 0 M en 100 mL, colectando fracciones de 2 mL para identificar la presencia de la proteína mediante SDS-PAGE 12 %.

Posteriormente, las fracciones que presentaban la proteína fueron dializadas para eliminar el exceso de $(\text{NH}_4)_2\text{SO}_4$. Se realizó una cromatografía de exclusión molecular en una columna Superdex 75 prep. 26/60, colectando 50 fracciones de 3 mL y utilizando como solución amortiguadora Tris 50 mM pH 8.0, adicionado con 150 mM de NaCl. Las fracciones eluidas se analizaron mediante SDS-PAGE al 12 % y las fracciones correspondientes a una masa molecular de 40 kDa,

correspondientes a la proteína SUMO- β A1 cristalina humana, se reunieron y concentraron en una unidad para filtración con centrifuga (Amicon Ultra) con tamaño de filtro de 10 kDa. La concentración se estimó mediante la absorción UV 280 nm en un equipo NanoDrop.

9.1.9 Cristalización de las Proteínas de Fusión

Con el fin de encontrar condiciones donde se promoviera la formación de cristales debida a la presencia de la etiqueta, a partir de las proteínas con una pureza superior al 90% y concentradas, se pusieron pruebas de cristalización por el método de *microbatch* de cada una de las 4 PFs purificadas, utilizando los kits de cristalización *Wizard I, II, III y IV, Index I y II, Crystal Screen I y II, Crystal Screen Cryo, Crystal Screen lite Quick Screen, Natrrix, MembFac y PEG-Ion*, en dos diferentes temperaturas 4 y 18 °C. Las pruebas de cristalización se observaron periódicamente para evidenciar la presencia de cristales. Las condiciones con cristales con características comunes a cristales de proteína se repitieron y se realizaron matrices de optimización a partir de éstas.

9.2 Resultados

9.2.1 Expresión y purificación de TRX-HilE

Se expresó la PF TRX-HilE de 34 kDa en un cultivo de 100 mL de cepa de *E. coli* BL21(DE3) transformada con el vector pET32-HilE (Figura 9.2.1-1). El proceso de purificación de esta proteína se reportó previamente por Paredes [66] ; Sin embargo,

con este protocolo, la mayor parte de la proteína se encontraba en la fracción insoluble (indicada por la flecha azul en la Figura 9.2.1-1) y no se lograban obtener concentraciones suficientes para poner pruebas de cristalización.

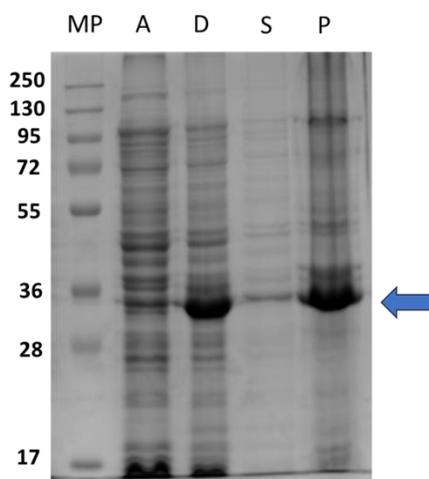


Figura 9.2.1-1: SDS-PAGE 12 % de la expresión y lisis de TRX-HilE. MP: Marcador de peso molecular, A: Perfil de proteínas del cultivo de *E. coli* BL21(DE3)+pET32-HilE antes de inducir la expresión de la PF. D: Perfil de proteínas del cultivo de *E. coli* BL21(DE3)+pET32-HilE después de inducir la expresión de la PF con IPTG. S: Muestra soluble resultante de lisis celular por sonicación. P: Pastilla que corresponde a la fracción insoluble obtenida de la lisis celular por sonicación. Masa molecular esperada para TRX-HilE de 34 kDa indicada por la flecha azul.

Por dicha razón, se realizó un proceso de optimización que permitió obtener una mayor cantidad de proteína en la fracción soluble y en mayores concentraciones. Para esto se evaluó la solubilidad de la proteína posterior al proceso de sonicado en un intervalo de pH 4 a 8 (Amortiguadores: citrato de sodio 50 mM pH 4 y 5, MES 50 mM pH 6 y HEPES 50 mM pH 7 y 8). Se pudo evidenciar una mayor cantidad de

proteína soluble en los pH 6, 7 y 8, de los que se eligió el pH de 7.0 por ser en el que se presentó una mayor cantidad de proteína soluble. Posteriormente, se ensayó si la adición de NaCl a concentraciones de 50, 100 y 150 mM, mejoraba la solubilidad de la proteína; se encontró que la adición de esta última concentración mejoraba la solubilidad de la PF (Figura 9.2.1-2).

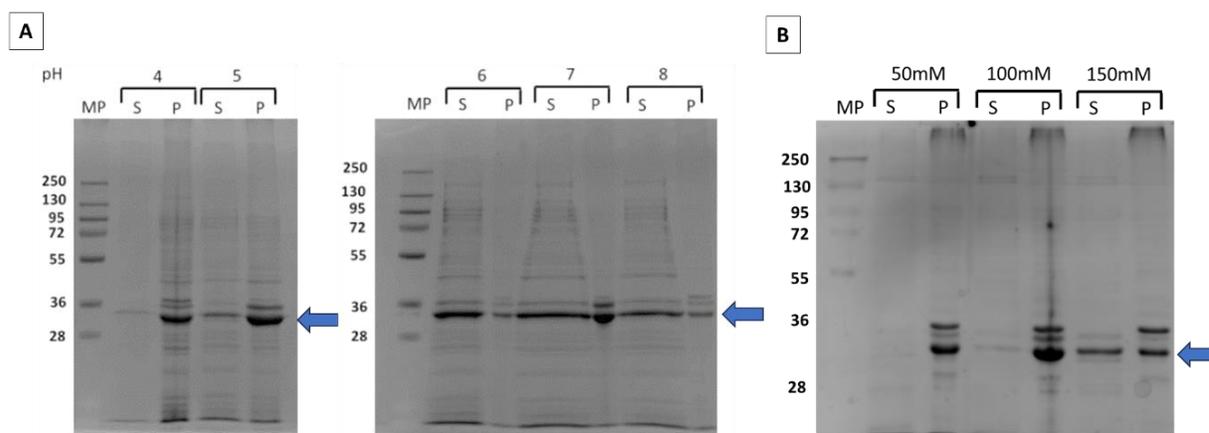


Figura 9.2.1-2: SDS-PAGE 12 % de la evaluación de la solubilidad de TRX-HilE posterior a la lisis celular en diferentes valores de pH (Panel A) y diferentes concentraciones de NaCl (Panel B). MP: Marcador de peso molecular, S: Muestra soluble resultante de lisis celular por sonicación. P: Pastilla que corresponde a la fracción insoluble obtenida de la lisis celular por sonicación. Masa molecular esperada para TRX-HilE de 34 kDa indicada por la flecha azul.

El nuevo protocolo desarrollado consistió en una separación inicial mediante cromatografía de afinidad a níquel, utilizando la solución amortiguadora HEPES 50 mM pH 7.0, adicionada con NaCl a una concentración final de 150mM e imidazol a una concentración final de 5 mM, eluyendo la proteína con un gradiente de 5 a 500 mM de imidazol. Fue posible obtener la proteína en un volumen de elución de

alrededor de 24 mL, obteniendo una PF con una pureza aproximada del 70 % (Figura 10.1.1-3).

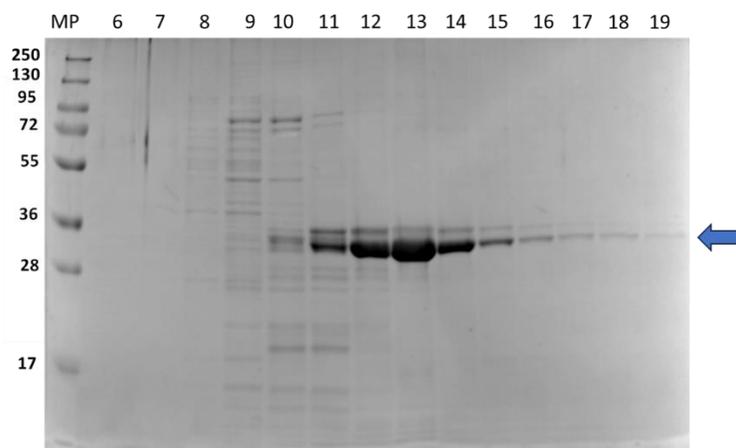


Figura 9.2.1-3: SDS-PAGE 12 % de las fracciones obtenidas mediante separación por cromatografía de afinidad a níquel del extracto total de proteína de 100 mL de cultivo *E. coli* BL21(DE3)+pET32-HiIE. MP: Marcador de peso molecular, 6-19: Fracciones obtenidas en la cromatografía de afinidad a níquel. Masa molecular esperada para TRX-HiIE de 34 kDa indicada por la flecha azul.

Después de reunir y concentrar las fracciones que contenían la proteína, ésta se llevó a una columna de desalado SP-10 (Sephadex 25), con el fin de eliminar el exceso de imidazol proveniente de la cromatografía de afinidad a níquel, y de obtener una muestra con mayor pureza. Para evitar la precipitación de la proteína debido al cambio brusco en la concentración de imidazol, se realizaron cambios escalonados de 100 en 100 desde 400 mM hasta 0 mM de imidazol. Con este protocolo fue posible obtener concentraciones estimadas por el método de Bradford de hasta 15 mg/mL, mucho mayores a las obtenidas con el protocolo anterior, donde estas no superaban concentraciones de 3 mg/mL, y con una pureza mayor al 90 % (Figura 9.2.1-4).

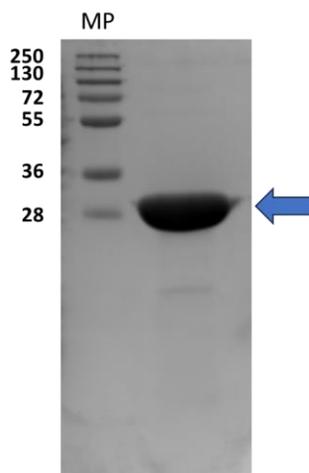


Figura 9.2.1-4: SDS-PAGE 12 % de la proteína final utilizada para pruebas de cristalización. MP: Marcador de peso molecular. Masa molecular esperada para TRX-HiIE de 34 kDa indicada por la flecha azul.

9.2.2 Pruebas de cristalización de TRX-HiIE

Con la proteína TRX-HiIE a una concentración de 12 mg/mL, se realizaron pruebas de cristalización por el método de *microbatch* de los kits comerciales *Wizard I, II, III y IV, Index I y II, Crystal Screen I y II, Crystal Screen Cryo, Crystal Screen Lite, Quick Screen, Natrix, MembFac, PEG-Ion y pH Screen* a 4 y 18 °C, con las muestras previo a retirar el imidazol y posterior a este proceso (por duplicado). Adicionalmente, durante este proyecto se probaron dos diferentes matrices y réplicas de condiciones que en su momento parecían ser prometedoras. Estas pruebas significaron más de 5,000 pruebas de cristalización evaluadas. En ninguna de las pruebas fue posible obtener cristales de la proteína TRX-HiIE, por lo que no se logró una cristalización asistida por la presencia de TRX.

9.2.3 Expresión y purificación de MBP-HiID

Se expresó la proteína de fusión MBP-HiID de 78 kDa en un cultivo de 1 L de la cepa de *E. coli* TOB10 β transformada con el vector PMAL-HiID (Figura 9.2.3-1). El proceso de purificación de esta proteína se realizó como fue reportado previamente por Paredes [66].

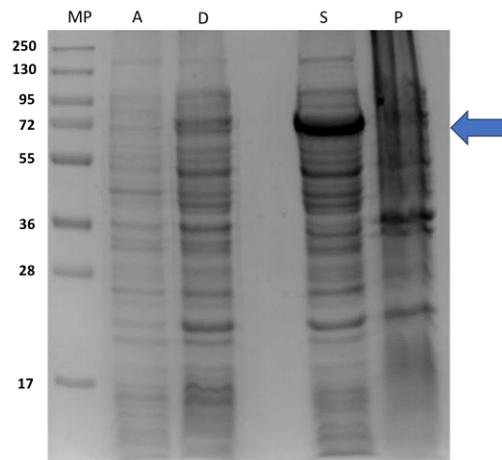


Figura 9.2.3-1: SDS-PAGE 12 % de la expresión y lisis de MBP-HiID. MP: Marcador de peso molecular, A: Perfil de proteínas del cultivo de *E. coli* TOB10 β +PMAL-HiID antes de inducir la expresión de la PF. D: Perfil de proteínas del cultivo de *E. coli* TOB10 β +PMAL-HiID después de inducir la expresión de la PF con IPTG. S: Muestra soluble resultante de lisis celular por sonicación. P: Pastilla que corresponde a la fracción insoluble obtenida de la lisis celular por sonicación. Masa molecular esperada para MBP-HiID de 78 kDa indicada por la flecha azul.

El protocolo de purificación consistió en una separación inicial mediante cromatografía de afinidad a amilosa como se describió en la sección de métodos y materiales. La proteína eluyó en un volumen alrededor de 20 mL del gradiente de amilosa correspondiente a las fracciones 10 a 15 (Figura 9.2.3-2 A).

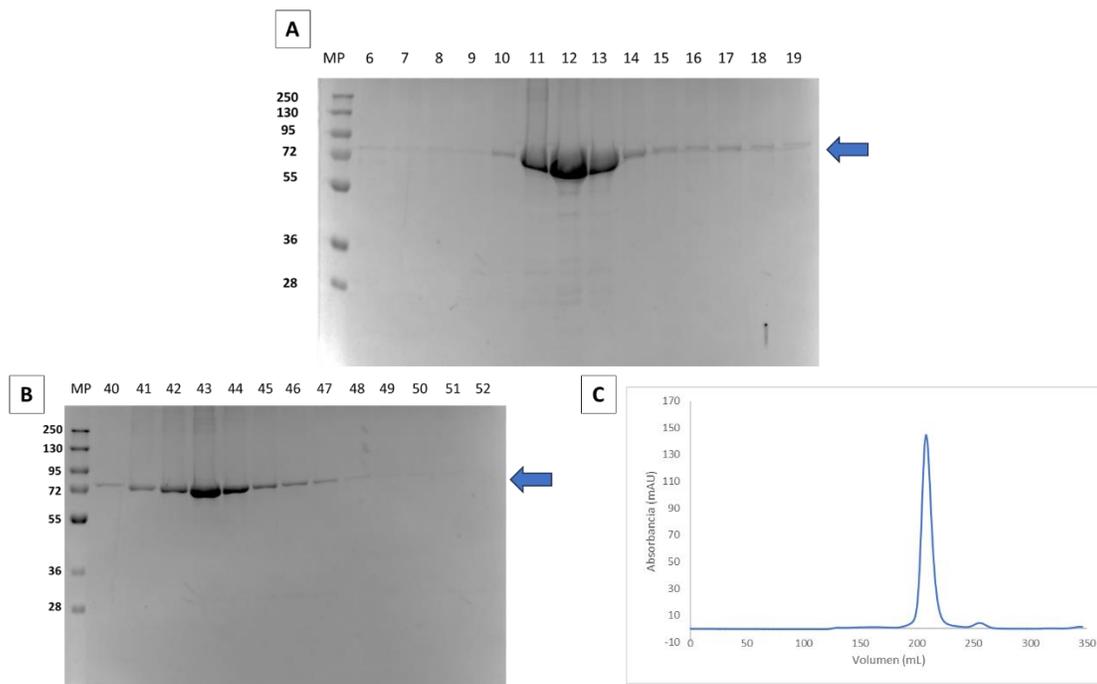


Figura 9.2.3-2: Pasos de purificación de MBP-HilD a partir de 1 L de cultivo. En A se puede observar el SDS-PAGE 12 % de las fracciones obtenidas de la cromatografía de afinidad a amilosa. En B se observa el SDS-PAGE 12 % de las fracciones obtenidas de la cromatografía de exclusión molecular. En C se observa el cromatograma obtenido en la cromatografía de exclusión molecular. MP: Marcador de peso molecular, 6-19: Fracciones obtenidas en la cromatografía de afinidad a amilosa, 40-52: Fracciones obtenidas en la cromatografía de exclusión molecular. Masa molecular esperada para MBP-HilD de 78 kDa indicada por la flecha azul.

Las fracciones donde se encontraba la proteína fueron reunidas y dializadas para someterlas a cromatografía de exclusión molecular en una columna Superdex 200. Se observó que la proteína eluye en un volumen aproximado de 200 mL y que es una muestra homogénea con un solo estado oligomérico (Figura 9.2.3-2 B y C). Las fracciones que presentaban la proteína fueron nuevamente reunidas y dializadas

para eliminar el NaCl. Se logró obtener la proteína MBP-HiID con una pureza entre el 80 y 90 % (Figura 9.2.3-3).

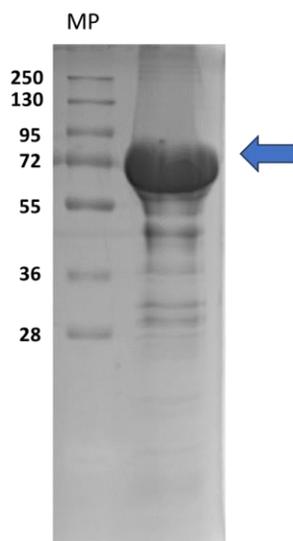


Figura 9.2.3-3: SDS-PAGE 12 % de la MBP- HiID final utilizada para pruebas de cristalización. MP: Marcador de peso molecular. Masa molecular esperada para MBP-HiID de 78 kDa indicada por la flecha azul.

9.2.4 Pruebas de cristalización de MBP-HiID

A partir de la proteína a una concentración de 20 mg/mL se prepararon más de 4,000 pruebas de cristalización por el método de *microbatch* de los kits comerciales *Wizard I, II, III y IV, Index I y II, Crystal Screen I y II, Crystal Screen Cryo, Crystal Screen Lite, Quick Screen, Natrix, MembFac, PEG-Ion, y pH Screen* a 18 °C y a 4°C por duplicado. Además, se realizaron matrices y réplicas de condiciones que, en su momento, parecieron ser prometedoras. En ninguna de las pruebas fue posible obtener cristales de la proteína MBP-HiID, por lo que no se logró una cristalización asistida por la presencia de MBP.

9.2.5 Expresión y purificación de SUMO- γ S cristalina

Se expresó la proteína de fusión SUMO- γ S cristalina humana de 43 kDa en un cultivo de 500 mL de la cepa de *E. coli* BL21(DE3) transformada con el vector pE-SUMO CRYGS (Figura 9.2.5-1). Para su purificación el extracto total de proteínas de la biomasa colectada y sonicada, se sometió a una cromatografía de afinidad a níquel como se describió en la sección de materiales y métodos.

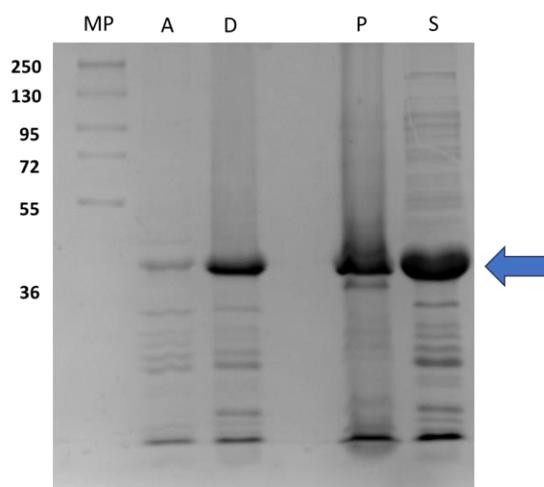


Figura 9.2.5-1: SDS-PAGE 12 % de la expresión y lisis de SUMO- γ S cristalina. MP: Marcador de peso molecular, A: Perfil de proteínas del cultivo de *E. coli* BL21(DE3)+pE-SUMO CRYGS γ S antes de inducir la expresión de la PF. D: Perfil de proteínas del cultivo de *E. coli* BL21(DE3)+pE-SUMO CRYGS después de inducir la expresión de la PF con IPTG. S: Muestra soluble resultante de lisis celular por sonicación. P: Pastilla que corresponde a la fracción insoluble obtenida de la lisis celular por sonicación. Masa molecular esperada para SUMO- γ S cristalina de 43 kDa indicada por la flecha azul.

La PF eluyó en un volumen aproximado de 24 mL de gradiente con imidazol correspondientes a las fracciones 6 en adelante (Figura 9.2.5-2 A). Posteriormente, después de reunir, concentrar y dializar las fracciones que en el SDS-PAGE 12%

mostraron la presencia de la proteína, éstas se sometieron a una cromatografía de exclusión molecular. En el cromatograma, fue posible identificar dos señales con volúmenes de elución de 50 y 75 mL aproximadamente (Figura 9.2.5-2 C). Mediante SDS-PAGE se confirmó que ambas correspondían a la SUMO- γ S cristalina, indicando que para esta proteína se obtienen dos estados oligoméricos diferentes (Figura 9.2.5-2 B y C).

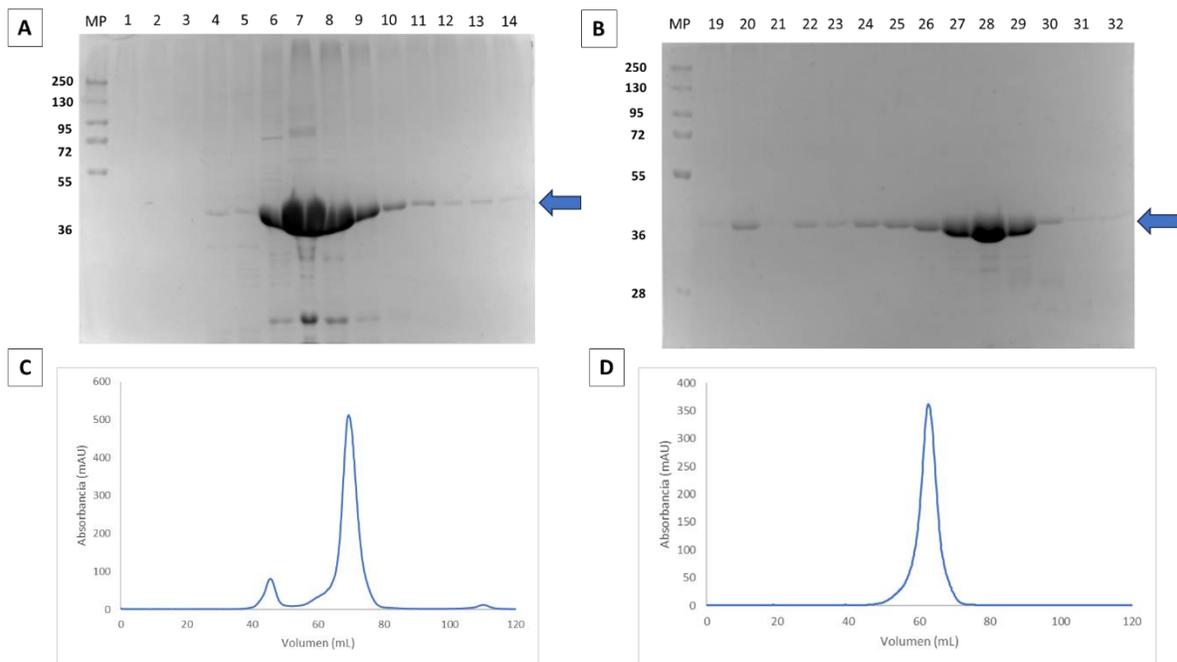


Figura 9.2.5-2: Pasos de purificación de SUMO- γ S cristalina a partir de 500 mL de cultivo. En A se puede observar el SDS-PAGE 12 % de las fracciones obtenidas de la cromatografía de afinidad a níquel. En B se observa el SDS-PAGE 12 % de las fracciones obtenidas de la cromatografía de exclusión molecular. En C se observa el cromatograma obtenido en la cromatografía de exclusión molecular. En D se observa el cromatograma obtenido en la cromatografía de exclusión molecular 24 horas después de la primera exclusión molecular. MP: Marcador de peso molecular, 1-14: Fracciones obtenidas en la cromatografía de afinidad a níquel, 19-32: Fracciones obtenidas en la cromatografía de exclusión molecular. Masa molecular esperada para SUMO- γ S cristalina de 43 kDa indicada por la flecha azul.

Debido a la cantidad de proteína obtenida, se decidió trabajar con aquella que, según el volumen de elución en la columna utilizada, correspondiera al monómero. Con el fin de confirmar si esta proteína alcanza un equilibrio de estos dos estados oligoméricos conforme pasa el tiempo, o si estos eran resultados del proceso de expresión y purificación, se decidió tomar las fracciones correspondientes al monómero, y 24 horas después someterlo a una nueva exclusión molecular. Se encontró que se conserva el estado monomérico al observar una única señal en el cromatograma (Figura 9.2.5-2 D). La proteína obtenida se encontraba entonces homogénea, en forma de monómero y con una pureza mayor al 90 % (Figura 9.2.5-3).

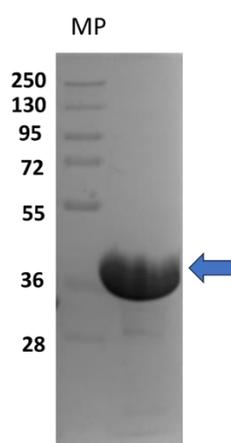


Figura 9.2.5-3: SDS-PAGE 12 % de la SUMO- γ S cristalina final utilizada para pruebas de cristalización. MP: Marcador de peso molecular. Masa molecular esperada para SUMO- γ S cristalina de 43 kDa indicada por la flecha azul.

9.2.6 Pruebas de cristalización de SUMO- γ S cristalina

A partir de la proteína concentrada a 20 mg/mL, se prepararon más de 4,000 pruebas de cristalización por el método de *microbatch* de los kits comerciales

Wizard I, II, III y IV, Index I y II, Crystal Screen I y II, Crystal Screen Cryo, Quick Screen, Crystal Screen Lite, Natrrix, MembFac, PEG-Ion y pH Screen a 18 °C y a 4°C por duplicado. Además, se realizaron matrices y réplicas de condiciones que parecieron ser prometedoras. En ninguna de las pruebas fue posible obtener cristales de la PF, por lo que no se logró una cristalización asistida por la presencia de SUMO.

9.2.7 Expresión y purificación de SUMO-βA1 cristalina

Se expresó la proteína de fusión SUMO-βA1 cristalina humana de 40 kDa en un cultivo de 500 mL de la cepa de *E. coli* BL21(DE3) transformada con el vector pE-SUMO CRYGS (Figura 9.2.7-1).

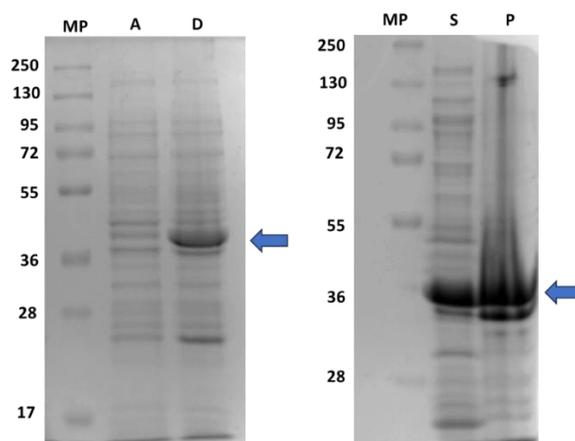


Figura 9.2.7-1: SDS-PAGE 12 % de la expresión y lisis de SUMO-βA1 cristalina. MP: Marcador de peso molecular, A: Perfil de proteínas del cultivo de *E. coli* BL21(DE3)+pE-SUMO CRYGS βA1 antes de inducir la expresión de la PF. D: Perfil de proteínas del cultivo de *E. coli* BL21(DE3)+pE-SUMO CRYGS βA1 después de inducir la expresión de la PF con IPTG. S: Muestra soluble resultante de lisis celular por sonicación. P: Pastilla que corresponde a la fracción insoluble obtenida de la lisis celular por sonicación. Masa molecular esperada para SUMO- βA1 cristalina de 40 kDa indicada por la flecha azul.

El primer paso de purificación de esta proteína consistió en una cromatografía de afinidad a níquel como se describió previamente. La proteína eluyó en un volumen de 27 mL de gradiente de 10 a 500 mM de imidazol (Figura 9.2.7-2 A). Este proceso dio como resultado proteína con una cantidad considerable de contaminantes, por lo que las fracciones que contenían la proteína se sometieron a una cromatografía de interacción hidrofóbica. La adición de 1.5 M $(\text{NH}_4)_2\text{SO}_4$ y la separación con una butil sefarosa permitió obtener una proteína más pura. En este proceso, la proteína eluyó en un volumen de 50 mL de un gradiente de 1.5 a 0M de $(\text{NH}_4)_2\text{SO}_4$ (Figura 9.2.7-2 B).

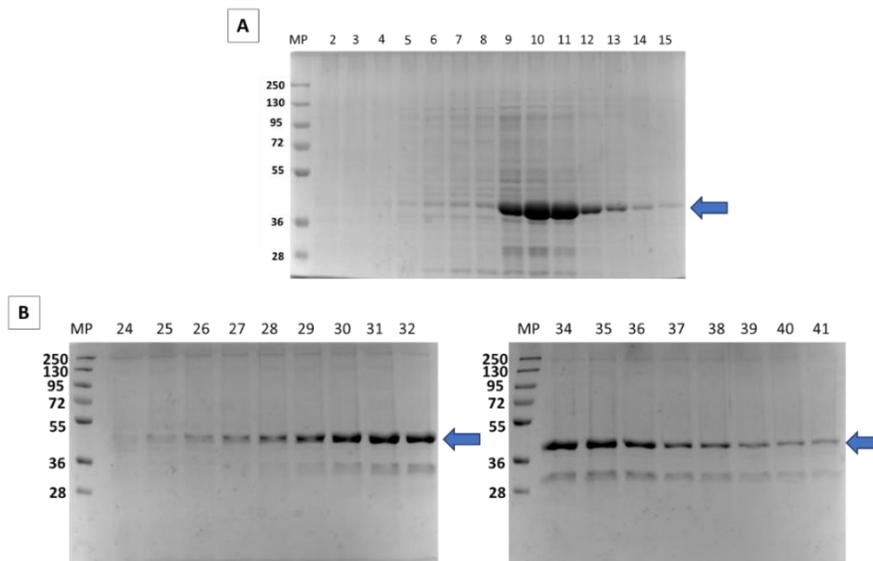


Figura 9.2.7-2: Pasos de purificación iniciales de SUMO-βA1 cristalina a partir de 500 mL de cultivo. En A se puede observar el SDS-PAGE 12 % de las fracciones obtenidas de la cromatografía de afinidad a níquel. En B se observa el SDS-PAGE 12 % de las fracciones obtenidas de la cromatografía de interacciones hidrofóbicas. MP: Marcador de peso molecular, 2-15: Fracciones obtenidas en la cromatografía de afinidad a níquel, 24-41: Fracciones obtenidas en la cromatografía de interacciones hidrofóbicas. Masa molecular esperada para SUMO- βA1 cristalina de 40 kDa indicada por la flecha azul.

Las fracciones obtenidas en las que se presentaba la SUMO-βA1 cristalina fueron reunidas y dializadas para eliminar el exceso de $(\text{NH}_4)_2\text{SO}_4$, y se sometieron a una cromatografía de exclusión molecular. De las fracciones obtenidas, se tomaron aquellas que correspondían a la señal con mayor intensidad que según el volumen de elución correspondía al monómero (Figura 9.2.7-3).

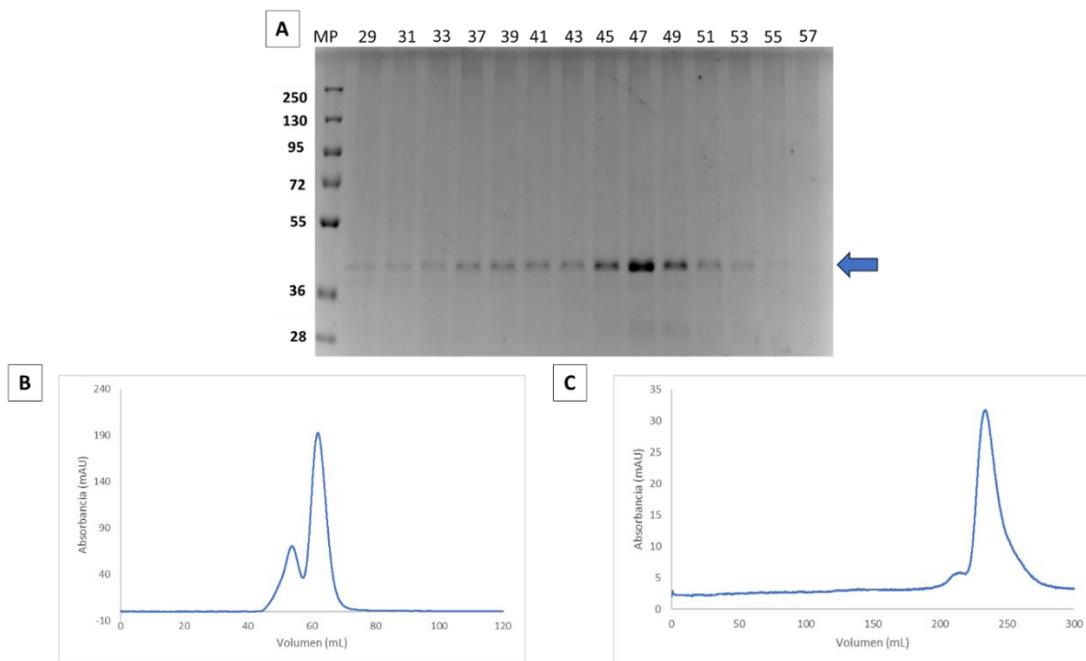


Figura 9.2.7-3: Resultados del proceso de cromatografía de exclusión molecular de la SUMO-βA1 cristalina. En A se puede observar el SDS-PAGE 12 % de las fracciones obtenidas de la cromatografía de exclusión molecular. En B se observa el cromatograma obtenido en la cromatografía de exclusión molecular en una columna sup75. En C se observa el cromatograma obtenido en la cromatografía de exclusión molecular 24 horas después de la primera exclusión molecular en una columna sup200. MP: Marcador de peso molecular, 29-57: Fracciones obtenidas en la cromatografía de exclusión molecular. Masa molecular esperada para SUMO- βA1 cristalina de 40 kDa indicada por la flecha azul.

Debido a que, en ensayos previos a implementar la cromatografía de interacciones hidrofóbicas, se presentaban señales muy intensas de diferentes estados oligoméricos de la PF, se decidió evaluar la estabilidad de la proteína obtenida. Esto se realizó sometiendo las fracciones con la proteína a una segunda cromatografía de exclusión molecular 24 horas después, en una columna Sephadex con un corte de 200 KDa con el fin de separar mejor las señales. En esta se observó que la mayor parte corresponde a la misma señal (Figura 9.2.7-3 C). La proteína obtenida de este proceso corresponde a una proteína con una pureza alrededor del 80 % (Figura 9.2.7-4).

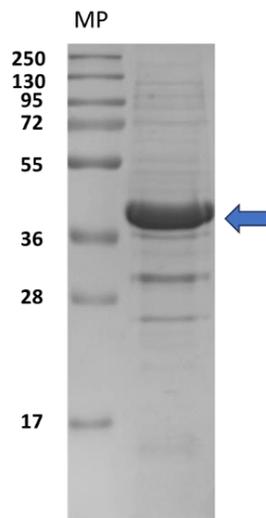


Figura 9.2.7-4: SDS-PAGE 12 % de la SUMO- β A1 cristalina final utilizada para pruebas de cristalización. MP: Marcador de peso molecular. Masa molecular esperada para SUMO- β A1 cristalina de 40 kDa indicada por la flecha azul.

9.2.8 Pruebas de cristalización de SUMO-βA1 cristalina

A partir de la proteína concentrada a 20 mg/mL, se prepararon más de 4,000 pruebas de cristalización por el método de *microbatch* por duplicado de los kits comerciales *Wizard I, II, III y IV, Index I y II, Crystal Screen I y II, Crystal Screen Cryo, Crystal Screen Lite, Quick Screen, Natrix, MembFac, PEG-Ion y pH Screen* a 18 °C y a 4°C. Además, se realizaron matrices y réplicas de condiciones que, en su momento, parecieron ser prometedoras. En ninguna de las pruebas puestas fue posible obtener cristales de SUMO-βA1 cristalina; por lo que, no se logró una cristalización asistida por la presencia de SUMO.

9.3 Discusión

Cada vez es más común encontrar publicaciones donde se enfatizan los beneficios del uso de las PFs para obtener estructuras cristalográficas [12,32,41,82]. En la mayoría de estas, la más destacada es la capacidad que poseen las etiquetas de promover los contactos cristalográficos necesarios para la formación de un cristal, actuando como 'chaperonas' de cristalización en proteínas de difícil cristalización [32,42].

En este trabajo, se evaluó esta capacidad en algunas proteínas de difícil cristalización, en las cuales su sistema de expresión había sido diseñado previamente para generar una proteína de fusión. Como se describió en los resultados, la presencia de la etiqueta si generó un efecto en el proceso de purificación. En todos los casos fue posible obtener cantidad de proteína suficiente

y purezas aceptables para iniciar procesos de cristalogénesis, usando no más de tres pasos de purificación. Además, fue posible obtener las proteínas en altas concentraciones para las pruebas de cristalización, que en los casos particulares de HilE y HilD, no sería posible alcanzar sin la presencia de la etiqueta, debido a la poca solubilidad que presentan estos reguladores transcripcionales. Además, para estas dos proteínas, también existe evidencia que la presencia de las etiquetas no impide su actividad de regulación [66]. Es así como en las proteínas estudiadas, la presencia de la etiqueta permitió llegar más fácilmente a obtener las características necesarias para probar miles de condiciones de cristalización.

Sin embargo, a pesar de las miles de condiciones de cristalización evaluadas, no fue posible obtener cristales de proteína de ninguna de las PFs probadas en esta tesis. Este resultado nos indica que la cristalización asistida con etiquetas no es un método infalible para cristalizar, y que al igual que la cristalización de una proteína sin etiqueta, existen muchos factores que pueden influir (ya sea promoviendo o limitando) en el proceso de formación de la red cristalina.

El primer factor para tomar en cuenta es la etiqueta que se debe utilizar. Así como cualquier proteína a cristalizar, éste posee características diferentes que generan que las condiciones de cristalización difieran entre sí [83] ; no cualquier etiqueta puede promover la cristalización de cualquier proteína. Una prueba indirecta de esto es el número de estructuras cristalográficas depositadas en el PDB para diferentes

etiquetas. En este trabajo se realizó una búsqueda de PFs en el PDB, encontrando que existen 252 estructuras con MBP, 69 con GFP, 38 con SUMO, 13 con GST y 13 con TRX (Ver resultados sección 10.2.1). La diferencia en la cantidad de las estructuras depositadas con las diferentes etiquetas puede deberse a varios factores: que no todas promueven la cristalización, que las características de difícil cristalización de la proteína de interés no son compensadas de igual manera por las diferentes etiquetas, o bien, que algunas etiquetas se han usado más preferentemente que otras para cristalización de proteínas. Es de destacar el aumento en el número de estructuras depositadas con las etiquetas TRX, GST y MBP, con el reporte hecho por Waugh [42] hace 7 años. El número de nuevas estructuras de PFs depositadas fue de 10 más para TRX, siete para más GST y 150 más para MBP. Acorde a lo previamente mencionado, MBP parece ser una de las etiquetas más reportadas para cristalización y por lo tanto podría ser la primera elección de muchos. Sin embargo, esta diferencia notable también podría deberse a que MBP sea una mejor 'chaperona' de cristalización que otras etiquetas, probablemente debido a que promueve mayores niveles de solubilidad que las etiquetas TRX y GST [84]. Sin embargo, esto tampoco es garantía, pues para cada proteína o condiciones específicas, la etiqueta adecuada para cristalización puede variar y es necesario evaluar diferentes opciones, como fue reportado por Ecsédi [50], donde el uso de MBP no promovió la cristalización, a diferencia del uso de la proteína annexina A2 como etiqueta, que si generó cristales.

El segundo factor a considerar para promover la cristalización asistida es la elección del enlazador. Distintos autores han resaltado la importancia de elegir una

secuencia adecuada dependiendo la PF y la aplicación que tendrá [2,3,42,85]. Los enlazadores pueden clasificarse en dos categorías: rígidos o flexibles, y cada uno aporta características diferentes a la PF [2,3]. Los enlazadores rígidos (2 a 5 aminoácidos de longitud) pueden aumentar el área superficial y promover la formación de contactos para formar la red cristalina al reducir la movilidad y, por tanto, la heterogeneidad estructural. Por el contrario, los enlazadores flexibles (de más de 10 aminoácidos de longitud) pueden incrementar la concentración local al promover más movilidad y la interacción entre la proteína etiqueta y la proteína de interés [3]. No existen suficientes evidencias experimentales que demuestren tendencias claras para determinar cuáles son los parámetros a tomar en cuenta para elegir el enlazador. Aunque puede observarse que la mayor parte de los enlazadores de las proteínas de fusión depositadas en el PDB no tiene una longitud mayor a 10-12 residuos, existen múltiples ejemplos con enlazadores largos (más de 20 aminoácidos) que han dado lugar a cristales de proteína (algunos ejemplos son las estructuras con el código PDB: 4O4B, 6DKS, 4JKM, 6LNP, 1B8X). Además, la longitud parece no ser el único factor determinante del éxito de una secuencia enlazadora; existen ejemplos donde ha sido necesario probar diferentes secuencias de aminoácidos para alcanzar la cristalización asistida [3,86]. Lo que sí es claro es que tienen una influencia sobre el proceso de cristalización asistida, además de que pueden ofrecer muchas otras ventajas como mejorar la actividad y expresión [2]. Adicionalmente, es posible agregar otras moléculas o realizar mutaciones que promuevan más fácilmente la cristalización [42,49,87]. En este caso de estudio, las proteínas TRX-HilE y MBP-HilD poseían enlazadores largos y flexibles, y en las SUMO- γ S cristalina y SUMO- β A1 cristalina, la región de unión era corta,

acercándose a las dos aproximaciones propuestas, pero en ningún caso se obtuvieron cristales. Junto con todos los ejemplos descritos en la literatura, y todos los diferentes enlazadores existentes en las PFs depositadas en el PDB, queda claro que no hay una regla general que permita predecir cual será la mejor etiqueta o enlazador para lograr la cristalización asistida.

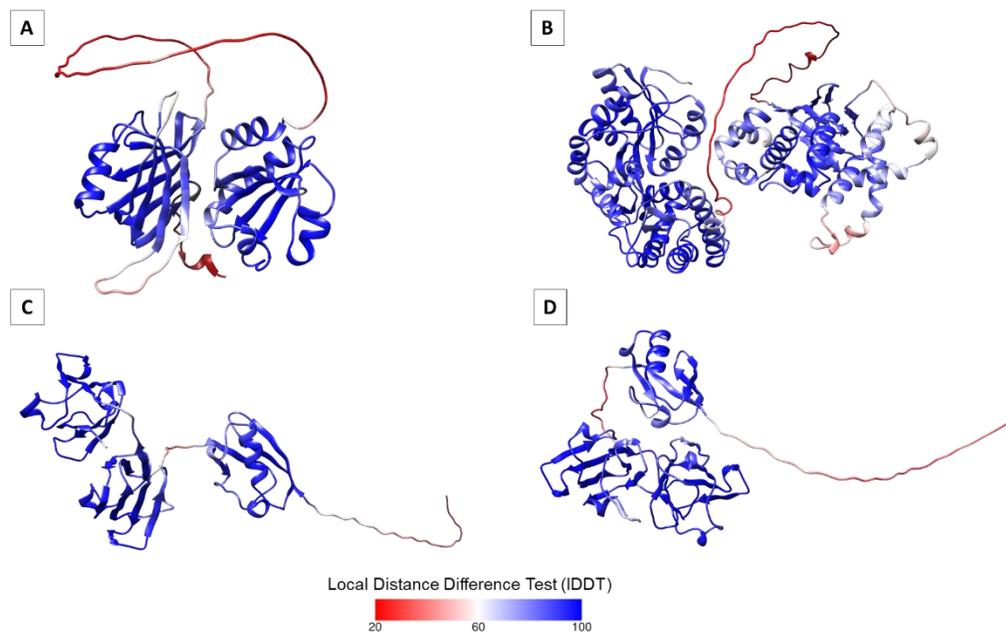


Figura 9.3-1: Predicción de la estructura de las PFs evaluadas en este estudio, realizada con AlphaFold2. Se colorea según el valor de IDDT (*Local Distance Difference Test*), el cual evalúa las diferencias de distancia local de todos los átomos en el modelo, y puede ser tomado como un indicador de confianza de la estructura predicha (rojo = poca confianza; azul = mayor confianza). A: TRX-HiIE, B: MBP-HiID, C: SUMO- γ S cristalina, D: SUMO- β A1 cristalina.

Una predicción de la estructura de las PFs utilizadas, en esta tesis, realizada con Alphafold2 [88] mostró la presencia de regiones sin estructura secundaria definida y con baja confianza en el modelo, en las 4 PFs, principalmente en los enlazadores o en algunas secciones localizadas de las etiquetas, particularmente en el caso de SUMO (Figura 9.3-1). Además, un análisis de predicción de regiones intrínsecamente desordenadas realizado con IUpred3 [89] mostró que estas mismas regiones poseen mayor probabilidad de ser intrínsecamente desordenadas (Figura 9.3-2).

A pesar de que esto podría ser un indicador de que estas proteínas son difíciles de cristalizar, existen evidencias de que no es necesariamente así en todos los casos. Se realizó la predicción de la probabilidad de cristalizar estas PFs con el servidor XtalPRED (Figura 9.3-3) [90], la cual clasifica con dos parámetros, el EP, basado en la longitud, punto isoeléctrico, índice *gravy* (hidrofobicidad e hidrofiliidad promedio), predicción de regiones desordenadas, índice de inestabilidad, estructura secundaria prevista, estructura prevista de regiones *coiled coil* y puntuación de inserción; y el RF, basado en los mismos parámetros anteriores sumado a la predicción de la rugosidad de la superficie, hidrofobicidad, entropía de la cadena lateral de los residuos de la superficie y composición de aminoácidos de la superficie. En ambos casos, un valor mayor indica que son menos prometedores para la formación de cristales. En todas la PFs se obtuvieron valores altos, indicando que podrían no ser favorables a cristalizar.

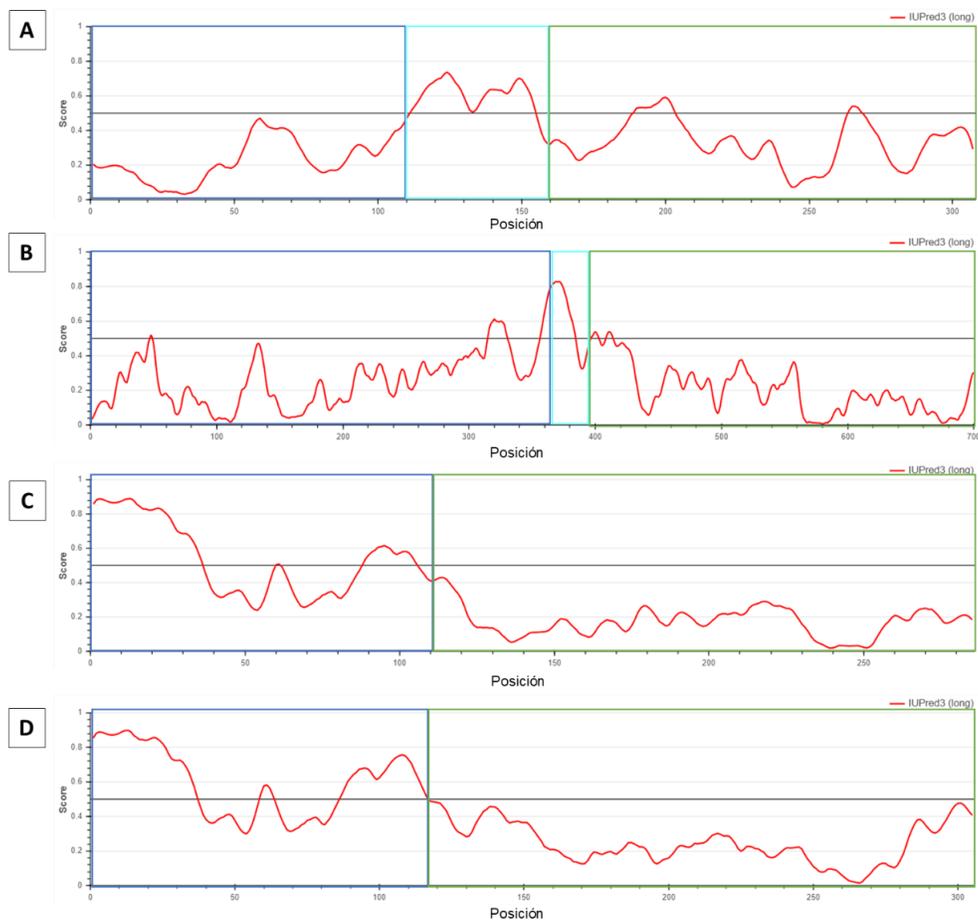


Figura 9.3-2: Predicción de las regiones intrínsecamente desordenadas en las PFs evaluadas en este estudio, realizada con IUPred3. El *score* puede tener valores entre 0 y 1 para cada residuo, correspondiente a la probabilidad de que el residuo dado sea parte de una región desordenada. A: TRX-HilE, B: MBP-HilD, C: SUMO-γS cristalina, D: SUMO-βA1 cristalina. En los recuadros azules se marca la sección de la etiqueta, en cian del enlazador y en verde la proteína de interés.

Sin embargo, este resultado no es algo definitivo que indique que una proteína no se podrá cristalizar. Existen varios ejemplos de proteínas que han sido cristalizadas a pesar de que su probabilidad en este servidor es baja. En este trabajo presentamos el ejemplo de la proteína SUMO, que posee múltiples regiones

predichas como intrínsecamente desordenadas, y que según XtalPRED es poco probable de cristalizar. Sin embargo, existen más de 60 estructuras cristalográficas donde la proteína SUMO estaba presente y en la muchas de ellas las coordenadas de SUMO pudieron ser determinadas. Igualmente está el caso de la PF MBP-EHIP6KA, la cual posee un enlazador con una longitud de 31 residuos y que según la predicción no es muy propensa a formar cristales; a pesar de esto, fue posible cristalizar esta proteína y su estructura se encuentra depositada con el código 4O4B en el PDB.

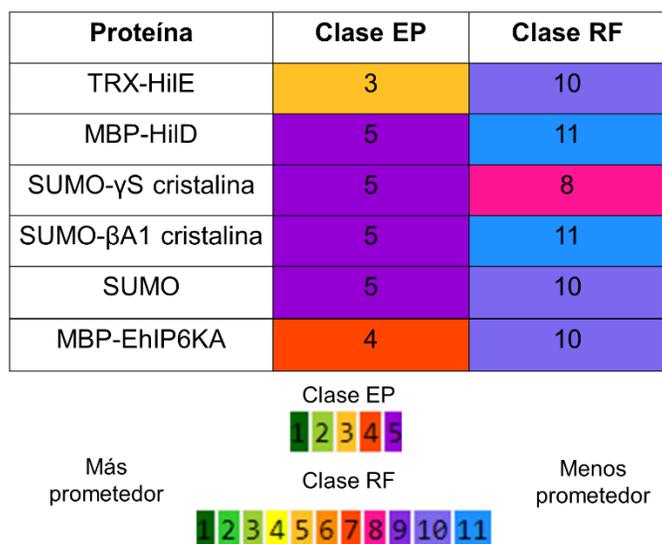


Figura 9.3-3: Predicción de probabilidad de cristalizar, usando el servidor XtalPRED. Las PFs de este estudio, la proteína SUMO y la PF MBP-EHIP6KA. En el inferior se presenta los valores que pueden tomar cada predicción en las clases EP y RF.

En el presente estudio se trabajó con una sola variante de etiqueta y enlazador para cada una de las proteínas elegidas. Los resultados obtenidos, por lo tanto, no indican necesariamente que el uso de PFs para promover la formación de cristales

no sea una aproximación adecuada. Por el contrario, indica que, de querer usarse, deben diseñarse y evaluarse diferentes etiquetas y enlazadores para cada una de las proteínas de difícil cristalización.

10. Determinación estructural mediante la expansión de fases en proteínas de fusión depositadas en el PDB

10.1 Métodos y materiales

10.1.1 Estructuras cristalográficas de PFs depositados en PDB

La búsqueda de PFs en el PDB se realizó mediante un alineamiento de la secuencia de las etiquetas con estructura tridimensional definida más comúnmente utilizadas: MBP, TRX, GST, SUMO y GFP. El alineamiento se realizó contra las estructuras depositadas en el PDB utilizando BLAST [91], y cada resultado se examinó manualmente para confirmar que correspondía con la estructura de una PF resuelta mediante cristalografía de rayos X. Para cada PF se enumeraron el porcentaje que representa la etiqueta en la PF, la resolución y el método de resolución de fases (material suplementario, Tabla 1).

10.1.2 Determinación de la estructura de la proteína de interés

Para determinar la estructura de la proteína de interés mediante la expansión de fases de la proteína etiqueta, se eligieron 50 PFs con MBP y 50 con GFP. En todas las PFs seleccionadas, la etiqueta no constituía más del 77 % de toda la PF. Inicialmente, se seleccionaron para el análisis las 38 estructuras de PFs depositadas que contenían SUMO; sin embargo, una de las entradas fue excluida debido a la ausencia de intensidades en los documentos depositados en el PDB, lo que impedía resolver la estructura por el método propuesto (entrada PDB 3IX6).

Además, 21 estructuras fueron eliminadas debido a que SUMO no estaba presente en las coordenadas ni era visible en el mapa de densidad electrónica (entradas PDB: 6UKM, 6UKU, 6UKV, 6UKW, 6UKX, 6UKZ, 6UL0, 7PWH, 7PWI, 6BR9, 3IX6, 6FMP, 6FMQ, 4DA1, 5VWM, 5M23, 5M25, 7P7I, 6PTH, 6OTE, 6CFP). Finalmente, fueron analizadas 16 PFs con la etiqueta SUMO. El número de PFs que contenían TRX y GST fue insuficiente para realizar un análisis estadístico y, por lo tanto, se excluyeron del análisis (Figura 10.1.2-1 panel A). Dentro de las PFs seleccionadas, varias entradas eran redundantes, ya que correspondían a la misma PF, difiriendo solamente en la presencia de ligandos o mutaciones. Estas PFs redundantes fueron incluidas en el análisis con el fin de evaluar si la proteína de interés genera una influencia en obtener la estructura final.

La aproximación propuesta fue evaluada mediante dos métodos. Para el método 1, se utilizaron las intensidades de cada PF y las coordenadas de la proteína etiqueta, para realizar un remplazo molecular (RM) parcial [92]. A continuación, se utilizó el mapa de densidad electrónica resultante y la secuencia de cada PF para construir los residuos faltantes usando ARP/wARP [93]. El proceso se repitió con un nuevo RM utilizando las coordenadas resultantes de ARP/wARP (después de eliminar las moléculas de agua y los átomos "*dummy*"). Los nuevos factores de estructura del RM se utilizaron para un nuevo ciclo de ARP/wARP. Estos ciclos de RM y construcción automática se repitieron hasta obtener las coordenadas más completas de la sección enlazador-proteína de interés (Figura 10.1.2-1 panel B). El proceso se realizó siguiendo los modos predeterminados de ambos programas y evitando la modificación manual.

El método 2 consistió en el mismo proceso descrito en el método 1, pero usando como coordenadas iniciales para el RM parcial las correspondientes a la estructura de la proteína etiqueta no fusionada depositada en el PDB. Para las PFs con MBP, el RM se realizó con las coordenadas de MBP en la conformación cerrada (entrada PDB 1ANF) y en la conformación abierta (entrada PDB 1OMP). Para GFP se utilizaron las coordenadas de la entrada 1KYS. En todos los casos, las coordenadas elegidas correspondieron a las estructuras de cada proteína (silvestre) en la mejor resolución disponible en el PDB. Para las PFs con SUMO, se emplearon las entradas 1EUV, 2EKE, 3V60, 5D6J y 1L2N. Estas estructuras se seleccionaron debido a la ausencia de depósitos de estructuras cristalográficas individuales de SUMO. En todos los casos, se seleccionaron para la construcción automática los resultados de RM parcial con el mayor valor de *Log Likelihood Gain* (LLG) (Figura 10.1.2-1 panel C).

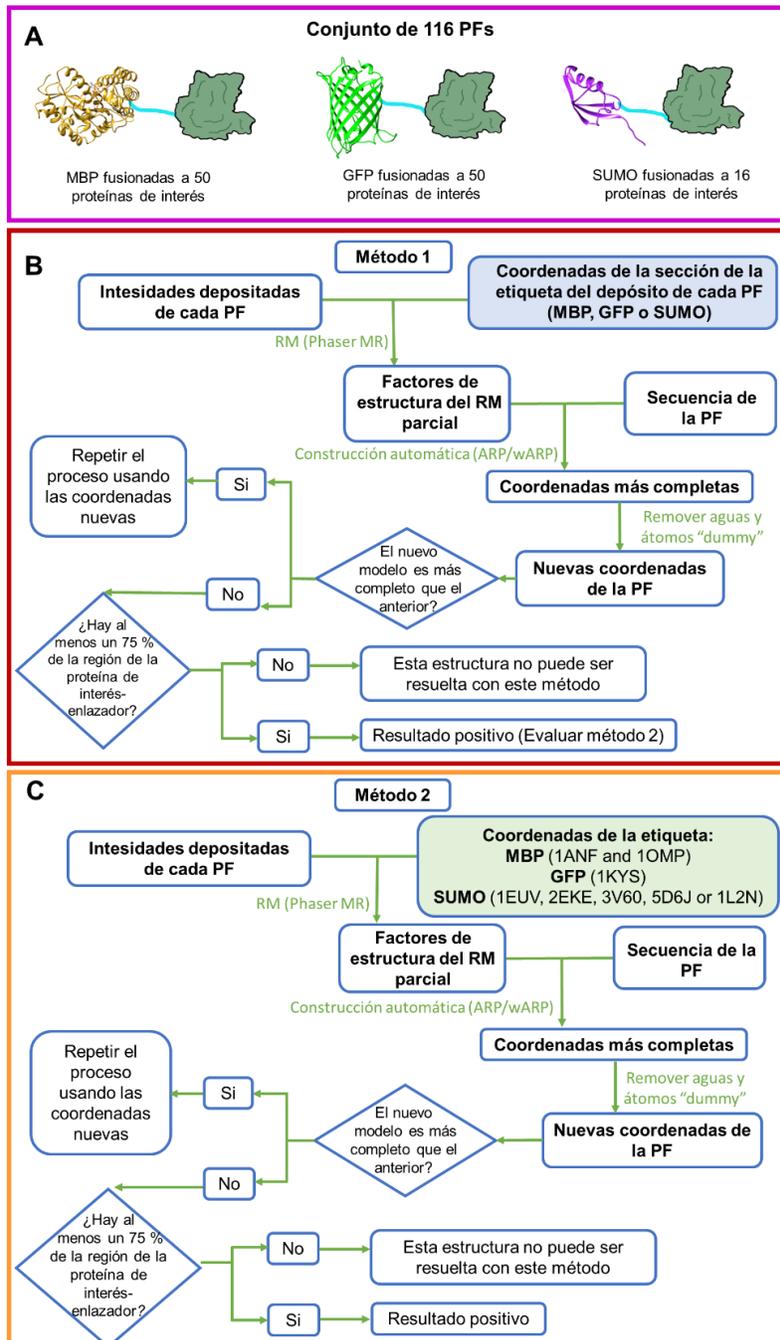


Figura 10.1.2-1: Métodos utilizados para construir la estructura faltante de las PFs seleccionadas. A) Se eligió un conjunto de 116 PFs con las etiquetas MBP, GFP y SUMO. B) Método 1 - Las 116 PFs se construyeron utilizando las coordenadas correspondientes a la sección de la etiqueta de cada depósito. C) Método 2: los resultados positivos obtenidos en el Método 1 se construyeron utilizando las coordenadas de estructuras de las etiquetas no fusionadas disponibles en el PDB.

10.1.3 Análisis estadístico de los resultados obtenidos con ARP/wARP

Los resultados del método 1 fueron analizados para determinar la influencia de algunas variables en la capacidad de obtener las estructuras más completas. Se examinaron variables como resolución, porcentaje que representa la proteína etiqueta en la PF, número de aminoácidos en la unidad asimétrica (UA), porcentaje de solvente, completitud, multiplicidad, I/σ (I), Factor B de Wilson, LLG del RM inicial, longitud del enlazador, tipo de fuente de rayos X, tipo de detector y el grupo espacial.

Los casos en los que fue posible construir el 75 % o más de la sección enlazador-proteína de interés, se consideraron resultados positivos. Los análisis se realizaron inicialmente utilizando correlaciones lineales simples: histogramas para variables cuantitativas y gráficos de barras para variables cualitativas. Además, se realizaron una serie de Análisis de Componentes Principales (PCA). El PCA es un análisis multivariado que permite reducir el número de variables, y determinar cuáles tienen mayor peso y significancia [85]. En el caso de las variables cualitativas, se realizaron cuatro análisis mixtos de componentes principales (PCA mix). Para estos análisis se utilizaron las variables cuantitativas en conjunto con: 1) todas las variables cualitativas, 2) fuente de rayos X, 3) tipo de detector y 4) grupo espacial. Para las variables cuantitativas se ejecutó un PCA. Todos los PCAs se llevaron a cabo utilizando XLSTAT (© Addinsoft).

10.2 Resultados

10.2.1 Estructuras cristalográficas de PFs depositados en PDB

Varios autores han reportado a MBP como la etiqueta más utilizada en las PFs depositadas en el PDB, seguida de GFP, TRX y GST [38, 78]. En 2016, Waugh reportó la presencia de seis estructuras con GST, tres con TRX y 102 con MBP [42]. Cuando se escribió este manuscrito, se encontraron 385 depósitos de PF resueltas por cristalografía de rayos X en el PDB: 252 con MBP, 69 con GFP, 38 con SUMO, 13 con GST y 13 con TRX (material suplementario Tabla 1).

La mayoría de las PFs encontradas reportan el uso de RM parcial (Ver material suplementario Tabla 1), utilizando las coordenadas de la etiqueta para obtener las fases iniciales: representan el 52.2 % de las PFs (201 estructuras en el PDB). Para 156 de estas 201 estructuras, la expansión de fases se logró mediante construcción manual; mientras que solo 45 utilizaron construcción automática en algún punto. Para 167 (43.4 %) de las 385 PFs, se realizó RM utilizando las coordenadas tanto de la etiqueta como de la proteína de interés como modelo inicial. El resto de las PFs (4.4 % del total), utilizaron SAD, MIR o una fusión de estos con RM.

10.2.2 Determinación de la estructura de la proteína de interés

Para determinar las coordenadas de la proteína de interés, partiendo únicamente de las fases iniciales de la proteína etiqueta, se siguió el método 1 (ver materiales y métodos). Se consideró un resultado positivo cuando se construyó al menos el 75 %

de la sección enlazador-proteína de interés. Este umbral se determinó en función del punto en el que se observó un cambio significativo en el porcentaje de resultados construido de todas las PFs evaluadas (Figura 10.2.2-1).

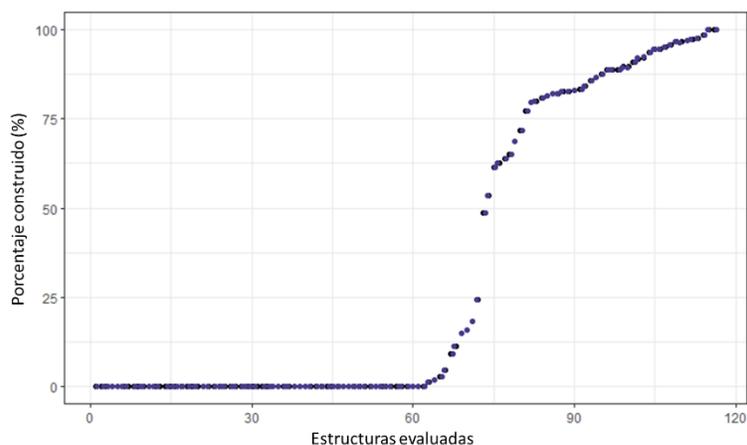


Figura 10.2.2-1: Porcentaje de la sección enlazador-proteína de interés construida a partir de 116 estructuras de PFs en el PDB mediante RM parcial de la etiqueta seguido de construcción automática usando ARP/wARP.

De las 50 estructuras con la etiqueta MBP, fue posible construir 18 (36 % de resultados positivos). Cuando se utilizó GFP para generar las fases iniciales, se construyeron con éxito 14 de 50 (28 % de resultados positivos). Además, cuatro de las 16 PFs con SUMO fueron construidas con éxito (25 % de resultados positivos). En general, el 31 % de las PFs estudiadas se construyeron con éxito (Material suplementario Tabla 2). Con el fin de comparar nuestras respuestas positivas con las coordenadas originalmente depositadas en el PDB, se calcularon los valores de RMSD entre ambos conjuntos de datos. El RMSD promedio de la región de la proteína de interés-enlazador en las 36 estructuras construidas en comparación con

la estructura depositada fue de 0.246 Å, mostrando que el uso del método propuesto en esta tesis produjo coordenadas casi idénticas a las depositadas en el PDB original. Además, considerando que nuestro método no incluyó afinamiento de los resultados de ARP/wARP, es probable que los valores de RMSD obtenidos puedan llegar a ser aún más bajos de aplicarse un afinamiento tradicional (Tabla suplementaria 3). Los resultados del método 1 indicarían que la naturaleza de la etiqueta no tiene una influencia considerable sobre el porcentaje de resultados positivos. Además, se producen resultados positivos y negativos en varios casos en los que la única diferencia es la presencia de mutaciones puntuales o ligandos, indicando que la proteína de interés no tiene una influencia en los resultados. A pesar de comenzar con las coordenadas correctas de la etiqueta, la mayoría de los casos no produjeron un resultado positivo, indicando que existen otros factores además del RM inicial que influyen en el éxito del método evaluado. El uso de las coordenadas finales de la etiqueta para el RM parcial (método 1), nos permitió analizar el impacto de la calidad de los datos, eliminando la influencia negativa de un RM con LLG bajo.

Sin embargo, si un cristalógrafo se enfrenta a la tarea de construir una PF desde cero, las coordenadas iniciales deben derivarse de una estructura de la etiqueta obtenida previamente. Por lo tanto, la eficacia de este enfoque se verificó iniciando el RM parcial con las coordenadas de las estructuras de las etiquetas disponibles en el PDB (Método 2). Esta verificación se realizó solo en los casos en los que se lograron resultados positivos en el método 1. De los 18 casos positivos identificados en las PFs con MBP, 13 de ellos se resolvieron con éxito utilizando las coordenadas

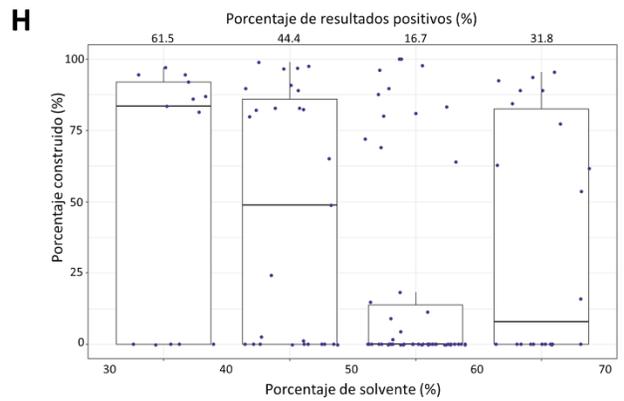
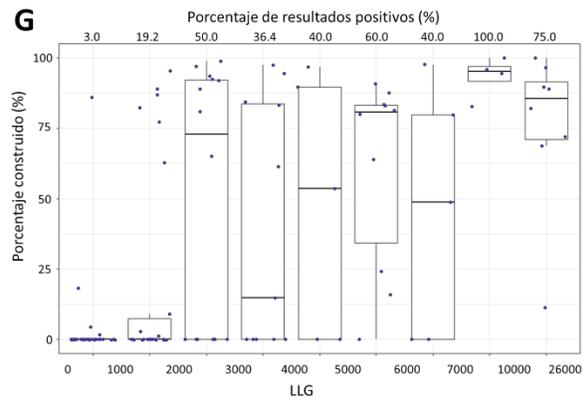
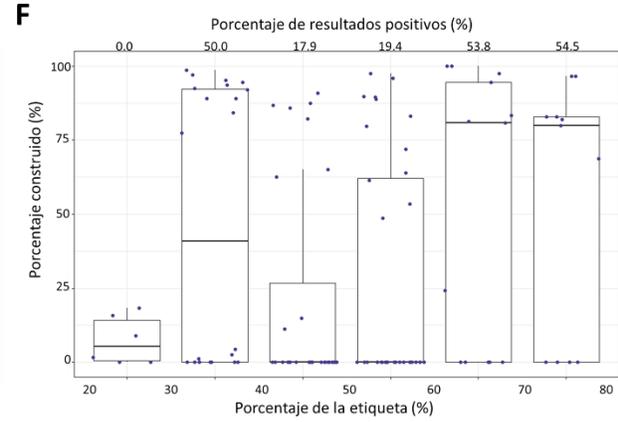
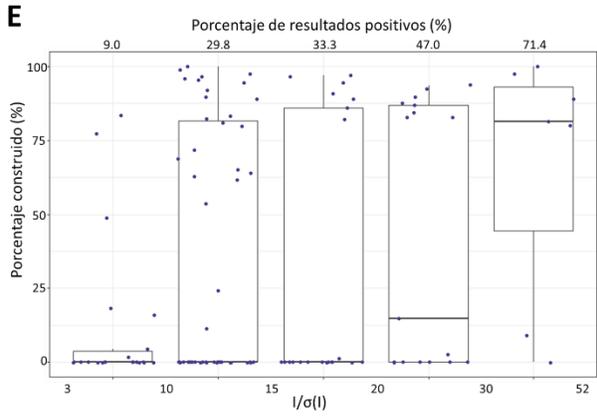
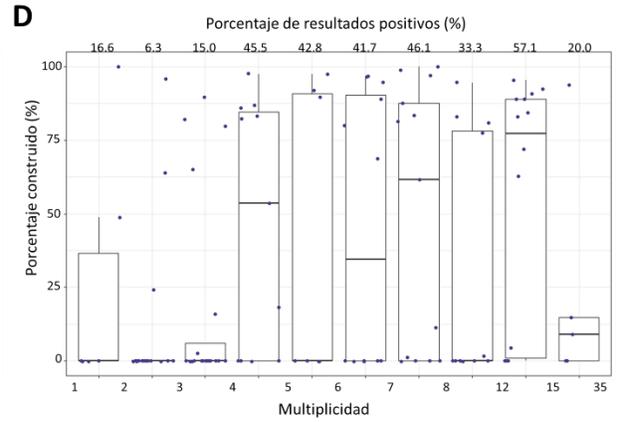
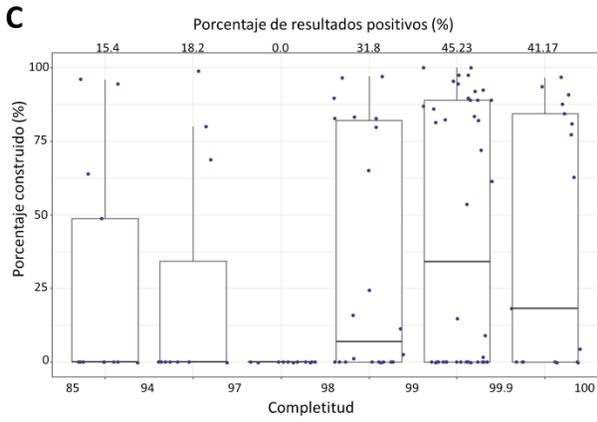
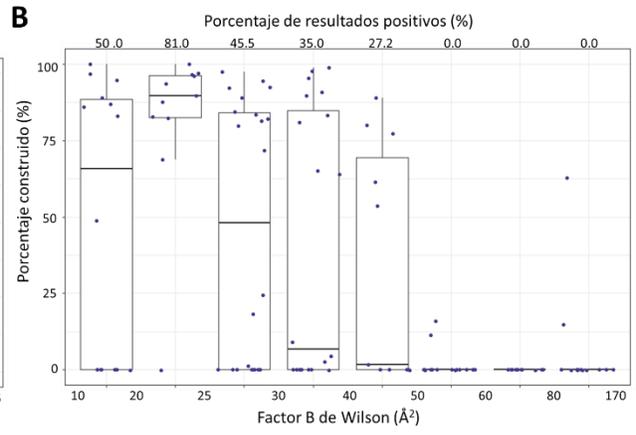
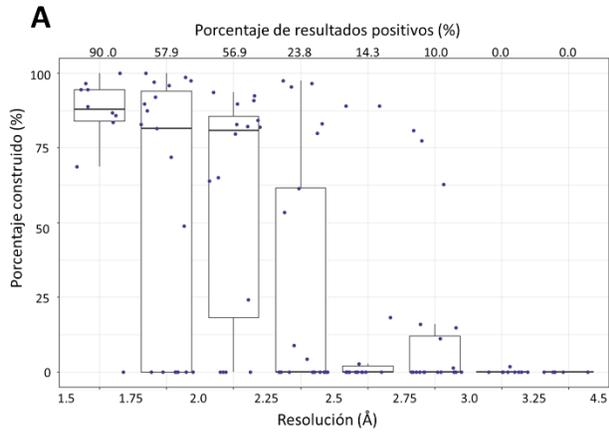
individuales del MBP. Debido a la naturaleza dinámica de MBP y la existencia de dos estados principales, se evaluaron dos modelos iniciales para el RM. En los 13 casos con resultados positivos, las coordenadas empleadas para RM correspondieron al estado cerrado de MBP. En todos estos casos, como se esperaba, el resultado del RM mostró un valor de LLG superior a 1,000 y de TFZ (*translation function Z-score*) superior a 40, lo que indica la importancia de un buen resultado del RM para garantizar el éxito del proceso de construcción automática.

En el caso de las PFs con GFP, las 14 estructuras con resultados positivos en el método 1 fueron resueltas utilizando el método 2. En todos estos casos, los resultados de RM también mostraron valores de LLG sobrepasando los 2,400 y de TFZ mayor a 45. Lamentablemente, cuando se trata de PFs con SUMO, no se lograron resultados positivos independientemente del modelo utilizado para realizar el RM parcial (en todos los casos los valores de LLG fueron inferiores a 100 y los valores de TFZ fueron menores a 10).

10.2.3 Análisis estadístico de los resultados obtenidos con ARP/wARP

Se eligieron diferentes variables correspondientes a propiedades de las proteínas, estadísticas de calidad de los datos, características de los cristales y de colecta de datos, con el fin de analizar su impacto en los resultados obtenidos en la sección anterior. En concreto fueron elegidas: resolución, porcentaje que representa la etiqueta en la PF, número de residuos de aminoácidos en la UA, porcentaje de solvente, completitud, multiplicidad, $I/\sigma(I)$, factor B de Wilson, LLG del RM parcial

inicial, longitud del enlazador, además de la fuente de rayos X, el tipo de detector y el grupo espacial. Una correlación lineal simple reveló tendencias positivas en ciertas variables para obtener un resultado positivo; los casos con una resolución mejor que 2.80 Å y un factor B de Wilson inferior a 50 Å² fueron los más exitosos en la expansión de fases a la proteína de interés (Figura 10.2.3-1: paneles A y B). Otras variables mostraron tendencias positivas discernibles: estructuras con altos valores de completitud, multiplicidad, $I/\sigma(I)$, porcentaje de la etiqueta y LLG en el RM inicial (Figura 10.2.3-1: paneles C, D, E, F y G). Además, porcentajes más bajos de solvente en el cristal, menos residuos en la UA y un enlazador con una longitud media a corta correspondieron a un mayor porcentaje de resultados positivos (Figura 10.2.3-1: paneles H, I y J). Las estructuras obtenidas utilizando un ánodo rotatorio presentaron una mayor proporción de resultados positivos (Figura 3: panel K). Sin embargo, cabe señalar que el número de estructuras colectadas con esta fuente de rayos X es inferior a las obtenidas mediante un sincrotrón en las proteínas estudiadas. No se observaron patrones discernibles en las variables tipo de detector ni grupo espacial (Fig. 3: paneles L y M).



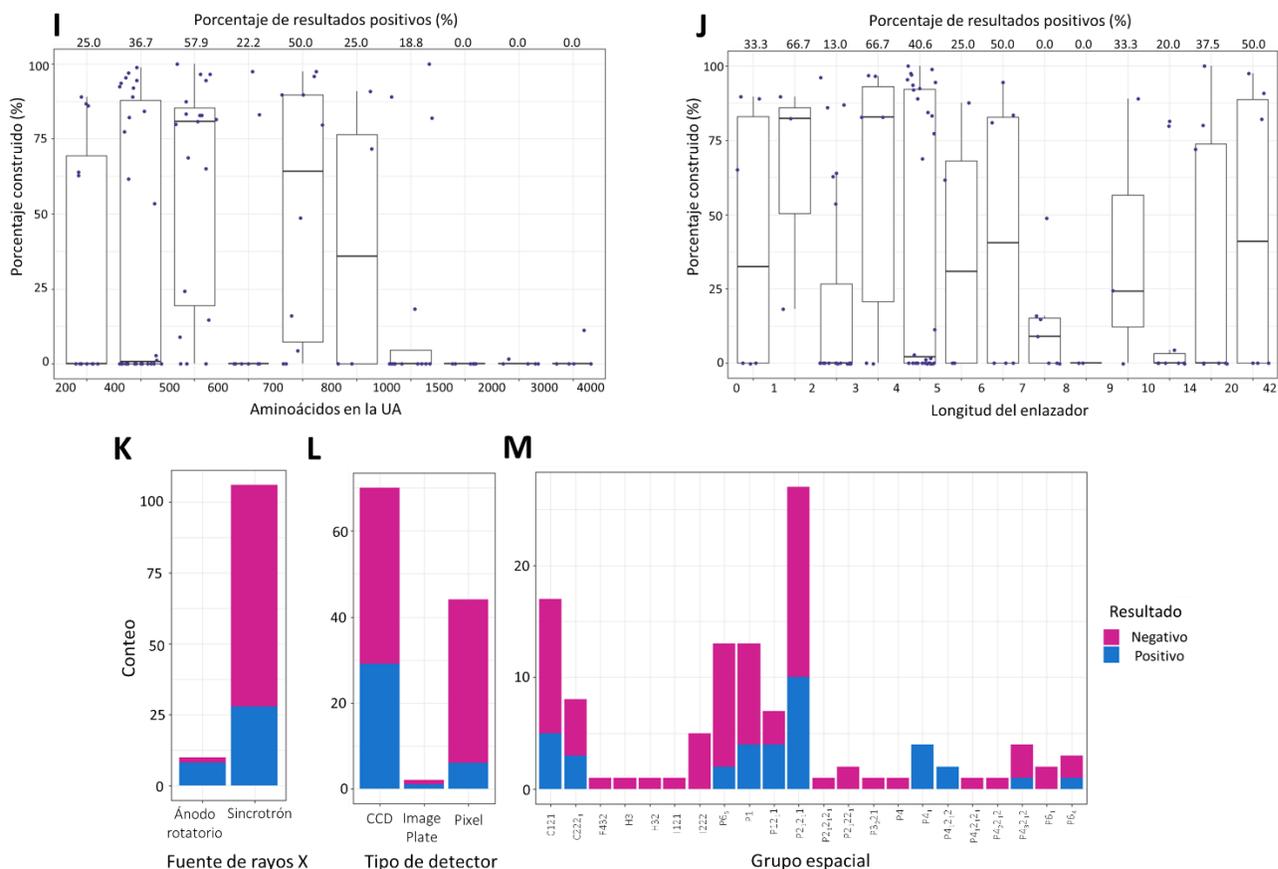


Figura 10.2.3-1: Análisis de correlación lineal simple de los resultados considerando A) Resolución, B) Factor B de Wilson, C) Completitud, D) Multiplicidad, E) l/σ (I), F) Porcentaje de la etiqueta, G) LLG del RM parcial, H) Porcentaje de solvente, I) Aminoácidos en la UA, J) Longitud del enlazador. K) Fuente de rayos X, L) Tipo de detector, y M) Grupo espacial.

Con el fin de validar las tendencias observadas en los análisis lineares simples, además de identificar las variables más significativas en obtener resultados positivos, se realizaron PCAs. Para analizar las variables cualitativas, se realizó un PCA mix que incluía todas las variables cuantitativas y cualitativas (fuente de rayos X, tipo de detector y grupo espacial). Posteriormente, se realizaron una serie de PCAs mix que incluían todas las variables cuantitativas y las variables cualitativas

individualmente. Estos análisis mostraron que era posible explicar un mayor porcentaje de la varianza de los resultados al no incluir las variables cualitativas. Esos hallazgos sugieren que las variables cualitativas evaluadas tienen una influencia mínima en los resultados positivos (Figura 10.2.3-2).

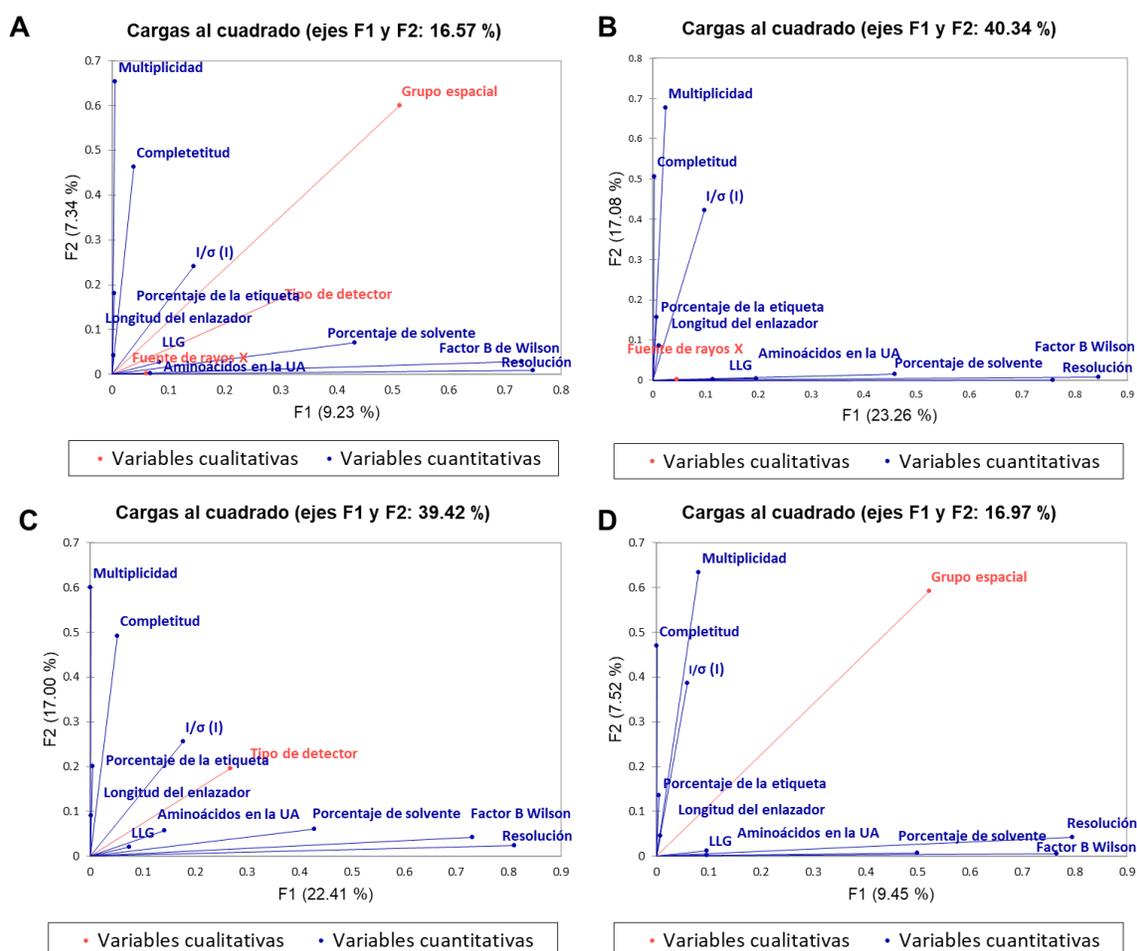


Figura 10.2.3-2: Cargas al cuadrado identificadas en los análisis mixtos de componentes principales de 116 PFs del PDB: A) Análisis que incluye todas las variables cualitativas y cuantitativas, B) Análisis que incluye la fuente de rayos X y las variables cuantitativas, C) Análisis que incluye el tipo de detector y las variables cuantitativas, D) Análisis que incluye el grupo espacial y las variables cuantitativas.

Para determinar la influencia de variables cuantitativas se realizó un PCA. Los tres ejes principales calculados pueden explicar el 63.93 % de la variabilidad. Este análisis mostró las mismas tendencias identificadas en el análisis lineal simple (Figura 10.2.3-3).

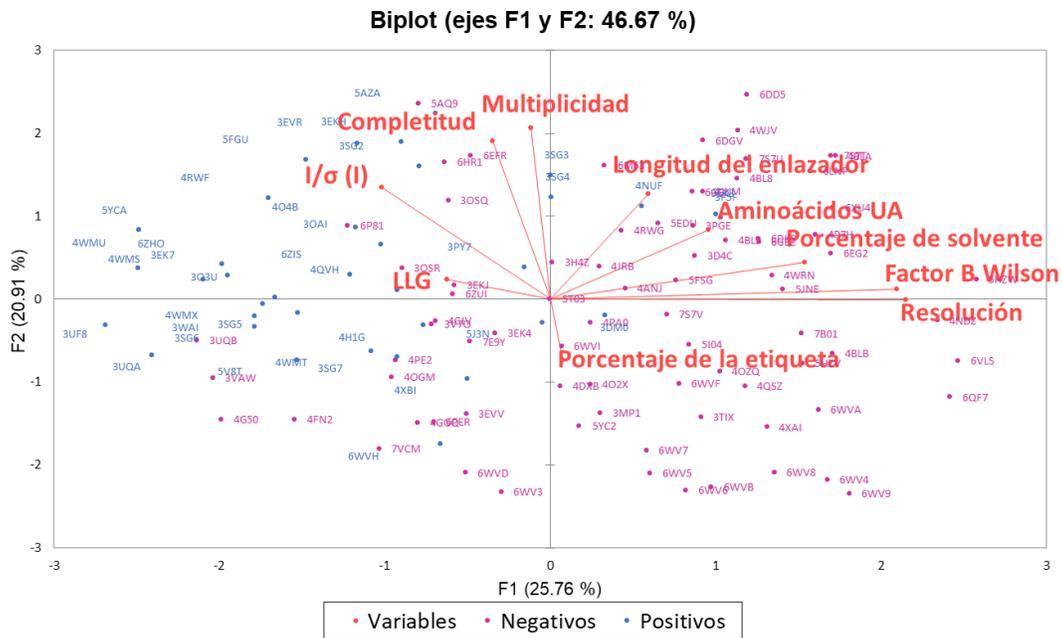


Figura 10.2.3-3: *Biplot* de los resultados del PCA para las 116 proteínas de fusión analizadas. Las proteínas de fusión que dieron resultados positivos se muestran en azul y las que dieron resultados negativos se muestran en violeta. Los resultados positivos se agrupan en el lado izquierdo y los resultados negativos se agrupan en el lado derecho del *biplot*. Los vectores en rojo representan las variables evaluadas y su dirección y longitud indican el peso y la tendencia que presentan en el grupo de datos y en los ejes calculados. Por ejemplo, variables como la resolución, el porcentaje de solvente y el factor B de Wilson, contribuyen más al eje F1. Además, su dirección y longitud indican que valores más bajos de estas tres variables tienden a dar resultados positivos.

Sin embargo, no todas las variables mostraron la misma importancia para determinar el éxito del método evaluado. Los resultados estuvieron influenciados principalmente por la resolución, el factor B de Wilson y el porcentaje de solvente (Tabla 10.2.3-1: Eje F1). A estos les siguieron la multiplicidad y la completitud de los datos (Tabla 10.2.3-1: Eje F2). Además, el porcentaje de la etiqueta, el valor LLG del RM inicial y la longitud del enlazador también contribuyen a explicar la variabilidad (Tabla 10.2.3-1: Eje F3).

Tabla 10.2.3-1: Análisis de componentes principales de las variables cuantitativas que contribuyen a explicar el éxito de la expansión de fase de FP asistida por AMB mediante ARP/wARP. El porcentaje de contribución de las variables cuantitativas en los ejes F1, F2 y F3 y la contribución total del eje. Los valores en negrita corresponden para cada variable al factor para el cual el coseno al cuadrado es mayor. La sumatoria del aporte de los tres ejes explica el 63.93 % de la variabilidad.

Contribución de las variables (%)			
Ejes	F1	F2	F3
Resolución	32.561	0.000	0.707
Porcentaje de la etiqueta	0.042	4.163	30.652
Porcentaje de solvente	16.678	1.514	1.232
Factor B de Wilson	30.861	0.113	0.032
Aminoácidos en la UA	6.416	5.449	13.995
Completitud	0.846	28.475	2.023
Multiplicidad	0.096	33.186	2.890
l/σ (l)	7.268	14.100	0.440
LLG	2.755	0.432	29.717
Longitud del enlazador	2.478	12.567	18.312
Contribución del eje a la variabilidad	25.759	20.907	17.266

10.3 Discusión

Las PFs son utilizadas cada vez más como herramientas en aplicaciones biotecnológicas [2,41,94,95], y las publicaciones que enfatizan los beneficios del uso de PFs para la obtención de estructuras cristalográficas son cada vez más comunes [12,82]. Esta tendencia se refleja en el creciente número de estructuras cristalográficas de PFs en el PDB en comparación con el último reporte realizado por Waugh en 2016 [42]. Hasta agosto de 2023 se encontraron 385 estructuras de PFs depositadas: 252 con MBP, 69 con GFP, 38 con SUMO, 13 con GST y 13 con TRX.

Uno de los principales beneficios del uso de PFs en cristalografía, cuando se utiliza RM, es que no es necesario un modelo inicial para la proteína de interés [96,97]. De hecho, la proteína etiqueta puede actuar como modelo para un RM parcial y las fases pueden extenderse a la proteína de interés [12,32]. Esto permite obtener estructuras de proteínas sin homólogos. Esto es fundamental ya que inclusive los modelos generados con AlphaFold [88] y RosettaFold [98], suelen tener regiones con baja confianza o precisión estructural [99–101], lo que puede dificultar su uso en la determinación de una estructura por RM [102].

Sin embargo, emplear PFs para determinar las fases es una tarea compleja, ya que la selección de una etiqueta adecuada es crucial para obtener un RM inicial útil. Esta limitación surge del hecho de que, como lo demostraron anteriormente Cohen y colaboradores [103], el resultado del RM desempeña un papel vital en la eficacia de la construcción automática. En el presente trabajo se logró determinar que la

opción más favorable parece ser el uso de etiquetas con alta estabilidad y mínima movilidad. En este estudio, GFP surgió como la etiqueta proteica preferida para un RM parcial útil, una proteína caracterizada por su alta estabilidad [104].

Por otro lado, la tasa de éxito del RM con MBP depende de la conformación que adopta en el cristal. Esta proteína tiene la capacidad de cambiar entre una conformación abierta y cerrada, mostrando ocasionalmente estados intermedios, particularmente cuando la proteína está en su estado apo [105,106]. Todos los resultados positivos con MBP presentaron una conformación cerrada, la conformación más estable y común en PDB debido al método de purificación empleado (cromatografía de afinidad de amilosa) que promueve el estado holo [107].

Por el contrario, no se recomienda el uso de SUMO debido a su propensión a presentar regiones de inestabilidad estructural y alta movilidad; estas características generan desafíos tanto en la cristalización como en la resolución de fases. De hecho, no existen depósitos de estructuras cristalográficas independientes de SUMO en el PDB. Todas las estructuras cristalográficas de SUMO implican fusión o complejos con otras proteínas, y varios ejemplos carecen de una densidad electrónica clara en segmentos sustanciales de SUMO.

Un análisis de las PFs depositadas en el PDB muestra que la mayoría de las estructuras fueron determinadas mediante RM parcial con las coordenadas de la etiqueta (Material suplementario, Tabla 1). Cabe destacar que el 77 % de estas

estructuras emplearon construcción manual para ampliar el modelo y calcular nuevas fases. De las PFs reportadas en este trabajo, sólo 45 se determinaron utilizando programas de construcción automática en algún punto (Material suplementario, Tabla 1). Según nuestros resultados, podemos atribuir la adopción limitada del uso de estos programas a su baja tasa de éxito, ya que en con el método evaluado fue posible resolver el 31 % de las proteínas estudiadas. La tasa de éxito puede ser incluso menor si el modelo inicial utilizado para el RM parcial tiene una baja estabilidad estructural generando un pobre resultado de RM parcial. Es recomendable el uso de una combinación de métodos automáticos y manuales para mejorar la tasa de éxito.

El uso de construcción automática basado en fases calculadas a partir de RM no es nuevo y varios estudios han identificado factores que contribuyen al éxito [93,103,108]. Sin embargo, este trabajo es el primer análisis de múltiples casos en los que el modelo de RM exhibe una identidad de secuencia del 100 % pero representa un segmento de la UA. Un antecedente clave fue el realizado por Cohen y colaboradores en 2007 [103], los cuales evaluaron la aplicación de ARP/wARP para ampliar fases derivadas de RM. Los autores determinaron con éxito la estructura de la PF Ubc9-SUMO, que se determinó utilizando únicamente las coordenadas Ubc9, además, determinaron la influencia de diversos factores en los resultados de construcción automática de 129 estructuras cristalográficas. El objetivo de este estudio fue desarrollar un estimador empírico de calidad para indicar la probabilidad de éxito del uso de Flex ARP/wARP. El estimador propuesto se basó en el valor R inicial del modelo, la resolución y el porcentaje de solvente del

cristal, ya que los autores encontraron que estos eran los factores más importantes de éxito. Sus hallazgos son consistentes con otros estudios que evalúan el uso de construcción automática, donde la resolución y el valor R parecen ser los factores más críticos [109–111].

El análisis estadístico realizado en nuestro estudio identificó varios factores que pueden influir en el éxito del uso de RM parcial y ARP/wARP para determinar la estructura completa de PFs. Como informan Cohen y colaboradores, la calidad del modelo de RM inicial, la resolución y el porcentaje de solvente son fundamentales. Sin embargo, es necesario agregar que los cristales con alto orden y baja movilidad de proteínas, indicados por valores bajos del factor Wilson B, pueden mejorar los resultados. Si bien estos factores pueden no ser controlables, nuestro estudio revela que una estrategia de colecta de datos centrada en mejorar la completitud y la multiplicidad pueden aumentar la probabilidad de obtener resultados positivos. Además, el diseño de la PF podría potencialmente conducir a mejores resultados, la selección de la etiqueta apropiada para RM parcial, que represente una porción significativa de la PF, complementada con un enlazador corto, es fundamental.

11. Conclusiones

Las PFs son una herramienta de gran utilidad en la cristalografía de proteínas. El uso de PFs en los procesos de expresión y purificación genera grandes cantidades de proteína soluble, y con purezas suficientes para poner pruebas de cristalización de forma relativamente simple. Además, aumenta considerablemente la solubilidad de proteínas difíciles de trabajar en condiciones acuosas. Sin embargo, cuando se desean usar como herramientas para facilitar la cristalización, la presencia de una etiqueta no garantiza la cristalogénesis. Es necesario evaluar diferentes etiquetas, enlazadores y condiciones; por lo que esta aproximación puede sumarse a un conjunto de variables a evaluar y modificar cuando se busca obtener un cristal de proteína. Se puede recomendar su uso como otra posible forma de cristalizar, incluso evaluando en paralelo con pruebas de cristalización de la proteína de interés de forma independiente, para aumentar las posibilidades de obtener un cristal.

A pesar de esto, puede observarse un aumento significativo en el número de estructuras cristalográficas de PFs en el PDB en la última década. Entre estas estructuras, la utilización de RM parcial junto con construcción automática se destaca como un método raramente empleado para resolver fases. Este empleo poco frecuente podría deberse a una baja tasa de éxito, como se observó en este estudio (aproximadamente 30 %). Sin embargo, encontramos que es posible aumentar la tasa de éxito al obtener cristales con buena resolución, valores bajos del factor Wilson B y un bajo porcentaje de contenido de solvente. Además, los resultados positivos se pueden mejorar empleando datos de difracción que posean

altos valores de multiplicidad y completitud. Si bien varias de estas características son deseables en casi todos los métodos de resolución de fases, en los métodos evaluados aquí su combinación es esencial. Además, un mejor diseño de PFs puede ayudar a mejorar la tasa de éxito. La elección de un enlazador de poca longitud parece mejorar los resultados positivos. Además, es preferible elegir una etiqueta con alta estabilidad estructural (GFP en contraposición a SUMO) y que constituya una porción significativa de la PF.

12. Bibliografía

1. Yu K, Liu C, Kim BG, Lee DY. Synthetic fusion protein design and applications. *Biotechnol Adv.* 2015 Jan 1;33(1):155–64.
2. Chen X, Zaro JL, Shen WC. Fusion protein linkers: Property, design and functionality. *Adv Drug Deliv Rev.* 2013 Oct 15;65(10):1357–69.
3. Kobe B, Ve T, Williams SJ. Fusion-protein-assisted protein crystallization. *Acta Crystallogr F Struct Biol Commun.* 2015 Jan 1;71:861–9.
4. Schmidt SR. Fusion-proteins as biopharmaceuticals-applications and challenges. *Curr Opin Drug Discov Devel.* 2009 Mar;12(2):284–95.
5. Aroul-Selvam R, Hubbard T, Sasidharan R. Domain Insertions in Protein Structures. *J Mol Biol.* 2004 May;338(4):633–41.
6. Czajkowsky DM, Hu J, Shao Z, Pleass RJ. Fc-fusion proteins: new developments and future perspectives. *EMBO Mol Med.* 2012 Oct 26;4(10):1015–28.
7. Terpe K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol.* 2003;60(5):523–33.
8. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B. Crystal structures of fusion proteins with large-affinity tags. *Protein Science.* 2003 Jul;12(7):1313–22.
9. Krauss IR, Merlino A, Vergara A, Sica F. An overview of biological macromolecule crystallization. *Int J Mol Sci.* 2013;14(6):11643–91.

10. McPherson A. Introduction to protein crystallization. *Acta Crystallogr.* 2014;2–20.
11. Chayen NE, Saridakis E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods.* 2008 Feb 30;5(2):147–53.
12. Holcomb J, Spellmon N, Zhang Y, Doughan M, Li C, Yang Z. Protein crystallization: Eluding the bottleneck of X-ray crystallography. *AIMS Biophys.* 2017;4(4):557–75.
13. Martens JHA, Stunnenberg HG. The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett.* 2010 Jun 18;584(12):2662–9.
14. Alcalay M, Meani N, Gelmetti V, Fantozzi A, Fagioli M, Orleth A, et al. Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. *Journal of Clinical Investigation.* 2003 Dec 1;112(11):1751–61.
15. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA. Protein interaction maps for complete genomes based on gene fusion events. *Nature.* 1999 Nov;402(6757):86–90.
16. Osterman A. Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol.* 2003 Apr;7(2):238–51.
17. Yanai I, Derti A, DeLisi C. Genes linked by fusion events are generally of the same functional category: A systematic analysis of 30 microbial genomes. *Proceedings of the National Academy of Sciences.* 2001 Jul 3;98(14):7940–5.
18. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting Protein Function and Protein-Protein Interactions from Genome Sequences. *Science (1979).* 1999 Jul 30;285(5428):751–3.

19. Saadat F. A review on chimeric xylanases: methods and conditions. 3 Biotech. 2017 May 27;7(1):67.
20. Punt PJ, Levasseur A, Visser H, Wery J, Record E. Fungal Protein Production: Design and Production of Chimeric Proteins. Annu Rev Microbiol. 2011 Oct 13;65(1):57–69.
21. Yuste R. Fluorescence microscopy today. Nat Methods. 2005 Dec;2(12):902–4.
22. Hwang PM, Pan JS, Sykes BD. Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. FEBS Lett. 2014 Jan 21;588(2):247–52.
23. Sassenfeld H. Engineering proteins for purification. Trends Biotechnol. 1990;8:88–93.
24. Derewenda ZS. The use of recombinant methods and molecular engineering in protein crystallization. Methods. 2004;34(3):354–63.
25. Smyth MS, Martin JHJ. X Ray crystallography. J Clin Pathol. 2000;53:8–14.
26. Rodríguez DD, Grosse C, Himmel S, González C, de Ilarduya IM, Becker S, et al. Crystallographic ab initio protein structure solution below atomic resolution. Nat Methods. 2009;6(9):651–3.
27. Ladd MFC, Marcus FC, Palmer RA, Rex A. Structure determination by X-ray crystallography. 2003. 1–756 p.

28. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS Journal*. 2013 Nov;280(22):5705–36.
29. Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal*. 2008;275(1):1–21.
30. Terwilliger TC, Grosse-Kunstleve RW, Afonine P V., Moriarty NW, Zwart PH, Hung LW, et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr*. 2007 Dec 4;64(1):61–9.
31. Taylor GL. Introduction to phasing. *Acta Crystallogr D Biol Crystallogr*. 2010;66(4):325–38.
32. Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Science*. 2008 Dec;17(12):2070–9.
33. Booth WT, Schlachter CR, Pote S, Ussin N, Mank NJ, Klapper V, et al. Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. *ACS Omega*. 2018 Jan 31;3(1):760–8.
34. Li DF, Feng L, Hou YJ, Liu W. The expression, purification and crystallization of a ubiquitin-conjugating enzyme E2 from *Agrocybe aegerita* underscore the impact of His-tag location on recombinant protein properties. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2013 Feb 1;69(2):153–7.

35. Ledent P, Duez C, Vanhove M, Lejeune A, Fonzé E, Charlier P, et al. Unexpected influence of a C-terminal-fused His-tag on the processing of an enzyme and on the kinetic and folding parameters. *FEBS Lett.* 1997 Aug 18;413(2):194–6.
36. Eschenfeldt WH, Maltseva N, Stols L, Donnelly MI, Gu M, Nocek B, et al. Cleavable C-terminal His-tag vectors for structure determination. *J Struct Funct Genomics.* 2010 Mar 6;11(1):31–9.
37. Huh I, Gene R, Kumaran J, MacKenzie CR, Brooks CL. In situ proteolysis, crystallization and preliminary X-ray diffraction analysis of a VHH that binds listeria internalin B. *Acta Crystallogr F Struct Biol Commun.* 2014 Nov 1;70:1532–5.
38. Kim KM, Yi EC, Baker D, Zhang KYJ. Post-translational modification of the N-terminal His tag interferes with the crystallization of the wild-type and mutant SH3 domains from chicken src tyrosine kinase. *Acta Crystallogr D Biol Crystallogr.* 2001 May 1;57(5):759–62.
39. Carson M, Johnson DH, McDonald H, Brouillette C, DeLucas LJ. His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr.* 2007 Feb 21;63(3):295–301.
40. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow G V, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* 2021 Jan 8;49(D1):D437–51.
41. Wild R, Hothorn M. The macro domain as fusion tag for carrier-driven crystallization. *Protein Science.* 2017 Feb 1;26(2):365–74.

42. Waugh DS. Crystal structures of MBP fusion proteins. *Protein Science*. 2016 Mar 1;25(3):559–71.
43. Momin AA, Hameed UFS, Arold ST. Passenger sequences can promote interlaced dimers in a common variant of the maltose-binding protein. *Sci Rep*. 2019 Dec 1;9(1):1–9.
44. Liu Y, Manna AC, Li R, Martin WE, Murphy RC, Cheung AL, et al. Crystal structure of the SarS protein from *Staphylococcus aureus*. *J Bacteriol*. 2003;185(14):4219–25.
45. Suzuki N, Hiraki M, Yamada Y, Matsugaki N, Igarashi N, Kato R, et al. Crystallization of small proteins assisted by green fluorescent protein. *Acta Crystallogr D Biol Crystallogr*. 2010 Oct 1;66(10):1059–66.
46. Waugh DS. Crystal structures of MBP fusion proteins. *Protein Science*. 2016;25(3):559–71.
47. Batyuk A, Wu Y, Honegger A, Heberling MM, Plückthun A. DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography. *J Mol Biol*. 2016 Apr;428(8):1574–88.
48. Chun E, Thompson AA, Liu W, Roth CB, Griffith MT, Katritch V, et al. Fusion Partner Toolchest for the Stabilization and Crystallization of G Protein-Coupled Receptors. *Structure*. 2012 Jun;20(6):967–76.
49. Gumpena R, Lountos GT, Waugh DS. MBP-binding DARPins facilitate the crystallization of an MBP fusion protein. *Acta Crystallogr F Struct Biol Commun*. 2018 Sep 1;74(9):549–57.

50. Ecsédi P, Gógl G, Hóf H, Kiss B, Harmat V, Nyitray L. Structure Determination of the Transactivation Domain of p53 in Complex with S100A4 Using Annexin A2 as a Crystallization Chaperone. *Structure*. 2020 Aug 4;28(8):943-953.e4.
51. Zhang K, Wu H, Hoppe N, Manglik A, Cheng Y. Fusion protein strategies for cryo-EM study of G protein-coupled receptors. *Nat Commun*. 2022 Jul 28;13(1):4366.
52. Baxter MA, Jones BD. Two-component regulators control hilA expression by controlling fimZ and hilE expression within *Salmonella enterica* serovar Typhimurium. *Infect Immun*. 2015;83(3):978–85.
53. Boddicker JD, Knosp BM, Jones BD. Transcription of the *Salmonella* invasion gene activator, hilA, requires HilD activation in the absence of negative regulators. *J Bacteriol*. 2003;185(2):525–33.
54. Ellermeier CD, Ellermeier JR, Slauch JM. HilD, HilC and RtsA constitute a feed forward loop that controls expression of the SPI1 type three secretion system regulator hilA in *Salmonella enterica* serovar Typhimurium. *Mol Microbiol*. 2005 Aug;57(3):691–705.
55. Hubiz C, E. J, Golubeva YA, Lin D, Miller LD, Slauch JM. FliZ regulates expression of the *Salmonella* pathogenicity island 1 invasion locus by controlling HilD protein activity in *Salmonella enterica* serovar typhimurium. *J Bacteriol*. 2010;192(23):6261–70.
56. Jones BD. *Salmonella* invasion gene regulation: A story of environmental awareness. *Journal of Microbiology*. 2005;43(July):110–7.

57. Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. AraC/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev.* 1997;61(4):393–410.
58. Egan SM. Growing Repertoire of AraC/XylS Activators. *J Bacteriol.* 2002 Oct 15;184(20):5529–32.
59. Baxter MA, Fahlen TF, Wilson RL, Jones BD. HilE interacts with hilD and negatively regulates hilA transcription and expression of the *Salmonella enterica* serovar typhimurium invasive phenotype. *Infect Immun.* 2003 Mar 1;71(3):1295–305.
60. Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejero M, Sukhan A, et al. Supramolecular Structure of the *Salmonella* typhimurium Type III Protein Secretion System. *Science (1979).* 1998;280:602–5.
61. Paredes-Amaya CC, Valdés-García G, Juárez-González VR, Rudiño-Piñera E, Bustamante VH. The Hcp-like protein HilE inhibits homodimerization and DNA binding of the virulence-associated transcriptional regulator HilD in *Salmonella*. *Journal of Biological Chemistry.* 2018 Apr 27;293(17):6578–92.
62. Grenz JR, Cott Chubiz JE, Thaprawat P, Slauch JM. HilE Regulates HilD by Blocking DNA Binding in *Salmonella enterica* Serovar Typhimurium. *J Bacteriol.* 2018 Apr 15;200(8).
63. Hamed S, Shawky RM, Emara M, Slauch JM, Rao C V. HilE is required for synergistic activation of SPI-1 gene expression in *Salmonella enterica* serovar Typhimurium. *BMC Microbiol.* 2021 Dec 16;21(1):49.

64. Gong H, Vu GP, Bai Y, Chan E, Wu R, Yang E, et al. A *Salmonella* small non-coding rna facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. PLoS Pathog. 2011 Sep;7(9).
65. Lim S, Yun J, Yoon H, Park C, Kim B, Jeon B, et al. Mlc regulation of *Salmonella* pathogenicity island I gene expression via hilE repression. Nucleic Acids Res. 2007 Mar;35(6):1822–32.
66. Paredes CC. Caracterización de la proteína HilD: Un regulador central de la virulencia de *Salmonella enterica*. [Ciudad de México]: Universidad Nacional Autónoma de México; 2019.
67. Zigler JS, Sinha D. β A3/A1-crystallin: More than a lens protein. Prog Retin Eye Res. 2015 Jan;44:62–85.
68. Wistow G. The human crystallin gene families. Hum Genomics. 2012 Dec 1;6(1):26.
69. Wannemacher CF, Spector A. Protein synthesis in the core of calf lens. Exp Eye Res. 1968 Oct;7(4):623–5.
70. Bloemendal H, de Jong W, Jaenicke R, Lubsen NH, Slingsby C, Tardieu A. Ageing and vision: structure, stability and function of lens crystallins. Prog Biophys Mol Biol. 2004 Nov;86(3):407–85.
71. Ghosh KS, Chauhan P. Crystallins and Their Complexes. In 2019. p. 439–60.
72. Andley UP. Crystallins in the eye: Function and pathology. Prog Retin Eye Res. 2007 Jan;26(1):78–98.

73. Parthasarathy G, Ma B, Zhang C, Gongora C, Samuel Zigler J, Duncan MK, et al. Expression of β A3/A1-crystallin in the developing and adult rat eye. *J Mol Histol*. 2011 Feb 4;42(1):59–69.
74. Peterson CA, Piatigorsky J. Preferential conservation of the globular domains of the β A3/A1-crystallin polypeptide of the chicken eye lens. *Gene*. 1986 Jan;45(2):139–47.
75. Sinha D, Esumi N, Jaworski C, Kozak CA, Pierce E, Wistow G. Cloning and mapping the mouse Crygs gene and non-lens expression of [γ]S-crystallin. *Mol Vis*. 1998 Apr 30;4:8.
76. Vendra VPR, Khan I, Chandani S, Muniyandi A, Balasubramanian D. Gamma crystallins of the human eye lens. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2016 Jan;1860(1):333–43.
77. Slingsby C, Wistow GJ, Clark AR. Evolution of crystallins for a role in the vertebrate eye lens. *Protein Science*. 2013 Apr;22(4):367–80.
78. Vanita V, Singh JR, Singh D, Varon R, Sperling K. Novel mutation in the gamma-S crystallin gene causing autosomal dominant cataract. *Mol Vis*. 2009;15:476–81.
79. Thorn DC, Grosas AB, Mabbitt PD, Ray NJ, Jackson CJ, Carver JA. The Structure and Stability of the Disulfide-Linked γ S-Crystallin Dimer Provide Insight into Oxidation Products Associated with Lens Cataract Formation. *J Mol Biol*. 2019 Feb;431(3):483–97.
80. Wu Z, Delaglio F, Wyatt K, Wistow G, Bax A. Solution structure of γ S-crystallin by molecular fragment replacement NMR. *Protein Science*. 2005 Dec;14(12):3101–14.

81. Sagar V, Chaturvedi SK, Schuck P, Wistow G. Crystal Structure of Chicken γ S-Crystallin Reveals Lattice Contacts with Implications for Function in the Lens and the Evolution of the $\beta\gamma$ -Crystallins. *Structure*. 2017 Jul;25(7):1068-1078.e2.
82. Kuge M, Fujii Y, Shimizu T, Hirose F, Matsukage A, Hakoshima T. Use of a fusion protein to obtain crystals suitable for X-ray analysis: Crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. *Protein Science*. 1997;6(8):1783–6.
83. Dale GE, Oefner C, D'Arcy A. The protein as a variable in protein crystallization. *J Struct Biol*. 2003 Apr 1;142(1):88–97.
84. Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science*. 1999;8(8):1668–74.
85. Arai R, Ueda H, Kitayama A, Kamiya N, Nagamune T. Design of the linkers which effectively separate domains of a bifunctional fusion protein. Vol. 14, *Protein Engineering*. 2001.
86. Bucher MH, Evdokimov AG, Waugh DS. Biological Crystallography Differential effects of short affinity tags on the crystallization of *Pyrococcus furiosus* maltodextrin-binding protein. *Acta Crystallogr D Biol Crystallogr*. 2002;D58:392–7.
87. Mittl PR, Ernst P, Plückthun A. Chaperone-assisted structure elucidation with DARPins. *Curr Opin Struct Biol*. 2020 Feb 1;60:93–100.
88. Jumper J, Hassabis D. Protein structure predictions to atomic accuracy with AlphaFold. Vol. 19, *Nature Methods*. Nature Research; 2022. p. 11–2.

89. Erdős G, Pajkos M, Dosztányi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* 2021 Jul 2;49(W1):W297–303.
90. Slabinski L, Jaroszewski L, Rychlewski L, Wilson IA, Lesley SA, Godzik A. XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics.* 2007 Dec 15;23(24):3403–5.
91. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990 Oct;215(3):403–10.
92. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007 Jul 13;40(4):658–74.
93. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr.* 2001 Oct 1;57(10):1445–50.
94. Hoyano Y, Tamashiro I, Akimoto T. Fusion proteins of organophosphorus hydrolase and pHluorin for a whole-cell biosensor for organophosphorus pesticide measurement. *Analytical Sciences.* 2023 Sep 1;39(9):1515–20.
95. Waldow A, Beier LS, Arndt J, Schallenberg S, Vollbrecht C, Bischoff P, et al. cCPE Fusion Proteins as Molecular Probes to Detect Claudins and Tight Junction Dysregulation in Gastrointestinal Cell Lines, Tissue Explants and Patient-Derived Organoids. *Pharmaceutics.* 2023 Jul 19;15(7):1980.
96. Raimondo D, Giorgetti A, Giorgetti A, Bosi S, Tramontano A. Automatic procedure for using models of proteins in molecular replacement. *Proteins: Structure, Function, and Bioinformatics.* 2006 Nov 15;66(3):689–96.

97. Schwarzenbacher R, Godzik A, Grzechnik SK, Jaroszewski L. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr D Biol Crystallogr*. 2004 Jul 1;60(7):1229–36.
98. Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (1979). 2021 Aug 20;373(6557):871–6.
99. McDonald EF, Jones T, Plate L, Meiler J, Gulsevin A. Benchmarking AlphaFold2 on peptide structure prediction. *Structure*. 2023 Jan 5;31(1):111-119.e2.
100. Subramaniam S, Kleywegt GJ. A paradigm shift in structural biology. *Nat Methods*. 2022 Jan 11;19(1):20–3.
101. Terwilliger TC, Poon BK, Afonine P V., Schlicksup CJ, Croll TI, Millán C, et al. Improved AlphaFold modeling with implicit experimental information. *Nat Methods*. 2022 Nov 1;19(11):1376–82.
102. McCoy AJ, Sammito MD, Read RJ. Implications of AlphaFold 2 for crystallographic phasing by molecular replacement. *Acta Crystallogr D Struct Biol*. 2022 Jan 1;78(1):1–13.
103. Cohen SX, Ben Jelloul M, Long F, Vagin A, Knipscheer P, Lebbink J, et al. ARP/wARP and molecular replacement: The next generation. *Acta Crystallogr D Biol Crystallogr*. 2007 Dec 4;64(1):49–60.
104. Zimmer M. Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chem Rev*. 2002 Mar 1;102(3):759–82.

105. Bucher D, Grant BJ, Markwick PR, McCammon JA. Accessing a Hidden Conformation of the Maltose Binding Protein Using Accelerated Molecular Dynamics. *PLoS Comput Biol.* 2011 Apr 21;7(4):e1002034.
106. Telmer PG, Shilton BH. Insights into the Conformational Equilibria of Maltose-binding Protein by Analysis of High Affinity Mutants. *Journal of Biological Chemistry.* 2003 Sep;278(36):34555–67.
107. Tang C, Schwieters CD, Clore GM. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature.* 2007 Oct;449(7165):1078–82.
108. Terwilliger TC, DiMaio F, Read RJ, Baker D, Bunkóczi G, Adams PD, et al. phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta. *J Struct Funct Genomics.* 2012 Jun 15;13(2):81–90.
109. Alharbi E, Bond P, Calinescu R, Cowtan K. Predicting the performance of automated crystallographic model-building pipelines. *Acta Crystallogr D Struct Biol.* 2021 Dec 1;77:1591–601.
110. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.* 2008;3(7):1171–9.
111. Morris RJ, Zwart PH, Cohen S, Fernandez FJ, Kakaris M, Kirillova O, et al. Breaking good resolutions with ARP/wARP. *J Synchrotron Radiat.* 2004;11:56–9.

13. Glosario

Análisis de Componentes Principales (PCA): Es una técnica estadística utilizada para reducir la dimensionalidad de conjuntos de datos de alta dimensionalidad mientras se preserva la mayor cantidad posible de la variabilidad presente en los datos. Logra esto transformando las variables originales en un nuevo conjunto de variables no correlacionadas llamadas componentes principales. Estos componentes principales son combinaciones lineales de las variables originales y están ordenados por la cantidad de varianza que explican en los datos.

Análisis de componentes principales de datos mixtos (PCA Mix): Es una extensión del Análisis de Componentes Principales (PCA) diseñada para manejar conjuntos de datos con una combinación de variables continuas y categóricas. Utiliza diferentes métodos para evaluar las relaciones entre variables continuas y categóricas, permitiendo una exploración completa de las estructuras de datos heterogéneas.

Compleitud: corresponde al porcentaje de reflexiones cristalográficas colectadas en comparación con el número de reflexiones teóricamente posibles únicas para la simetría del cristal.

Cosenos cuadrados: Coeficientes calculados en PCA para representar la proporción de varianza de cada variable original que es explicada por cada componente principal. Son los cuadrados de los coeficientes de correlación entre las variables originales y los componentes principales.

Factor B de Wilson: valor que se obtiene a partir de un análisis del gráfico de la intensidad de reflexión promedio vs $\sin^2(\theta)/\lambda^2$ (el gráfico de Wilson). El factor Wilson B indica el grado de orden en el cristal. Es una aproximación de los valores B promedio que pueden obtenerse posterior al afinamiento.

Grupo espacial: es el conjunto de operaciones de simetría geométrica que forman un objeto periódico tridimensional (es decir, el cristal) en sí mismo.

$I/\sigma(I)$: Es la medida de la relación señal-ruido de las intensidades.

Log Likelihood Gain (LLG): es la probabilidad de que los datos se hubieran medido, dado el modelo, por lo que permite comparar qué tan bien concuerdan los diferentes modelos con los datos.

Multiplicidad: llamada también redundancia, es el número medio de mediciones por reflexión individual y simétricamente única.

Porcentaje de solvente: La fracción del volumen del cristal ocupada por el solvente. Suele ser de alrededor de 43%, pero se han reportado cristales con porcentajes de solvente desde 27 hasta 65%.

Resolución: En cristalografía de rayos X, la resolución es la distancia más pequeña entre los planos de la red cristalina que se resuelve en el patrón de difracción. Los valores numéricos altos de resolución, como 4 Å, significan una resolución deficiente, mientras que los valores numéricos bajos, como 1.5 Å, significan una buena resolución.

Translation Function Z-score (TFZ): medida estadística utilizada para evaluar la calidad del ajuste entre un modelo molecular y los datos experimentales. El TFZ se calcula durante el proceso de refinamiento de la estructura, y un valor alto de TFZ indica un buen ajuste entre el modelo y los datos experimentales.

Unidad asimétrica: es la porción más pequeña de una estructura cristalina a la que se pueden aplicar operaciones de simetría para generar la celda unitaria completa (la unidad repetitiva del cristal).

Variabilidad: Se refiere a la extensión de las diferencias o variaciones dentro de un conjunto de datos o una población. Describe la dispersión, la amplitud o la diversidad de valores dentro de un conjunto de observaciones. Comprender la variabilidad es esencial en el análisis estadístico ya que proporciona información sobre la distribución y las características de los datos, ayudando a hacer inferencias y sacar conclusiones.

14. Material suplementario

14.1 Tabla suplementaria 1

Lista de las proteínas de fusión (PF) reportadas en el PDB y su método de resolución de fases. RM-M: Reemplazo molecular con construcción manual. RM-AMB-M: Reemplazo molecular con construcción de modelos automatizada y manual. RM: Reemplazo molecular. SAD: *Single-wavelength anomalous dispersion*. SAD-RM: *Single-wavelength anomalous dispersion* y reemplazo molecular. SAD-RM-M: *Single-wavelength anomalous dispersion* con reemplazo molecular y construcción manual. MIR: *Multiple isomorphous replacement*.

Porcentaje de MBP en la PF (%)	Código PDB	Resolución (Å)	Método de resolución de fases	Referencia	Año de deposito
97.34	5B3W	2.40	RM-M	[1]	2016
97.23	6EQZ	2.29	RM-M	[2]	2017
97.10	5B3X	2.40	RM-M	[1]	2016
97.09	4XR8	2.25	RM-M	[3]	2016
97.05	6SIV	1.75	RM-M	No publicado	2019
97.04	6SJA	1.50	RM-M	No publicado	2019
96.40	1A7L	2.90	RM-M	[4]	1998
96.12	3W15	1.80	RM-AMB-M	[5]	2013
96.06	4GIZ	2.55	RM-M	[6]	2012
95.54	3LC8	2.00	RM-M	[7]	2011
95.50	3LBS	2.15	RM-M	[7]	2011
95.08	3Q25	1.90	RM-M	[8]	2010
95.08	3Q28	1.60	RM-M	[8]	2010
95.08	3Q27	1.30	RM-M	[8]	2010
95.08	3Q29	2.30	RM-M	[8]	2010
95.06	5B3Y	1.90	RM-M	[1]	2016
94.64	5ZCA	1.80	RM-M	No publicado	2018
94.61	5MM3	2.10	RM-M	[2]	2018
94.58	5E7U	2.80	RM-M	No publicado	2015
94.57	3G7W	1.75	RM-M	[9]	2009

94.39	6N85	2.50	RM	[10]	2018
94.22	6N84	1.75	RM	[10]	2018
94.09	5BMY	2.00	RM-M	[1]	2016
93.37	6LES	2.00	RM-AMB-M	[11]	2019
92.71	6QUG	2.70	SAD	[12]	2019
92.44	4JBZ	2.40	RM-M	[13]	2013
92.02	4RG5	1.70	RM-M	[14]	2014
92.02	4GLI	1.90	RM-M	[15]	2012
91.75	3Q26	1.54	RM-M	[8]	2010
91.29	5B3Z	2.30	RM-M	[1]	2016
91.04	2OK2	2.00	RM-M	[16]	2007
90.95	7BG0	2.89	RM-M	[17]	2021
90.84	3G7V	1.86	RM-M	[9]	2009
90.37	6M4W	3.11	RM-M	[18]	2020
88.86	3WOA	2.00	RM-AMB-M	[19]	2013
88.62	4KYD	2.21	RM-M	[20]	2013
88.41	4KYC	1.95	RM-M	[20]	2013
88.01	6LF3	3.20	RM-AMB-M	[11]	2019
87.77	4YS9	2.00	RM-M	[21]	2015
87.77	4WTH	2.25	RM-M	[21]	2014
87.35	4KYE	2.60	RM-M	[20]	2013
86.94	1MH3	2.10	RM	[22]	2002
86.94	1MH4	2.30	RM	[22]	2002
86.08	7KD4	1.31	RM-M	[23]	2020
85.65	3IO4	3.63	RM-M	[23]	2020
85.61	5AZ6	2.56	RM	[24]	2015
85.61	5AZ8	1.70	RM	[24]	2015
85.61	5AZ7	1.96	RM	[24]	2015
85.19	6WBH	2.46	RM-AMB-M	[25]	2020
84.42	5AZ9	1.82	RM	[24]	2015
83.81	3MQ9	2.80	RM	[26]	2010

83.79	5JST	2.20	RM	[27]	2016
83.75	6XDJ	2.20	RM	No publicado	2020
83.60	5IQZ	2.33	RM-M	[28]	2016
83.37	7MK7	2.43	RM	[29]	2021
83.22	4XZS	2.12	RM-M	[30]	2015
83.15	5Z0V	2.91	RM	[31]	2017
83.15	5Z0R	2.05	RM	[31]	2017
83.03	7FBB	2.31	RM-M	[32]	2021
82.92	6M4V	2.92	RM-M	[18]	2020
82.29	4XZV	3.58	RM-M	[30]	2015
82.22	4FE8	3.00	RM-M	[33]	2012
82.00	7FBD	2.85	RM-M	[32]	2021
81.74	5ZNZ	2.55	RM-M	[34]	2018
81.74	5ZNY	2.74	RM-M	[34]	2018
81.37	4IFP	1.99	RM-M	[35]	2012
81.33	1YTV	1.80	RM-M	[36]	2005
81.04	5ZD4	2.17	RM-AMB-M	[37]	2018
81.04	6KI0	2.00	RM-M	[38]	2019
80.98	5YEV	2.50	RM-M	[34]	2017
80.88	5GPQ	2.10	RM-M	[31]	2016
80.86	5C7R	1.94	RM	[39]	2015
80.84	4IKM	2.46	RM-M	[40]	2012
80.71	3A3C	2.50	RM-M	[41]	2009
80.71	6ANV	2.27	RM-M	[42]	2017
80.67	5GPP	2.00	RM-M	[31]	2016
80.66	5YGP	2.09	RM-M	[34]	2017
80.66	5YGS	2.69	RM-M	[34]	2017
80.56	2VGQ	2.10	RM-AMB-M	[43]	2007
80.44	1MG1	2.50	RM-M	[44]	1999
80.22	6NDJ	2.27	RM	[45]	2018
79.96	2XZ3	1.95	RM	[46]	2010

79.83	3CSG	1.80	RM	[47]	2008
79.83	2OBG	2.35	RM	[48]	2006
79.65	3N94	1.80	RM	[49]	2010
79.61	5WPZ	2.00	RM-M	[50]	2016
79.61	5H7Q	1.45	RM-M	[50]	2016
79.61	4KV3	2.20	RM-AMB-M	[51]	2013
79.44	6KXG	2.81	RM-M	[52]	2019
79.42	4IRL	1.47	RM-M	[35]	2013
79.31	3EHS	2.76	RM	[53]	2008
79.31	3EHT	3.40	RM	[53]	2008
79.31	3EHU	1.96	RM	[53]	2008
79.20	4FEB	2.80	RM-M	[33]	2012
79.20	4FEC	3.00	RM-M	[33]	2012
79.20	4FED	2.81	RM-M	[33]	2012
79.14	3CSB	2.00	RM	[47]	2008
79.10	5H7N	1.85	RM-M	[50]	2016
78.71	2ZXT	3.00	RM-M	[41]	2009
78.63	5GXT	2.25	RM-M	[54]	2016
78.60	1NMU	2.31	RM-M	[55]	2003
78.50	5CL1	3.80	RM	[56]	2015
78.40	1T0K	3.20	RM	[57]	2004
78.29	6PFO	1.78	RM	[58]	2019
78.26	7FBC	1.85	RM-M	[32]	2021
78.21	5GXV	2.10	RM-M	[54]	2016
78.06	4EXK	1.28	RM	No publicado	2012
77.99	6GFJ	3.30	RM-M	[59]	2018
77.97	4MY2	2.40	RM-M	[60]	2013
77.80	3D4C	2.90	RM-M	[61]	2008
77.80	3D4G	2.30	RM	[61]	2008
77.80	3EF7	3.10	RM-SAD	[61]	2008
77.80	6PGQ	2.85	RM	[58]	2019

77.63	1HSJ	2.30	RM-M	[62]	2000
77.47	3N93	2.50	RM	No publicado	2010
77.47	3N95	2.72	RM	No publicado	2010
77.47	3N96	2.75	RM	No publicado	2010
77.10	5WQ6	1.65	RM-M	[50]	2016
77.04	3VD8	2.07	RM-M	[35]	2012
76.65	4R0Y	2.00	RM-M	[63]	2014
76.56	6K7F	1.80	RM	[64]	2019
76.56	6K7D	2.00	RM	[64]	2019
76.56	6K7E	1.53	RM	[64]	2019
76.49	5IIC	2.90	RM-AMB-M	[65]	2016
76.46	4XHS	1.70	RM	No publicado	2015
76.30	5OSQ	2.05	RM-M	[61]	2017
76.07	5DIS	2.85	RM-M	[66]	2015
75.46	6XDS	1.47	RM-M	[67]	2020
75.36	4PQK	3.40	RM	[68]	2014
75.31	3OAI	2.10	RM-M	[69]	2010
75.30	5TTD	2.00	RM	[70]	2016
74.90	5HZ7	1.43	RM-AMB-M	[71]	2016
74.79	4EDQ	1.64	RM	No publicado	2012
74.75	4O2X	2.70	RM	[72]	2013
74.14	5CFV	1.80	RM	[73]	2015
74.14	5IHJ	2.20	RM	[73]	2015
74.09	1Y4C	1.90	RM-AMB-M	[74]	2004
73.79	3OB4	2.71	RM-M	[75]	2010
73.69	4XA2	1.98	RM	No publicado	2014
73.55	5VAW	1.69	RM	[76]	2017
73.30	3HST	2.25	RM-M	[77]	2010
73.11	4PE2	1.72	RM	[78]	2014
73.11	6SMV	2.14	RM	No publicado	2019
72.96	1R6Z	2.80	RM-AMB-M	[79]	2003

72.82	3PY7	2.29	RM-M	[6]	2010
72.67	5II4	2.00	RM-AMB-M	[65]	2016
72.44	5K94	2.10	RM	[80]	2016
72.30	5WVM	2.90	RM-M	[81]	2016
72.18	5II5	1.80	RM-AMB-M	[65]	2016
72.10	6APX	2.49	RM-M	[82]	2017
71.96	3OSQ	1.90	RM-M	[83]	2010
71.91	5WVN	2.90	RM-M	[81]	2016
71.48	6SLM	2.80	RM	[84]	2019
71.40	5JON	2.04	RM-AMB-M	[85]	2016
71.40	4WMU	1.55	RM-M	[86]	2014
71.26	4WMS	1.90	RM-M	[86]	2014
71.26	4WMW	1.90	RM-M	[86]	2014
71.26	4WMT	2.35	RM-M	[86]	2014
71.12	4WGI	1.85	RM-M	[87]	2014
71.12	6DM8	2.70	RM	[88]	2018
71.12	5LOF	2.20	RM	[89]	2016
70.71	6SJV	2.03	RM	No publicado	2019
70.66	4WMV	2.40	RM-M	[86]	2014
70.58	4TSM	1.90	RM	[78]	2014
70.44	6TZC	2.41	RM	[90]	2019
70.44	3MP1	2.60	RM	[91]	2010
70.44	3MP6	1.48	RM	[91]	2010
70.31	3MP8	1.92	RM	[91]	2010
69.83	3C4M	1.95	RM-AMB-M	[92]	2008
69.79	4WVH	2.10	RM-M	[93]	2014
69.73	3L2J	3.24	RM-M	[92]	2009
69.64	6WH0	1.99	RM-M	[94]	2020
69.62	4OGM	2.23	RM	[78]	2014
69.51	5JJ4	2.81	RM-M	[95]	2016
69.33	4WVG	2.05	RM-AMB-M	[93]	2014

69.04	3H3G	1.94	RM-M	[96]	2009
68.47	5W0Z	3.61	RM	[97]	2017
67.77	4RWF	1.76	RM-M	[98]	2014
67.73	4WVJ	1.95	RM-M	[93]	2014
67.73	4WVI	1.90	RM-M	[93]	2014
67.46	5W1C	3.18	RM	[97]	2017
67.04	5DFM	2.30	RM	[99]	2015
66.97	5TJ2	2.80	RM	[100]	2016
66.73	5JQE	3.16	RM-M	[101]	2016
66.67	5W0R	2.40	RM	[97]	2017
66.67	5W0U	2.90	RM	[97]	2017
66.61	5TIB	2.60	RM-M	[100]	2016
66.61	6XRX	1.95	RM-M	[102]	2020
66.31	5TJ4	3.50	RM	[100]	2016
65.95	4EGC	1.99	SAD	[103]	2012
65.60	6AEO	2.30	RM-M	[104]	2018
65.60	4LOG	2.70	RM	[105]	2013
64.84	6KEA	2.35	RM-M	[106]	2019
64.79	4RWG	2.44	RM-M	[98]	2014
64.66	4XAI	2.60	RM	[107]	2014
64.56	4XAJ	3.55	RM-M	[107]	2014
64.50	4WMX	2.00	RM-M	[86]	2014
64.40	6ZHO	1.60	RM-M	[108]	2020
64.40	6ZIS	1.73	RM-M	[108]	2020
64.23	3H4Z	2.35	RM	[109]	2009
64.22	4NUF	2.80	RM-M	[110]	2013
63.72	5Y2G	3.00	RM-M	[111]	2017
63.62	4B3N	3.30	RM	[112]	2012
63.25	3O3U	1.50	RM-M	[113]	2010
62.78	4KEG	2.50	SAD	No publicado	2013
62.78	4N4X	2.50	RM	No publicado	2013

62.31	6X91	3.51	SAD-RM-M	[114]	2020
61.78	4QVH	1.75	RM-AMB-M	[115]	2014
61.30	4O4B	1.80	RM-M	[116]	2013
61.02	4DXB	2.29	RM	[117]	2012
60.33	5HZV	2.70	RM-AMB-M	[118]	2016
60.20	7B01	2.80	RM-M	[119]	2021
59.55	6QF7	4.00	RM-M	[120]	2019
59.32	6EG2	2.98	RM	[121]	2018
58.94	6EG3	2.84	RM	[121]	2018
56.51	4DXC	2.30	RM	[117]	2012
56.03	4NDZ	3.45	RM	No publicado	2013
55.62	3F5F	2.65	RM	[122]	2008
54.65	5HZW	4.45	RM-AMB-M	[118]	2016
54.32	5I04	2.42	RM-AMB-M	[118]	2016
54.15	6DD5	2.85	RM-M	[123]	2018
54.10	4WRN	3.20	RM-AMB-M	[124]	2014
53.40	4QSZ	2.86	RM	No publicado	2014
53.11	5FSG	3.21	SAD-RM-M	[125]	2016
53.03	5T03	2.10	RM	[126]	2016
52.97	3DM0	2.40	RM-M	[127]	2008
52.89	5T05	1.95	RM	[126]	2016
52.89	5T0A	1.95	RM	[126]	2016
51.89	5EDU	2.79	RM-M	[128]	2015
51.60	4OZQ	2.71	RM-M	[129]	2014
51.14	4H1G	2.15	RM-M	[130]	2012
50.69	3OSR	2.00	RM-M	[83]	2010
50.14	3WAI	1.90	SAD-RM-M	[131]	2013
50.00	4WJV	3.20	RM-M	[132]	2014
49.66	4BLB	2.80	RM-M	[133]	2013
49.46	4BLA	3.50	RM-M	[133]	2013
49.10	6DKS	2.78	RM-M	[134]	2018

48.74	4BLD	2.80	RM	[133]	2013
48.35	5AQ9	1.86	RM	[135]	2015
46.46	4BL9	2.80	RM-M	[133]	2013
46.21	2NVU	2.80	RM	[136]	2006
44.88	4BL8	3.04	RM-M	[133]	2013
44.47	5E24	2.14	RM	[137]	2015
43.62	5AZA	2.08	RM	[24]	2015
42.82	6HD8	2.40	RM-AMB-M	[138]	2017
40.82	7CY4	2.20	SAD-RM-M	[139]	2020
39.92	4JKM	2.26	RM	[140]	2015
39.91	7CY5	2.20	RM	[139]	2020
39.91	7CY6	2.10	RM	[139]	2020
39.91	7CY7	2.15	RM	[139]	2020
39.91	7CY8	2.40	RM	[139]	2020
38.35	6VLS	3.20	SAD-RM-M	[141]	2020

Porcentaje de GFP en la PF (%)	Código PDB	Resolución (Å)	Método de resolución de fases	Referencia	Año de deposito
87.69	4Z4M	2.15	RM-M	[142]	2015
85.00	7BWN	2.40	RM-M	[143]	2020
84.13	3VHT	2.40	RM	[142]	2011
82.07	4Z4K	2.80	RM-M	[142]	2011
80.57	3AI4	1.60	RM-AMB-M	[144]	2010
79.80	4NDJ	1.85	RM-M	[145]	2013
79.80	4NDK	2.30	RM-M	[145]	2013
74.27	3AI5	1.40	RM-M	[144]	2010
73.79	4BDU	3.00	RM-M	[146]	2012
73.79	5KTG	2.80	RM	[147]	2016
73.40	6WVE	2.43	RM-AMB-M	[148]	2020
72.84	4U2V	2.30	RM	[149]	2014

67.44	3U8P	2.75	RM	[150]	2011
66.48	5WTS	3.00	RM	No publicado	2016
66.47	4PFE	2.60	RM	[151]	2014
65.19	2JAD	2.70	RM-M	[152]	2006
60.51	6WV5	2.20	RM-AMB-M	[153]	2020
59.84	7VCM	1.85	RM-M	[154]	2021
59.58	4XBI	2.01	RM-M	[72]	2014
59.34	5J3N	2.45	RM-AMB-M	[155]	2016
58.72	6WV7	2.48	RM-AMB-M	[153]	2020
58.72	6WV6	2.70	RM-AMB-M	[153]	2020
58.72	6WVH	1.99	RM-AMB-M	[153]	2020
58.72	6WV3	2.20	RM-AMB-M	[153]	2020
58.35	6XU4	3.18	RM	[156]	2020
57.39	6WVF	2.90	RM-AMB-M	[148]	2020
57.00	6WV8	3.01	RM-AMB-M	[153]	2020
57.00	6WV9	3.35	RM-AMB-M	[153]	2020
57.00	6WVA	3.35	RM-AMB-M	[153]	2020
57.00	6WVB	2.87	RM-AMB-M	[153]	2020
57.00	6WVI	2.40	RM-AMB-M	[153]	2020
54.57	4JRB	2.41	RM-M	[157]	2013
54.52	6WVD	2.25	RM-AMB-M	[148]	2020
53.92	6HR1	1.90	RM-M	[158]	2018
46.40	6LNP	2.99	RM	[159]	2019
42.88	6PER	2.10	RM-M	[160]	2019
42.88	6URU	2.60	RM-M	No publicado	2019
42.05	6DGV	2.80	RM	[161]	2018
40.76	6GEL	2.51	RM-AMB-M	[162]	2018
40.76	6GEZ	2.47	RM-AMB-M	[162]	2018
40.18	5FGU	1.90	RM-M	[163]	2015
39.66	3EVR	2.00	RM	[164]	2008
39.26	7E9Y	2.25	RM	[165]	2021

39.02	3EVV	2.60	RM	[164]	2008
37.95	6WV4	3.01	RM-AMB-M	[153]	2020
36.45	3OSR	2.00	RM	[83]	2010
36.30	3EK7	1.85	RM	[166]	2008
36.30	3EKH	2.00	RM	[166]	2008
36.30	3EK4	2.65	RM	[166]	2008
36.18	6M63	2.25	RM-M	[167]	2020
36.16	3SG4	2.40	RM	[168]	2011
36.16	3SG7	1.90	RM	[168]	2011
36.01	3OSQ	1.90	RM	[83]	2010
31.61	6ZUI	2.20	RM	[169]	2020
31.40	3EK8	2.80	RM	[166]	2008
31.40	3EKJ	2.80	RM	[166]	2008
31.40	3SG2	2.00	RM	[168]	2011
31.33	3SG6	1.70	RM	[168]	2011
31.25	3SG5	1.90	RM	[168]	2011
31.18	3SG3	2.70	RM	[168]	2011
31.18	3EVU	1.75	RM	[164]	2008
27.50	6EFR	2.40	RM	[170]	2018
26.82	7S7T	3.20	SAD-RM	[171]	2021
26.82	7S7U	2.95	RM	[171]	2021
26.82	7S7V	2.50	RM	[171]	2021
22.58	4P7H	3.20	RM	[172]	2014
22.27	4PA0	2.25	RM-AMB-M	[173]	2014
22.24	4ANJ	2.60	RM-M	[174]	2012
6.90	4RH7	3.41	MIR	[175]	2014

Porcentaje de SUMO en la PF (%)	Código PDB	Resolución (Å)	Método de resolución de fases	Referencia	Año de deposito
49.00	3PGE	2.80	RM-M	[176]	2010

46.38	3TIX	2.90	SAD	[177]	2011
45.45	3UF8	1.50	RM	[178]	2011
45.45	4GGQ	1.95	RM	[178]	2011
45.45	3UQA	1.55	RM	[178]	2011
45.45	3VAW	1.55	RM	[178]	2011
45.45	4FN2	1.95	RM	[178]	2011
45.45	4G50	1.75	RM	[178]	2011
45.45	4GIV	2.45	RM	[178]	2011
45.45	3UQB	1.90	RM	[178]	2011
41.85	3V7O	2.25	RM	[179]	2011
40.38	5V8T	2.10	RM	No publicado	2017
36.23	5YC2	2.70	RM	[180]	2017
36.23	5YCA	1.57	RM	[180]	2017
31.67	6UKM	1.74	RM	[181]	2019
31.67	6UKU	1.68	RM	[181]	2019
31.67	6UKV	1.83	RM	[181]	2019
31.67	6UKW	1.97	RM	[181]	2019
31.67	6UKX	1.93	RM	[181]	2019
31.67	6UKZ	1.52	RM	[181]	2019
31.67	6UL0	1.76	RM	[181]	2019
30.16	7PWH	1.90	RM	[182]	2021
30.16	7PWI	1.33	RM	[182]	2021
29.61	6VEL	2.65	RM	[183]	2020
27.86	6BR9	2.20	SAD	[120]	2017
26.67	3IX6	2.20	RM	No publicado	2009
26.57	6FMP	2.92	RM	[184]	2018
26.57	6FMQ	2.10	RM	[184]	2018
25.71	4DA1	2.38	RM	[185]	2012
24.00	5VWM	1.80	RM	No publicado	2017
22.58	5M23	1.97	RM	[186]	2016
22.58	5M25	2.43	RM	[186]	2016

22.54	7P7I	1.70	RM-AMB-M	No publicado (BioRxiv)	2021
21.53	5JNE	2.85	RM-AMB-M	[187]	2016
21.21	6P81	1.75	RM-AMB-M	No publicado	2019
20.73	6PTH	3.05	RM	No publicado	2019
17.48	6OTE	2.95	RM	No publicado	2019
15.38	6CFP	2.45	RM	No publicado	2018

Porcentaje de GST en la PF (%)	Código PDB	Resolución (Å)	Método de resolución de fases	Referencia	Año de deposito
95.61	4ECB	2.20	RM-M	[188]	2012
94.37	4ECC	2.20	RM-M	[188]	2012
93.53	1GNE	2.50	RM-M	[189]	1994
92.74	1DUG	1.80	RM-M	[190]	2000
85.83	1BG5	2.60	RM-M	[191]	1998
77.50	1B8X	2.70	RM-M	[192]	1999
55.64	7AL7	1.80	RM	[193]	2020
36.45	7ESG	2.53	RM	No publicado	2021
29.90	7NT8	2.22	RM	[194]	2021
8.05	4AI6	3.40	RM	[195]	2012
8.05	4AKG	3.30	RM	[195]	2012
8.05	4AKH	3.60	RM	[195]	2012
8.05	4AKI	2.70	SAD	[195]	2012

Porcentaje de TRX en la PF (%)	Código PDB	Resolución (Å)	Método de resolución de fases	Referencia	Año de deposito
52.43	6GD1	2.01	RM-M	[196]	2018
48.65	3DXB	2.20	RM	[197]	2008
29.51	6Y4Y	1.75	RM	[198]	2020
29.51	6Y73	1.70	RM	[198]	2020
29.51	6Y4Z	1.90	RM	[198]	2020

29.03	7SCD	2.90	RM	[199]	2021
28.50	6YT3	2.85	RM-M	[158]	2020
25.96	6IBL	2.70	RM	[200]	2018
25.96	6H7L	2.70	RM	[201]	2018
25.96	6H7J	2.80	RM	[201]	2018
25.96	6H7M	2.76	RM	[201]	2018
25.96	6H7O	2.80	RM	[201]	2018
15.61	4KCA	1.90	RM	[202]	2013

14.2 Tabla suplementaria 2

Entradas de PDB, valores de las variables evaluadas y resultados del Método 1 de las 116 proteínas de fusión evaluadas. Los códigos PDB resaltados en amarillo corresponden a las proteínas de fusión con MBP, en verde con GFP y en naranja con SUMO. Los porcentajes resaltados en azul indican resultados positivos, mientras que los violetas indican resultados negativos. NA: valor no disponible.

Código PDB	Porcentaje construido	Resolución	LLG	Aminoácidos en la UA	Porcentaje de la etiqueta	Porcentaje de solvente	Complejidad	Multiplicidad	I/σ	Factor B de Wilson	Fuente de rayos X	Tipo de detector	Grupo espacial
4WMU	96.6	1.55	15930	518	70.8	42.7	100.0	6.1	17.91	15.3	Ánodo rotatorio	CCD	P2 ₁ 2 ₁ 2
4PE2	68.8	1.72	11505	516	71.1	59.3	95.3	6.7	11.99	24.4	Sincrotrón	Pixel	I121
4RWF	81.4	1.76	5123	591	62.3	34.8	99.9	7.2	30.13	28.4	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4WMS	82.8	1.90	7617	518	70.8	43.3	98.2	13.4	27.93	17.3	Ánodo rotatorio	CCD	P2 ₁ 2 ₁ 2
4WMX	82.8	2.00	5719	518	70.8	43.2	98.7	8.4	20.58	23.0	Ánodo rotatorio	CCD	P2 ₁ 2 ₁ 2
3OAI	82.0	2.10	12618	1014	72.2	43.1	99.9	3.5	17.80	26.7	Sincrotrón	Image Plate	P12 ₁ 1
4OGM	24.3	2.23	5337	520	69.6	42.1	98.7	2.0	14.14	26.6	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
3PY7	79.9	2.29	5329	523	70.2	53.0	96.1	6.4	33.60	42.2	Sincrotrón	CCD	C222 ₁
4WMT	96.6	2.35	4296	518	70.8	44.1	98.2	6.9	14.78	24.7	Ánodo rotatorio	CCD	P2 ₁ 2 ₁ 2
3MP1	0.0	2.60	2241	522	70.3	43.2	96.9	3.2	9.50	38.4	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4O2X	0.0	2.70	6706	1014	72.4	32.6	97.3	2.7	15.80	78.0	Sincrotrón	Pixel	P6 ₁
4NUF	80.8	2.80	2965	580	63.4	59.6	100.0	9.9	12.80	33.5	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
3D4C	0.0	2.90	1495	481	76.5	57.6	99.3	11.1	33.10	92.5	Sincrotrón	CCD	I222
3WAI	89.6	1.90	4948	739	50.1	45.4	99.5	3.6	25.00	23.6	Sincrotrón	CCD	C121
4H1G	79.7	2.15	6069	715	50.2	45.7	98.8	3.2	12.02	25.2	Sincrotrón	CCD	P12 ₁ 1
3DM0	83.1	2.40	3754	694	52.7	59.5	98.8	4.2	12.40	36.3	Ánodo rotatorio	CCD	P1
3F5F	88.9	2.65	12328	1316	54.9	67.9	99.6	6.3	14.10	46.8	Ánodo rotatorio	CCD	P6 ₁
4OZQ	0.0	2.71	1803	720	51.5	53.8	94.6	3.1	10.72	40.8	Sincrotrón	CCD	P1
4NDZ	0.0	3.45	99	3948	55.8	53.8	99.1	6.0	NA	88.5	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4BLA	0.0	3.50	99	2976	49.5	56.3	99.9	11.7	28.10	110.8	Sincrotrón	CCD	P2 ₁ 2 ₁ 2
4BLB	0.0	2.80	583	3024	48.7	53.1	98.2	3.5	10.40	80.0	Sincrotrón	Pixel	P12 ₁ 1
4BL9	11.3	2.80	25966	3324	44.3	55.8	98.6	7.7	12.90	59.2	Sincrotrón	CCD	P1
6QF7	0.0	4.00	226	1256	59.6	67.6	94.8	2.5	3.80	39.9	Sincrotrón	Pixel	P4
5I04	0.0	2.42	1304	692	53.6	58.6	99.0	3.4	11.80	53.2	Sincrotrón	Pixel	P6 ₅
6EG2	0.0	2.98	941	621	58.9	69.3	97.9	12.7	18.20	91.5	Sincrotrón	Pixel	P4 ₂ 2 ₁ 2

5HZW	0.0	4.45	153	691	54.4	68.4	99.7	9.5	3.80	163.2	Sincrotrón	Pixel	P6 ₄
6DD5	0.0	2.85	4235	1356	52.9	61.8	100.0	14.8	17.70	54.0	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4WRN	0.0	3.20	705	1390	54.1	48.0	99.0	6.3	16.20	97.7	Sincrotrón	Pixel	H3 ₂
4QSZ	0.0	2.86	785	1374	52.5	50.3	97.7	1.3	NA	38.0	Sincrotrón	CCD	P12 ₁ 1
5FSG	0.0	3.21	763	698	51.4	51.3	99.0	4.3	14.21	29.1	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
5T03	0.0	2.10	3294	1384	53.0	53.0	99.7	6.0	6.30	29.6	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
5EDU	0.0	2.79	688	1490	51.5	49.0	100.0	2.5	23.70	55.6	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4WJV	0.0	3.20	3791	3048	50.0	49.8	100.0	11.4	15.60	64.9	Sincrotrón	CCD	C121
6DKS	0.0	2.78	2116	1658	49.1	55.3	99.3	3.7	14.90	59.7	Sincrotrón	CCD	P12 ₁ 1
5AQ9	0.0	1.86	1458	1834	43.1	49.8	100.0	13.1	21.83	29.7	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
4BL8	0.0	3.04	1805	1662	44.3	51.1	99.9	10.9	17.00	90.1	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
5AZA	90.8	2.08	5154	872	42.3	49.0	100.0	12.0	18.70	34.5	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4JKM	0.0	2.26	246	2004	39.9	55.6	99.6	7.4	12.00	45.8	Sincrotrón	CCD	C222 ₁
6VLS	0.0	3.20	985	3864	38.0	59.1	97.3	2.3	6.11	70.2	Sincrotrón	Pixel	P12 ₁ 1
7B01	0.0	2.80	1592	608	60.2	59.0	92.2	6.5	12.00	61.0	Sincrotrón	Pixel	P3 ₂ 2 ₁
5HZV	0.0	2.72	2006	626	59.3	57.8	98.4	3.3	7.70	63.0	Sincrotrón	Pixel	P6 ₅
4DXB	0.0	2.29	3500	1274	56.5	51.9	95.9	1.9	12.90	32.2	Sincrotrón	CCD	P1
4O4B	97.4	1.80	3467	646	61.3	49.8	99.4	5.0	39.00	27.5	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4QVH	100.0	1.75	7781	1196	61.4	53.3	99.3	7.4	10.50	23.2	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
3O3U	100.0	1.50	16347	581	61.6	55.9	99.3	1.1	32.90	16.2	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
6ZIS	83.4	1.73	5617	594	62.1	37.7	99.9	7.6	6.31	27.0	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
6ZHO	94.5	1.60	8750	594	62.1	37.7	99.6	9.3	12.30	13.5	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
4XAI	0.0	2.60	3541	1146	63.9	54.3	92.8	2.4	7.80	58.5	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
4RWG	0.0	2.44	1469	1779	62.1	51.9	99.8	4.2	24.30	56.6	Sincrotrón	CCD	C121
3H4Z	0.0	2.35	6790	1704	63.6	51.9	97.8	5.3	28.40	41.3	Ánodo rotatorio	CCD	C121
4XBI	89.6	2.01	10854	762	59.6	50.7	98.6	5.0	10.90	36.5	Sincrotrón	Pixel	P6 ₅
5J3N	97.5	2.45	6490	792	59.3	51.7	99.2	4.7	11.70	33.2	Sincrotrón	Pixel	C222 ₁
6WV7	0.0	2.48	1493	780	58.7	58.6	87.5	1.8	14.20	36.8	Sincrotrón	Pixel	P1
6WV6	0.0	2.70	1722	390	58.7	60.3	92.0	2.4	10.40	51.3	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2
6WVH	95.9	1.99	9789	780	58.7	58.0	90.7	2.3	12.50	23.6	Sincrotrón	Pixel	P1
6XU4	0.0	3.18	3626	1239	58.4	63.7	100.0	7.0	13.60	92.7	Sincrotrón	Pixel	P4 ₃ 2 ₁ 2
6WVF	0.0	2.90	1430	399	57.4	51.4	98.9	5.8	11.20	75.7	Sincrotrón	Pixel	P3 ₂ 2 ₁
6WV8	0.0	3.01	1940	414	57.0	59.0	93.9	2.7	8.20	63.0	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
6WV9	0.0	3.35	1431	414	57.0	61.1	93.1	2.3	7.80	78.5	Sincrotrón	Pixel	C121

6WVA	0.0	3.35	1468	414	57.0	60.9	98.2	3.7	10.80	84.8	Sincrotrón	Pixel	C121
6WVB	0.0	2.87	1695	414	57.0	53.5	93.7	2.6	7.80	57.6	Sincrotrón	Pixel	I222
6WVI	53.5	2.40	4490	414	57.0	60.4	99.5	4.0	13.10	46.7	Sincrotrón	Pixel	C121
4JRB	61.5	2.41	3921	416	54.6	60.7	99.3	7.1	14.90	48.7	Sincrotrón	CCD	C222 ₁
6WVD	0.0	2.25	2020	420	54.5	42.8	94.5	2.5	11.00	37.8	Sincrotrón	Pixel	C121
6HR1	71.8	1.90	12277	868	53.9	53.3	99.9	13.1	14.12	28.7	Sincrotrón	Pixel	P12 ₁ 1
6LNP	0.0	2.99	311	806	46.4	57.8	99.8	13.6	12.10	52.7	Sincrotrón	Pixel	P4 ₁
6PER	65.0	2.10	2554	555	42.9	48.9	98.1	3.1	11.40	33.1	Sincrotrón	Pixel	I222
6DGV	0.0	2.80	857	566	42.0	58.8	100.0	21.5	14.70	52.4	Sincrotrón	CCD	P6 ₄
6GEL	0.0	2.51	2189	1104	40.8	59.8	98.7	6.6	24.16	53.2	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
6GEZ	0.0	2.47	1553	1104	40.76	59.9	97.5	6.4	14.95	60.8	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
5FGU	87.5	1.90	5625	565	40.2	53.6	100.0	7.3	25.60	23.3	Sincrotrón	CCD	C222 ₁
3EVR	93.6	2.00	2886	411	34.1	68.5	100.0	17.4	23.60	21.2	Sincrotrón	CCD	P4 ₁ 2 ₁ 2 ₁
7E9Y	0.0	2.25	445	339	71.7	53.6	99.9	5.3	5.75	19.3	Sincrotrón	Pixel	C222 ₁
3EVB	0.0	2.60	675	451	39.0	36.6	98.0	3.3	8.00	21.7	Sincrotrón	CCD	C121
6WV4	0.0	3.01	189	390	37.9	64.7	91.1	2.6	5.40	57.6	Sincrotrón	Pixel	P2 ₁ 22 ₁
6WV5	0.0	2.20	1583	390	60.5	68.6	94.3	2.3	11.50	58.1	Sincrotrón	Pixel	H3
3OSR	0.0	2.00	858	1306	36.4	50.3	100.0	5.7	11.40	19.6	Sincrotrón	CCD	P12 ₁ 1
3EK7	96.9	1.85	2206	449	36.3	38.1	98.5	7.5	18.80	21.6	Sincrotrón	CCD	C121
3EKH	84.3	2.00	3422	449	36.3	64.5	100.0	14.1	21.50	26.6	Sincrotrón	CCD	P4 ₁ 2 ₁ 2 ₁
3EK4	2.7	2.65	1504	449	36.3	48.7	98.7	3.1	20.44	36.4	Sincrotrón	CCD	P2 ₁ 2 ₁ 2
6M63	4.4	2.25	869	774	36.2	53.8	100.0	14.5	8.30	34.0	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
3SG4	95.3	2.40	1810	448	36.2	64.3	99.8	13.7	14.70	33.4	Sincrotrón	CCD	P4 ₁ 2 ₁ 2
3SG7	98.7	1.90	2066	448	36.2	43.0	95.5	7.2	10.20	30.1	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
3OSQ	0.0	1.90	803	661	36.0	57.5	99.9	11.1	14.30	26.9	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
6ZUI	0.0	2.20	823	446	31.6	46.6	95.0	9.1	15.80	34.5	Sincrotrón	Pixel	C222 ₁
3EK8	77.3	2.80	1344	449	31.4	64.7	100.0	11.9	8.60	40.9	Ánodo rotatorio	CCD	P4 ₁ 2 ₁ 2
3EKJ	1.3	2.80	1107	449	31.4	42.3	98.9	7.1	15.40	26.5	Sincrotrón	CCD	C121
3SG2	92.3	2.00	2467	449	31.4	63.9	99.9	13.0	20.80	26.9	Sincrotrón	CCD	P4 ₁ 2 ₁ 2
3SG6	94.5	1.70	3351	450	31.3	39.5	93.9	6.5	15.90	25.3	Ánodo rotatorio	CCD	C121
3SG5	92.0	1.90	2372	448	31.3	35.8	99.6	5.5	10.90	25.7	Sincrotrón	CCD	C121
3SG3	88.9	2.70	2501	448	31.3	64.7	99.9	13.9	15.30	27.0	Sincrotrón	CCD	P4 ₁ 2 ₁ 2
6EFR	9.0	2.40	1905	520	27.5	50.8	99.1	32.8	47.79	36.6	Sincrotrón	CCD	I222

7S7T	0.0	3.20	2563	537	44.7	71.0	99.5	19.3	17.74	115.6	Sincrotrón	Pixel	P6 ₁
4P7H	1.7	3.20	105	2046	22.6	54.6	99.5	11.3	9.00	45.2	Ánodo rotatorio	Image Plate	P1
4PA0	0.0	2.25	136	2048	22.3	52.9	93.3	7.4	9.00	26.0	Sincrotrón	CCD	P12 ₁ 1
4ANJ	18.3	2.60	795	1052	22.2	59.4	100.0	4.8	7.20	26.9	Sincrotrón	CCD	C121
7S7U	14.8	2.95	3251	537	44.7	58.8	99.5	20.0	21.00	99.2	Sincrotrón	Pixel	P6 ₁
7VCM	48.7	1.85	6627	772	59.8	44.4	93.5	1.9	6.30	16.7	Sincrotrón	Pixel	P1
7S7V	0.0	2.50	4726	2148	44.7	52.5	98.5	3.5	10.69	40.0	Sincrotrón	Pixel	P1
6WV3	63.9	2.20	5092	390	58.7	56.3	92.0	2.4	10.40	33.9	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁
3PGE	62.7	2.80	1854	200	49.0	74.5	100.0	12.6	14.80	84.2	Sincrotrón	CCD	F432
3TIX	0.0	2.90	5203	414	46.4	58.4	99.7	1.9	8.04	64.0	Sincrotrón	CCD	C222 ₁
3UF8	86.8	1.50	1025	209	45.5	36.2	99.8	4.0	23.06	12.2	Sincrotrón	CCD	P12 ₁ 1
4GGQ	0.0	1.95	769	836	45.5	41.9	97.0	2.4	12.61	27.4	Sincrotrón	CCD	P1
3UQA	85.9	1.55	902	209	45.5	38.1	99.4	4.0	17.26	16.5	Sincrotrón	CCD	C121
3VAW	0.0	1.55	498	209	45.5	52.5	97.3	3.9	24.33	12.1	Sincrotrón	CCD	P12 ₁ 1
3V7O	0.0	2.25	328	454	41.9	50.6	99.7	4.8	15.20	29.8	Sincrotrón	CCD	P2 ₁ 2 ₁ 2 ₁
3UQB	0.0	1.90	514	209	45.5	36.1	99.8	4.0	16.56	18.1	Sincrotrón	CCD	P12 ₁ 1
5V8T	82.2	2.10	1110	416	40.4	40.6	99.8	4.1	11.51	20.8	Sincrotrón	CCD	P12 ₁ 1
5YC2	0.0	2.70	180	414	36.2	49.1	97.9	3.5	11.70	47.2	Sincrotrón	Pixel	P1
5JNE	15.9	2.85	5723	734	21.5	64.1	99.0	3.6	9.10	52.3	Sincrotrón	CCD	C121
6P81	0.0	1.75	39	462	21.2	52.9	99.9	7.4	18.00	26.9	Sincrotrón	CCD	I222
4FN2	0.0	1.95	174	418	45.5	37.5	97.9	3.8	12.27	18.9	Sincrotrón	CCD	P1
4G50	0.0	1.75	363	418	45.5	33.0	97.1	3.2	16.44	19.4	Sincrotrón	CCD	P1
4GIV	0.0	2.45	1154	418	45.5	44.7	99.9	3.7	15.96	36.3	Sincrotrón	CCD	C121
5YCA	88.9	1.57	1164	207	36.2	50.0	99.9	12.9	51.70	17.5	Sincrotrón	Pixel	P2 ₁ 2 ₁ 2 ₁

14.3 Tabla suplementaria 3

Desviación cuadrática media (RMSD) de la región de la proteína de interés en las 36 estructuras construidas en comparación con la estructura depositada.

Código PDB	Porcentaje construido	RMSD	Coincidencias
4QVH	100.0	0.067	226/226
3O3U	100.0	0.162	208/208
3SG7	98.7	0.306	149/149
5J3N	97.5	0.234	224/229
4O4B	97.4	0.266	238/240
3EK7	96.9	0.220	143/143
4WMU	96.6	0.069	146/146
4WMT	96.6	0.252	139/145
6WVH	95.9	0.431	151/152
3SG4	95.3	0.336	125/125
6ZHO	94.5	0.283	190/192
3SG6	94.5	0.122	100/119
3EVR	93.6	0.222	148/418
3SG2	92.3	0.336	147/148
3SG5	92.0	0.255	133/133
5AZA	90.8	0.173	380/400
3WAI	89.6	0.102	305/339
4XBI	89.6	0.614	226/229
3F5F	88.9	0.254	254/263
3SG3	88.9	0.273	125/127
5YCA	88.9	0.229	131/131
5FGU	87.5	0.209	226/229
3UF8	86.8	0.047	107/107
3UQA	85.9	0.229	112/112
3EKH	84.3	0.176	85/85
6ZIS	83.4	0.139	169/169
3DM0	83.1	0.295	269/275
4WMS	82.8	0.094	125/125
4WMX	82.8	0.095	120/121
5V8T	82.2	0.198	188/189
3OAI	82.0	0.287	117/119
4RWF	81.4	0.219	176/178
4NUF	80.8	0.952*	118/134
3PY7	79.9	0.246	117/117
4H1G	79.7	0.247	265/267
3EK8	77.3	0.220	143/143

* El valor alto se debe a que ARP/wARP construye las coordenadas en dos celdas unitarias diferentes, pero vecinas. Este valor disminuye por debajo de 0.300 si se comparan de forma independiente las secciones individuales que se separaron en la construcción con ARP/wARP.

14.4 Bibliografía suplementaria

1. Hanazono Y, Takeda K, Miki K. Structural studies of the N-terminal fragments of the WW domain: Insights into co-translational folding of a beta-sheet protein. *Sci Rep*. 2016 Dec 4;6(1):34654.
2. Nudelman H, Perez Gonzalez T, Kolushiva S, Widdrat M, Reichel V, Peigneux A, et al. The importance of the helical structure of a MamC-derived magnetite-interacting peptide for its function in magnetite formation. *Acta Crystallogr D Struct Biol*. 2018 Jan 1;74(1):10–20.
3. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, et al. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. *Nature*. 2016 Jan 28;529(7587):541–5.
4. Saul FA, Normand BV le, Lema F, Bentley GA. Crystal structure of a dominant B-cell epitope from the pres2 region of hepatitis B virus in the form of an inserted peptide segment in maltodextrin-binding protein. *J Mol Biol*. 1998 Jul;280(2):185–92.
5. Pan D, Nakatsu T, Kato H. Crystal structure of peroxisomal targeting signal-2 bound to its receptor complex Pex7p–Pex21p. *Nat Struct Mol Biol*. 2013 Aug 30;20(8):987–93.
6. Zanier K, Charbonnier S, Sidi AOMO, McEwen AG, Ferrario MG, Poussin-Courmontagne P, et al. Structural Basis for Hijacking of Cellular LxxLL Motifs by Papillomavirus E6 Oncoproteins. *Science* (1979). 2013 Feb 8;339(6120):694–8.
7. Zhang Y, Gao X, Michael Garavito R. Structural analysis of the intracellular domain of (pro)renin receptor fused to maltose-binding protein. *Biochem Biophys Res Commun*. 2011 Apr;407(4):674–9.
8. Zhao M, Cascio D, Sawaya MR, Eisenberg D. Structures of segments of α -synuclein fused to maltose-binding protein suggest intermediate states during amyloid formation. *Protein Science*. 2011 Jun;20(6):996–1004.
9. Wiltzius JJW, Sievers SA, Sawaya MR, Eisenberg D. Atomic structures of IAPP (amylin) fusions suggest a mechanism for fibrillation and the role of insulin in the process. *Protein Science*. 2009 Jul;18(7):1521–30.
10. Srivastava D, Gakhar L, Artemyev NO. Structural underpinnings of Ric8A function as a G-protein α -subunit chaperone and guanine-nucleotide exchange factor. *Nat Commun*. 2019 Dec 12;10(1):3084.
11. Momin AA, Hameed UFS, Arold ST. Passenger sequences can promote interlaced dimers in a common variant of the maltose-binding protein. *Sci Rep*. 2019 Dec 1;9(1):1–9.
12. Mehr A, Henneberg F, Chari A, Görlich D, Huyton T. The copper(II)-binding tripeptide GHK, a valuable crystallization and phasing tag for macromolecular crystallography. *Acta Crystallogr D Struct Biol*. 2020 Dec 1;76(12):1222–32.
13. Du W, Josephrajan A, Adhikary S, Bowles T, Bielinsky AK, Eichman BF. Mcm10 Self-Association Is Mediated by an N-Terminal Coiled-Coil Domain. *PLoS One*. 2013 Jul 23;8(7):e70518.

14. Gupta K, Martin R, Sharp R, Sarachan KL, Ninan NS, Van Duyne GD. Oligomeric Properties of Survival Motor Neuron·Gemin2 Complexes. *Journal of Biological Chemistry*. 2015 Aug;290(33):20185–99.
15. Martin R, Gupta K, Ninan NS, Perry K, Van Duyne GD. The Survival Motor Neuron Protein Forms Soluble Glycine Zipper Oligomers. *Structure*. 2012 Nov;20(11):1929–39.
16. Mendillo ML, Putnam CD, Kolodner RD. Escherichia coli MutS Tetramerization Domain Structure Reveals That Stable Dimers but Not Tetramers Are Essential for DNA Mismatch Repair in Vivo. *Journal of Biological Chemistry*. 2007 Jun;282(22):16345–54.
17. Kruse T, Hansen JL, Dahl K, Schäffer L, Sensfuss U, Poulsen C, et al. Development of Cagrilintide, a Long-Acting Amylin Analogue. *J Med Chem*. 2021 Aug 12;64(15):11183–94.
18. Wu HD, Kikuchi M, Dagliyan O, Aragaki AK, Nakamura H, Dokholyan N V., et al. Rational design and implementation of a chemically inducible heterotrimerization system. *Nat Methods*. 2020 Sep 3;17(9):928–36.
19. Hanazono Y, Takeda K, Miki K. Co-translational folding of α -helical proteins: structural studies of intermediate-length variants of the λ repressor. *FEBS Open Bio*. 2018 Aug 27;8(8):1312–21.
20. Yegambaram K, Bulloch EMM, Kingston RL. Protein domain definition should allow for conditional disorder. *Protein Science*. 2013 Nov;22(11):1502–18.
21. Zhemkov VA, Kulminskaya AA, Bezprozvanny IB, Kim M. The 2.2-Angstrom resolution crystal structure of the carboxy-terminal region of ataxin-3. *FEBS Open Bio*. 2016 Mar 18;6(3):168–78.
22. Ke A, Wolberger C. Insights into binding cooperativity of MATa1/MATalpha2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. *Protein Science*. 2003 Feb 1;12(2):306–12.
23. Webby MN, Herr N, Bulloch EMM, Schmitz M, Keown JR, Goldstone DC, et al. Structural Analysis of the Menangle Virus P Protein Reveals a Soft Boundary between Ordered and Disordered Regions. *Viruses*. 2021 Aug 31;13(9):1737.
24. Matsuoka R, Shimada A, Komuro Y, Sugita Y, Kohda D. Rational design of crystal contact-free space in protein crystals for analyzing spatial distribution of motions within protein molecules. *Protein Science*. 2016 Mar;25(3):754–68.
25. Chang TH, Hsieh FL, Smallwood PM, Gabelli SB, Nathans J. Structure of the RECK CC domain, an evolutionary anomaly. *Proceedings of the National Academy of Sciences*. 2020 Jun 30;117(26):15104–11.
26. Yang H, Wang J, Jia X, McNatt MW, Zang T, Pan B, et al. Structural insight into the mechanisms of enveloped virus tethering by tetherin. *Proceedings of the National Academy of Sciences*. 2010 Oct 26;107(43):18428–32.
27. Kim BW, Jung YO, Kim MK, Kwon DH, Park SH, Kim JH, et al. ACCORD: an assessment tool to determine the orientation of homodimeric coiled-coils. *Sci Rep*. 2017 Apr 7;7(1):43318.

28. Priyanka A, Solanki V, Parkesh R, Thakur KG. Crystal structure of the N-terminal domain of human SIRT7 reveals a three-helical domain architecture. *Proteins: Structure, Function, and Bioinformatics*. 2016 Oct;84(10):1558–63.
29. Li T, Stayrook SE, Tsutsui Y, Zhang J, Wang Y, Li H, et al. Structural basis for ligand reception by anaplastic lymphoma kinase. *Nature*. 2021 Dec 2;600(7887):148–52.
30. Miliara X, Garnett JA, Tatsuta T, Abid Ali F, Baldie H, Pérez-Dorado I, et al. Structural insight into the TRIAP1/PRELI -like domain family of mitochondrial phospholipid transfer complexes. *EMBO Rep*. 2015 Jul 13;16(7):824–35.
31. Li T, Zhao Q, Yang X, Chen C, Yang K, Wu C, et al. Structural insight into the Zika virus capsid encapsulating the viral genome. *Cell Res*. 2018 Apr 21;28(4):497–9.
32. Huang B, Xu Y, Hu X, Liu Y, Liao S, Zhang J, et al. A backbone-centred energy function of neural networks for protein design. *Nature*. 2022 Feb 17;602(7897):523–8.
33. Kim M. Beta conformation of polyglutamine track revealed by a crystal structure of Huntingtin N-terminal region with insertion of three histidine residues. *Prion*. 2013 May 27;7(3):221–8.
34. Yin X, Li W, Ma H, Zeng W, Peng C, Li Y, et al. Crystal structure and activation mechanism of DR3 death domain. *FEBS J*. 2019 Jul 11;286(13):2593–610.
35. Jin T, Curry J, Smith P, Jiang J, Xiao TS. Structure of the NLRP1 caspase recruitment domain suggests potential mechanisms for its association with procaspase-1. *Proteins: Structure, Function, and Bioinformatics*. 2013 Jul;81(7):1266–70.
36. Adikesavan NV, Mahmood SS, Stanley N, Xu Z, Wu N, Thibonnier M, et al. A C-terminal segment of the V 1R vasopressin receptor is unstructured in the crystal structure of its chimera with the maltose-binding protein. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2005 Apr 1;61(4):341–5.
37. Nosaki S, Miyakawa T, Xu Y, Nakamura A, Hirabayashi K, Asami T, et al. Structural basis for brassinosteroid response by BIL1/BZR1. *Nat Plants*. 2018 Oct 1;4(10):771–6.
38. Xu Z, Zhou Y, Liu M, Ma H, Sun L, Zahid A, et al. Homotypic CARD-CARD interaction is critical for the activation of NLRP1 inflammasome. *Cell Death Dis*. 2021 Jan 11;12(1):57.
39. Sun T, Gauthier SY, Campbell RL, Davies PL. Revealing Surface Waters on an Antifreeze Protein by Fusion Protein Crystallography Combined with Molecular Dynamic Simulations. *J Phys Chem B*. 2015 Oct 8;119(40):12808–15.
40. Jin T, Huang M, Smith P, Jiang J, Xiao TS. The structure of the CARD8 caspase-recruitment domain suggests its association with the FIIND domain and procaspases through adjacent surfaces. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2013 May 1;69(5):482–7.
41. Kawano S, Yamano K, Naoé M, Momose T, Terao K, Nishikawa S ichi, et al. Structural basis of yeast Tim40/Mia40 as an oxidative translocator in the mitochondrial intermembrane space. *Proceedings of the National Academy of Sciences*. 2009 Aug 25;106(34):14403–7.
42. Guo TW, Bartesaghi A, Yang H, Falconieri V, Rao P, Merk A, et al. Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell*. 2017 Oct;171(2):414-426.e12.

43. Potter JA, Randall RE, Taylor GL. Crystal structure of human IPS-1/MAVS/VISA/Cardif caspase activation recruitment domain. *BMC Struct Biol.* 2008;8(1):11.
44. Kobe B, Center RJ, Kemp BE, Pountourios P. Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proceedings of the National Academy of Sciences.* 1999 Apr 13;96(8):4319–24.
45. Shen C, Lu A, Xie WJ, Ruan J, Negro R, Egelman EH, et al. Molecular mechanism for NLRP6 inflammasome assembly and activation. *Proceedings of the National Academy of Sciences.* 2019 Feb 5;116(6):2052–7.
46. Lamb D, Schüttelkopf AW, van Aalten DMF, Brighty DW. Charge-Surrounded Pockets and Electrostatic Interactions with Small Ions Modulate the Activity of Retroviral Fusion Proteins. *PLoS Pathog.* 2011 Feb 3;7(2):e1001268.
47. Gilbreth RN, Esaki K, Koide A, Sidhu SS, Koide S. A Dominant Conformational Role for Amino Acid Diversity in Minimalist Protein–Protein Interfaces. *J Mol Biol.* 2008 Aug;381(2):407–18.
48. Koide A, Gilbreth RN, Esaki K, Tereshko V, Koide S. High-affinity single-domain binding proteins with a binary-code interface. *Proceedings of the National Academy of Sciences.* 2007 Apr 17;104(16):6632–7.
49. Kumar S, Pioszak A, Zhang C, Swaminathan K, Xu HE. Crystal Structure of the PAC1R Extracellular Domain Unifies a Consensus Fold for Hormone Recognition by Class B G-Protein Coupled Receptors. *PLoS One.* 2011 May 19;6(5):e19682.
50. Jin T, Chuenchor W, Jiang J, Cheng J, Li Y, Fang K, et al. Design of an expression system to enhance MBP-mediated crystallization. *Sci Rep.* 2017 Jan 23;7(1):40991.
51. Wagner JM, Chan S, Evans TJ, Kahng S, Kim J, Arbing MA, et al. Structures of EccB1 and EccD1 from the core complex of the mycobacterial ESX-1 type VII secretion system. *BMC Struct Biol.* 2016 Dec 27;16(1):5.
52. Liu M, Zhou K, Xu Z, Ma H, Cao X, Yin X, et al. Crystal structure of caspase-11 CARD provides insights into caspase-11 activation. *Cell Discov.* 2020 Oct 13;6(1):70.
53. Pioszak AA, Parker NR, Suino-Powell K, Xu HE. Molecular Recognition of Corticotropin-releasing Factor by Its G-protein-coupled Receptor CRFR1. *Journal of Biological Chemistry.* 2008 Nov;283(47):32900–12.
54. Zhang F, Wei Q, Tong H, Xu D, Wang W, Ran T. Crystal structure of MBP-PigG fusion protein and the essential function of PigG in the prodigiosin biosynthetic pathway in *Serratia marcescens* FS14. *Int J Biol Macromol.* 2017 Jun;99:394–400.
55. Chao JA, Prasad GS, White SA, Stout CD, Williamson JR. Inherent Protein Structural Flexibility at the RNA-binding Interface of L30e. *J Mol Biol.* 2003 Feb;326(4):999–1004.
56. Shen G, Ke J, Wang Z, Cheng Z, Gu X, Wei Y, et al. Structural basis of the Norrin-Frizzled 4 interaction. *Cell Res.* 2015 Sep 31;25(9):1078–81.

57. Chao JA, Williamson JR. Joint X-Ray and NMR Refinement of the Yeast L30e-mRNA Complex. *Structure*. 2004 Jul;12(7):1165–76.
58. Lee SM, Jeong Y, Simms J, Warner ML, Poyner DR, Chung KY, et al. Calcitonin Receptor N-Glycosylation Enhances Peptide Hormone Affinity by Controlling Receptor Dynamics. *J Mol Biol*. 2020 Mar;432(7):1996–2014.
59. Pellegrini E, Desfosses A, Wallmann A, Schulze WM, Rehbein K, Mas P, et al. RIP2 filament formation is required for NOD2 dependent NF- κ B signalling. *Nat Commun*. 2018 Oct 2;9(1):4043.
60. Ke J, Harikumar KG, Erice C, Chen C, Gu X, Wang L, et al. Structure and function of Norrin in assembly and activation of a Frizzled 4–Lrp5/6 complex. *Genes Dev*. 2013 Nov 1;27(21):2305–19.
61. Monné M, Han L, Schwend T, Burendahl S, Jovine L. Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature*. 2008 Dec 4;456(7222):653–7.
62. Liu Y, Manna A, Li R, Martin WE, Murphy RC, Cheung AL, et al. Crystal structure of the SarR protein from *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences*. 2001 Jun 5;98(12):6877–82.
63. Tong J, Yang H, Eom SH, Chun C, Im YJ. Structure of the GH1 domain of guanylate kinase-associated protein from *Rattus norvegicus*. *Biochem Biophys Res Commun*. 2014 Sep;452(1):130–5.
64. Bala S, Shinya S, Srivastava A, Ishikawa M, Shimada A, Kobayashi N, et al. Crystal contact-free conformation of an intrinsically flexible loop in protein crystal: Tim21 as the case study. *Biochimica et Biophysica Acta (BBA) - General Subjects*. 2020 Feb;1864(2):129418.
65. Raj I, Sadat Al Hosseini H, Dioguardi E, Nishimura K, Han L, Villa A, et al. Structural Basis of Egg Coat-Sperm Recognition at Fertilization. *Cell*. 2017 Jun;169(7):1315-1326.e17.
66. Port SA, Monecke T, Dickmanns A, Spillner C, Hofele R, Urlaub H, et al. Structural and Functional Characterization of CRM1-Nup214 Interactions Reveals Multiple FG-Binding Sites Involved in Nuclear Export. *Cell Rep*. 2015 Oct;13(4):690–702.
67. Byrne NJ, Lee AC, Kostas J, Reid JC, Partridge AT, So SS, et al. Development of a robust crystallization platform for immune receptor TREM2 using a crystallization chaperone strategy. *Protein Expr Purif*. 2021 Mar;179:105796.
68. Schumacher MA, Tonthat NK, Kwong SM, Chinnam N babu, Liu MA, Skurray RA, et al. Mechanism of staphylococcal multiresistance plasmid replication origin assembly by the RepA protein. *Proceedings of the National Academy of Sciences*. 2014 Jun 24;111(25):9121–6.
69. Liu Z, Wang Y, Yedidi RS, Brunzelle JS, Kovari IA, Sohi J, et al. Crystal structure of the extracellular domain of human myelin protein zero. *Proteins: Structure, Function, and Bioinformatics*. 2012 Jan;80(1):307–13.
70. Kwon H, Young PG, Squire CJ, Baker EN. Engineering a Lys-Asn isopeptide bond into an immunoglobulin-like protein domain enhances its stability. *Sci Rep*. 2017 Feb 16;7(1):42753.

71. Berry JL, Xu Y, Ward PN, Lea SM, Matthews SJ, Pelicic V. A Comparative Structure/Function Analysis of Two Type IV Pilin DNA Receptors Defines a Novel Mode of DNA Binding. *Structure*. 2016 Jun;24(6):926–34.
72. AhYoung AP, Koehl A, Vizcarra CL, Cascio D, Egea PF. Structure of a putative ClpS N-end rule adaptor protein from the malaria pathogen *Plasmodium falciparum*. *Protein Science*. 2016 Mar;25(3):689–701.
73. Piepenbrink KH, Lillehoj E, Harding CM, Labonte JW, Zuo X, Rapp CA, et al. Structural Diversity in the Type IV Pili of Multidrug-resistant *Acinetobacter*. *Journal of Biological Chemistry*. 2016 Oct;291(44):22924–35.
74. LaPorte SL, Forsyth CM, Cunningham BC, Miercke LJ, Akhavan D, Stroud RM. De novo design of an IL-4 antagonist and its structure at 1.9 Å. *Proceedings of the National Academy of Sciences*. 2005 Feb 8;102(6):1889–94.
75. Mueller GA, Gosavi RA, Pomés A, Wünschmann S, Moon AF, London RE, et al. Ara h 2: crystal structure and IgE binding distinguish two subpopulations of peanut allergic patients by epitope diversity. *Allergy*. 2011 Jul;66(7):878–85.
76. Ronish LA, Lillehoj E, Fields JK, Sundberg EJ, Piepenbrink KH. The structure of PilA from *Acinetobacter baumannii* AB5075 suggests a mechanism for functional specialization in *Acinetobacter* type IV pili. *Journal of Biological Chemistry*. 2019 Jan;294(1):218–30.
77. Watkins HA, Baker EN. Structural and Functional Characterization of an RNase HI Domain from the Bifunctional Protein Rv2228c from *Mycobacterium tuberculosis*. *J Bacteriol*. 2010 Jun;192(11):2878–86.
78. Piepenbrink KH, Maldarelli GA, Martinez de la Peña CF, Dingle TC, Mulvey GL, Lee A, et al. Structural and Evolutionary Analyses Show Unique Stabilization Strategies in the Type IV Pili of *Clostridium difficile*. *Structure*. 2015 Feb;23(2):385–96.
79. Song JJ, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, et al. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Mol Biol*. 2003 Dec 16;10(12):1026–32.
80. Khalili Yazdi A, Namjoshi S, Hackett J, Ghonaim N, Shilton BH. Characterization of a polypeptide-binding site in the DEAD Motor of the SecA ATPase. *FEBS Lett*. 2017 Oct;591(20):3378–90.
81. Zhang Y, Qiu S, Jia S, Xu D, Ran T, Wang W. Crystal structure of the sensor domain of BaeS from *Serratia marcescens* FS14. *Proteins: Structure, Function, and Bioinformatics*. 2017 Sep;85(9):1784–90.
82. Gumpena R, Lountos GT, Raran-Kurussi S, Tropea JE, Cherry S, Waugh DS. Crystal structure of the human dual specificity phosphatase 1 catalytic domain. *Protein Science*. 2018 Feb;27(2):561–7.
83. Marvin JS, Schreiter ER, Echevarría IM, Looger LL. A genetically encoded, high-signal-to-noise maltose sensor. *Proteins: Structure, Function, and Bioinformatics*. 2011 Nov 12;79(11):3025–36.

84. Conrady MC, Suarez I, Gogl G, Frecot DI, Bonhoure A, Kostmann C, et al. Structure of High-Risk Papillomavirus 31 E6 Oncogenic Protein and Characterization of E6/E6AP/p53 Complex Formation. *J Virol*. 2020 Dec 22;95(2).
85. Goldschen-Ohm MP, Klenchin VA, White DS, Cowgill JB, Cui Q, Goldsmith RH, et al. Structure and dynamics underlying elementary ligand binding events in human pacemaking channels. *Elife*. 2016 Nov 18;5(e20797):2–23.
86. Clifton MC, Dranow DM, Leed A, Fulroth B, Fairman JW, Abendroth J, et al. A Maltose-Binding Protein Fusion Construct Yields a Robust Crystallography Platform for MCL1. *PLoS One*. 2015 Apr 24;10(4):e0125010.
87. Fang C, D'Souza B, Thompson CF, Clifton MC, Fairman JW, Fulroth B, et al. Single Diastereomer of a Macrolactam Core Binds Specifically to Myeloid Cell Leukemia 1 (MCL1). *ACS Med Chem Lett*. 2014 Dec 11;5(12):1308–12.
88. Zhao B, Arnold AL, Coronel MA, Lee JH, Lee T, Olejniczak ET, et al. Understanding the Species Selectivity of Myeloid Cell Leukemia-1 (Mcl-1) Inhibitors. *Biochemistry*. 2018 Aug 14;57(32):4952–8.
89. Kotschy A, Szlavik Z, Murray J, Davidson J, Maragno AL, Le Toumelin-Braizat G, et al. The MCL1 inhibitor S63845 is tolerable and effective in diverse cancer models. *Nature*. 2016 Oct 19;538(7626):477–82.
90. Banjara S, Shimmon GL, Dixon LK, Netherton CL, Hinds MG, Kvensakul M. Crystal Structure of African Swine Fever Virus A179L with the Autophagy Regulator Beclin. *Viruses*. 2019 Aug 27;11(9):789.
91. Bian C, Xu C, Ruan J, Lee KK, Burke TL, Tempel W, et al. Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J*. 2011 Jul 20;30(14):2829–42.
92. Pioszak AA, Harikumar KG, Parker NR, Miller LJ, Xu HE. Dimeric Arrangement of the Parathyroid Hormone Receptor and a Structural Mechanism for Ligand-induced Dissociation. *Journal of Biological Chemistry*. 2010 Apr;285(16):12435–44.
93. Ting YT, Harris PWR, Batot G, Brimble MA, Baker EN, Young PG. Peptide binding to a bacterial signal peptidase visualized by peptide tethering and carrier-driven crystallization. *IUCrJ*. 2016 Jan 1;3(1):10–9.
94. Banjara S, D Sa J, Hinds MG, Kvensakul M. The structural basis of Bcl-2 mediated cell death regulation in hydra. *Biochemical Journal*. 2020 Sep 18;477(17):3287–97.
95. Pham P, Afif SA, Shimoda M, Maeda K, Sakaguchi N, Pedersen LC, et al. Structural analysis of the activation-induced deoxycytidine deaminase required in immunoglobulin diversification. *DNA Repair (Amst)*. 2016 Jul;43:48–56.
96. Pioszak AA, Parker NR, Gardella TJ, Xu HE. Structural Basis for Parathyroid Hormone-related Protein Binding to the Parathyroid Hormone Receptor and Design of Conformation-selective Peptides. *Journal of Biological Chemistry*. 2009 Oct;284(41):28382–91.
97. Qiao Q, Wang L, Meng FL, Hwang JK, Alt FW, Wu H. AID Recognizes Structured DNA for Class Switch Recombination. *Mol Cell*. 2017 Aug;67(3):361–373.e4.

98. Booe JM, Walker CS, Barwell J, Kuteyi G, Simms J, Jamaluddin MA, et al. Structural Basis for Receptor Activity-Modifying Protein-Dependent Selective Peptide Recognition by a G Protein-Coupled Receptor. *Mol Cell*. 2015 Jun;58(6):1040–52.
99. Jiang J, Chan H, Cash DD, Miracco EJ, Ogorzalek Loo RR, Upton HE, et al. Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions. *Science (1979)*. 2015 Oct 30;350(6260).
100. Chao KL, Kulakova L, Herzberg O. Gene polymorphism linked to increased asthma and IBD risk alters gasdermin-B structure, a sulfatide and phosphoinositide binding protein. *Proceedings of the National Academy of Sciences*. 2017 Feb 14;114(7).
101. Fu TM, Li Y, Lu A, Li Z, Vajjhala PR, Cruz AC, et al. Cryo-EM Structure of Caspase-8 Tandem DED Filament Reveals Assembly and Regulation Mechanisms of the Death-Inducing Signaling Complex. *Mol Cell*. 2016 Oct;64(2):236–50.
102. Foo ACY, Thompson PM, Chen SH, Jadi R, Lupo B, DeRose EF, et al. The mosquito protein AEG12 displays both cytolytic and antiviral properties via a common lipid transfer mechanism. *Proceedings of the National Academy of Sciences*. 2021 Mar 16;118(11).
103. Patrick AN, Cabrera JH, Smith AL, Chen XS, Ford HL, Zhao R. Structure-function analyses of the human SIX1–EYA2 complex reveal insights into metastasis and BOR syndrome. *Nat Struct Mol Biol*. 2013 Apr 24;20(4):447–53.
104. Wang X, Sun B, Xu M, Qiu S, Xu D, Ran T, et al. Crystal structure of the periplasmic domain of TssL, a key membrane component of Type VI secretion system. *Int J Biol Macromol*. 2018 Dec;120:1474–9.
105. Tan MHE, Zhou XE, Soon FF, Li X, Li J, Yong EL, et al. The Crystal Structure of the Orphan Nuclear Receptor NR2E3/PNR Ligand Binding Domain Reveals a Dimeric Auto-Repressed Conformation. *PLoS One*. 2013 Sep 12;8(9):e74359.
106. Wang X, Pernicone N, Pertz L, Hua D, Zhang T, Listovsky T, et al. REV7 has a dynamic adaptor region to accommodate small GTPase RAN/Shigella IpaB ligands, and its activity is regulated by the RanGTP/GDP switch. *Journal of Biological Chemistry*. 2019 Oct;294(43):15733–42.
107. Zhi X, Zhou XE, He Y, Searose-Xu K, Zhang CL, Tsai CC, et al. Structural basis for corepressor assembly by the orphan nuclear receptor TLX. *Genes Dev*. 2015 Feb 15;29(4):440–50.
108. Bucknell SJ, Ator MA, Brown AJH, Brown J, Cansfield AD, Cansfield JE, et al. Structure-Based Drug Discovery of N -((R)-3-(7-Methyl-1 H-indazol-5-yl)-1-oxo-1-(((S)-1-oxo-3-(piperidin-4-yl)-1-(4-(pyridin-4-yl)piperazin-1-yl)propan-2-yl)amino)propan-2-yl)-2'-oxo-1',2'-dihydrospiro[piperidine-4,4'-pyrido[2,3-d][1,3]oxazine]-1-carboxamide (HTL22562): A Calcitonin Gene-Related Peptide Receptor Antagonist for Acute Treatment of Migraine. *J Med Chem*. 2020 Jul 23;63(14):7906–20.
109. Mueller GA, Edwards LL, Aloor JJ, Fessler MB, Glesner J, Pomés A, et al. The structure of the dust mite allergen Der p 7 reveals similarities to innate immune proteins. *Journal of Allergy and Clinical Immunology*. 2010 Apr;125(4):909-917.e4.

110. Zhi X, Zhou XE, He Y, Zechner C, Suino-Powell KM, Kliewer SA, et al. Structural insights into gene repression by the orphan nuclear receptor SHP. *Proceedings of the National Academy of Sciences*. 2014 Jan 14;111(2):839–44.
111. Li Y, Zeng W, Li Y, Fan W, Ma H, Fan X, et al. Structure determination of the CAMP factor of *Streptococcus agalactiae* with the aid of an MBP tag and insights into membrane-surface attachment. *Acta Crystallogr D Struct Biol*. 2019;75:772–81.
112. Yang H, Ji X, Zhao G, Ning J, Zhao Q, Aiken C, et al. Structural insight into HIV-1 capsid recognition by rhesus TRIM5 α . *Proceedings of the National Academy of Sciences*. 2012 Nov 6;109(45):18372–7.
113. Park H, Boyington JC. The 1.5 Å Crystal Structure of Human Receptor for Advanced Glycation Endproducts (RAGE) Ectodomains Reveals Unique Features Determining Ligand Binding. *Journal of Biological Chemistry*. 2010 Dec;285(52):40762–70.
114. Wolfe AD, Li S, Goedderz C, Chen XS. The structure of APOBEC1 and insights into its RNA and DNA substrate selectivity. *NAR Cancer*. 2020 Dec 1;2(4).
115. Jung J, Bashiri G, Johnston JM, Brown AS, Ackerley DF, Baker EN. Crystal structure of the essential *Mycobacterium tuberculosis* phosphopantetheinyl transferase PptT, solved as a fusion protein with maltose binding protein. *J Struct Biol*. 2014 Dec;188(3):274–8.
116. Wang H, DeRose EF, London RE, Shears SB. IP6K structure and the molecular determinants of catalytic specificity in an inositol phosphate kinase family. *Nat Commun*. 2014 Sep 24;5(1):4178.
117. Ke W, Laurent AH, Armstrong MD, Chen Y, Smith WE, Liang J, et al. Structure of an Engineered β -Lactamase Maltose Binding Protein Fusion Protein: Insights into Heterotropic Allosteric Regulation. *PLoS One*. 2012 Jun 14;7(6):e39168.
118. Saito T, Bokhove M, Croci R, Zamora-Caballero S, Han L, Letarte M, et al. Structural Basis of the Human Endoglin-BMP9 Interaction: Insights into BMP Signaling and HHT1. *Cell Rep*. 2017 May;19(9):1917–28.
119. Kim HJ, Xu Y, Petri A, Vanhoorelbeke K, Crawley JTB, Emsley J. Crystal structure of ADAMTS13 CUB domains reveals their role in global latency. *Sci Adv*. 2021 Apr 16;7(16).
120. Pathak PK, Peng S, Meng X, Han Y, Zhang B, Zhang F, et al. Structure of a lipid-bound viral membrane assembly protein reveals a modality for enclosing the lipid bilayer. *Proceedings of the National Academy of Sciences*. 2018 Jul 3;115(27):7028–32.
121. Papillon JPN, Nakajima K, Adair CD, Hempel J, Jouk AO, Karki RG, et al. Discovery of Orally Active Inhibitors of Brahma Homolog (BRM)/SMARCA2 ATPase Activity for the Treatment of Brahma Related Gene 1 (BRG1)/SMARCA4-Mutant Cancers. *J Med Chem*. 2018 Nov 21;61(22):10155–72.
122. Bethea HN, Xu D, Liu J, Pedersen LC. Redirecting the substrate specificity of heparan sulfate 2- O -sulfotransferase by structurally guided mutagenesis. *Proceedings of the National Academy of Sciences*. 2008 Dec 2;105(48):18724–9.

123. Mohr G, Silas S, Stamos JL, Makarova KS, Markham LM, Yao J, et al. A Reverse Transcriptase-Cas1 Fusion Protein Contains a Cas6 Domain Required for Both CRISPR RNA Biogenesis and RNA Spacer Acquisition. *Mol Cell*. 2018 Nov;72(4):700-714.e8.
124. Bokhove M, Nishimura K, Brunati M, Han L, de Sanctis D, Rampoldi L, et al. A structured interdomain linker directs self-polymerization of human uromodulin. *Proceedings of the National Academy of Sciences*. 2016 Feb 9;113(6):1552-7.
125. Olal D, Daumke O. Structure of the Hantavirus Nucleoprotein Provides Insights into the Mechanism of RNA Encapsidation. *Cell Rep*. 2016 Mar;14(9):2092-9.
126. Xu Y, Moon AF, Xu S, Krahn JM, Liu J, Pedersen LC. Structure Based Substrate Specificity Analysis of Heparan Sulfate 6- O -Sulfotransferases. *ACS Chem Biol*. 2017 Jan 20;12(1):73-82.
127. Ullah H, Scappini EL, Moon AF, Williams LV, Armstrong DL, Pedersen LC. Structure of a signal transduction regulator, RACK1, from *Arabidopsis thaliana*. *Protein Science*. 2008 Oct;17(10):1771-80.
128. Hai Y, Christianson DW. Histone deacetylase 6 structure and molecular basis of catalysis and inhibition. *Nat Chem Biol*. 2016 Sep 25;12(9):741-7.
129. Arora K, Talje L, Asenjo AB, Andersen P, Atchia K, Joshi M, et al. KIF14 Binds Tightly to Microtubules and Adopts a Rigor-Like Conformation. *J Mol Biol*. 2014 Aug;426(17):2997-3015.
130. Delorme C, Joshi M, Allingham JS. Crystal structure of the *Candida albicans* Kar3 kinesin motor domain fused to maltose-binding protein. *Biochem Biophys Res Commun*. 2012 Nov;428(4):427-32.
131. Matsumoto S, Shimada A, Kohda D. Crystal structure of the C-terminal globular domain of the third paralog of the *Archaeoglobus fulgidus* oligosaccharyltransferases. *BMC Struct Biol*. 2013 Dec 1;13(1):11.
132. Baßler J, Paternoga H, Holdermann I, Thoms M, Granneman S, Barrio-Garcia C, et al. A network of assembly factors is involved in remodeling rRNA elements during preribosome maturation. *Journal of Cell Biology*. 2014 Nov 24;207(4):481-98.
133. Cherry AL, Finta C, Karlström M, Jin Q, Schwend T, Astorga-Wells J, et al. Structural basis of SUFU-GLI interaction in human Hedgehog signalling regulation. *Acta Crystallogr D Biol Crystallogr*. 2013 Dec 1;69(12):2563-79.
134. Yuan Z, VanderWielen BD, Giaimo BD, Pan L, Collins CE, Turkiewicz A, et al. Structural and Functional Studies of the RBPJ-SHARP Complex Reveal a Conserved Corepressor Binding Site. *Cell Rep*. 2019 Jan;26(4):845-854.e6.
135. Batyuk A, Wu Y, Honegger A, Heberling MM, Plückthun A. DARPin-Based Crystallization Chaperones Exploit Molecular Geometry as a Screening Dimension in Protein Crystallography. *J Mol Biol*. 2016 Apr;428(8):1574-88.
136. Huang DT, Hunt HW, Zhuang M, Ohi MD, Holton JM, Schulman BA. Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature*. 2007 Jan 14;445(7126):394-8.

137. Yuan Z, Praxenthaler H, Tabaja N, Torella R, Preiss A, Maier D, et al. Structure and Function of the Su(H)-Hairless Repressor Complex, the Major Antagonist of Notch Signaling in *Drosophila melanogaster*. *PLoS Biol.* 2016 Jul 12;14(7):e1002509.
138. Brunner JD, Jakob RP, Schulze T, Neldner Y, Moroni A, Thiel G, et al. Structural basis for ion selectivity in TMEM175 K⁺ channels. *Elife.* 2020 Apr 8;9(e53683.).
139. Li W, Zhang T, Sun M, Shi Y, Zhang XJ, Xu GL, et al. Molecular mechanism for vitamin C-derived C5-glyceryl-methylcytosine DNA modification catalyzed by algal TET homologue CMD1. *Nat Commun.* 2021 Feb 2;12(1):744.
140. Wallace BD, Roberts AB, Pollet RM, Ingle JD, Biernat KA, Pellock SJ, et al. Structure and Inhibition of Microbiome β -Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. *Chem Biol.* 2015 Sep;22(9):1238–49.
141. Jiang K, Zhang Y, Chen Z, Wu D, Cai J, Gao X. Structural and Functional Insights into the C-terminal Fragment of Insecticidal Vip3A Toxin of *Bacillus thuringiensis*. *Toxins (Basel).* 2020 Jul 5;12(7):438.
142. Suzuki N, Rohaim A, Kato R, Dikic I, Wakatsuki S, Kawasaki M. A novel mode of ubiquitin recognition by the ubiquitin-binding zinc finger domain of WRNIP1. *FEBS J.* 2016 Jun;283(11):2004–17.
143. Liu Y, Duan Z, Fang J, Zhang F, Xiao J, Zhang W. Cellular Synthesis and X-ray Crystal Structure of a Designed Protein Heterocatenane. *Angewandte Chemie International Edition.* 2020 Sep 7;59(37):16122–7.
144. Suzuki N, Hiraki M, Yamada Y, Matsugaki N, Igarashi N, Kato R, et al. Crystallization of small proteins assisted by green fluorescent protein. *Acta Crystallogr D Biol Crystallogr.* 2010 Oct;66(10):1059–66.
145. Mou Y, Huang PS, Thomas LM, Mayo SL. Using Molecular Dynamics Simulations as an Aid in the Prediction of Domain Swapping of Computationally Designed Protein Variants. *J Mol Biol.* 2015 Aug;427(16):2697–706.
146. Czabotar PE, Westphal D, Dewson G, Ma S, Hockings C, Fairlie WD, et al. Bax Crystal Structures Reveal How BH3 Domains Activate Bax and Nucleate Its Oligomerization to Induce Apoptosis. *Cell.* 2013 Jan;152(3):519–31.
147. Mandal T, Shin S, Aluvila S, Chen HC, Grieve C, Choe JY, et al. Assembly of Bak homodimers into higher order homooligomers in the mitochondrial apoptotic pore. *Sci Rep.* 2016 Aug 4;6(1):30763.
148. Liu S, Li S, Yang Y, Li W. Termini restraining of small membrane proteins enables structure determination at near-atomic resolution. *Sci Adv.* 2020 Dec 18;6(51).
149. Brouwer JM, Westphal D, Dewson G, Robin AY, Uren RT, Bartolo R, et al. Bak Core and Latch Domains Separate during Activation, and Freed Core Domains Form Symmetric Homodimers. *Mol Cell.* 2014 Sep;55(6):938–46.
150. Arpino JAJ, Czapinska H, Piasecka A, Edwards WR, Barker P, Gajda MJ, et al. Structural Basis for Efficient Chromophore Communication and Energy Transfer in a Constructed Didomain Protein Scaffold. *J Am Chem Soc.* 2012 Aug 22;134(33):13632–40.

151. Eshaghi M, Sun G, Grüter A, Lim CL, Chee YC, Jung G, et al. Rational Structure-Based Design of Bright GFP-Based Complexes with Tunable Dimerization. *Angewandte Chemie International Edition*. 2015 Nov 16;54(47):13952–6.
152. Håkansson KO, Winther JR. Structure of glutaredoxin Grx1p C30S mutant from yeast. *Acta Crystallogr D Biol Crystallogr*. 2007 Mar 1;63(3):288–94.
153. Liu S, Li S, Shen G, Sukumar N, Krezel AM, Li W. Structural basis of antagonizing the vitamin K catalytic cycle for anticoagulation. *Science* (1979). 2021 Jan;371(6524).
154. Wu SY, Wen Y, Serre NBC, Laursen CCH, Dietz AG, Taylor BR, et al. A sensitive and specific genetically-encoded potassium ion biosensor for in vivo applications across the tree of life. *PLoS Biol*. 2022 Sep 6;20(9):e3001772.
155. Grinkevich P, Sinha D, Iermak I, Guzanova A, Weiserova M, Ludwig J, et al. Crystal structure of a novel domain of the motor subunit of the Type I restriction enzyme EcoR124 involved in complex assembly and DNA binding. *Journal of Biological Chemistry*. 2018 Sep;293(39):15043–54.
156. Barykina N V., Sotskov VP, Gruzdeva AM, Wu YK, Portugues R, Subach OM, et al. FGCaMP7, an Improved Version of Fungi-Based Ratiometric Calcium Indicator for In Vivo Visualization of Neuronal Activity. *Int J Mol Sci*. 2020 Apr 24;21(8):3012.
157. Mueller GA, Pedersen LC, Lih FB, Glesner J, Moon AF, Chapman MD, et al. The novel structure of the cockroach allergen Bla g 1 has implications for allergenicity and exposure assessment. *Journal of Allergy and Clinical Immunology*. 2013 Dec;132(6):1420-1426.e9.
158. Collu G, Bierig T, Krebs AS, Engilberge S, Varma N, Guixà-González R, et al. Chimeric single α -helical domains as rigid fusion protein connections for protein nanotechnology and structural biology. *Structure*. 2022 Jan;30(1):95-106.e7.
159. Zhao Y, Shen Y, Wen Y, Campbell RE. High-Performance Intensiometric Direct- and Inverse-Response Genetically Encoded Biosensors for Citrate. *ACS Cent Sci*. 2020 Aug 26;6(8):1441–50.
160. Unger EK, Keller JP, Altermatt M, Liang R, Matsui A, Dong C, et al. Directed Evolution of a Selective and Sensitive Serotonin Sensor via Machine Learning. *Cell*. 2020 Dec;183(7):1986-2002.e26.
161. Marvin JS, Shimoda Y, Magloire V, Leite M, Kawashima T, Jensen TP, et al. A genetically encoded fluorescent sensor for in vivo imaging of GABA. *Nat Methods*. 2019 Aug 15;16(8):763–70.
162. Trigo-Mourino P, Thestrup T, Griesbeck O, Griesinger C, Becker S. Dynamic tuning of FRET in a green fluorescent protein biosensor. *Sci Adv*. 2019 Aug 2;5(8).
163. Moon AF, Krahn JM, Lu X, Cuneo MJ, Pedersen LC. Structural characterization of the virulence factor Sda1 nuclease from *Streptococcus pyogenes*. *Nucleic Acids Res*. 2016 May 5;44(8):3946–57.
164. Wang Q, Shui B, Kotlikoff MI, Sondermann H. Structural Basis for Calcium Sensing by GCaMP2. *Structure*. 2008 Dec;16(12):1817–27.

165. Nasu Y, Murphy-Royal C, Wen Y, Haidey JN, Molina RS, Aggarwal A, et al. A genetically encoded fluorescent biosensor for extracellular l-lactate. *Nat Commun.* 2021 Dec 6;12(1):7058.
166. Akerboom J, Rivera JDV, Guilbe MMR, Malavé ECA, Hernandez HH, Tian L, et al. Crystal Structures of the GCaMP Calcium Sensor Reveal the Mechanism of Fluorescence Signal Change and Aid Rational Design. *Journal of Biological Chemistry.* 2009 Mar;284(10):6455–64.
167. Wang L, Wu C, Peng W, Zhou Z, Zeng J, Li X, et al. A high-performance genetically encoded fluorescent indicator for in vivo cAMP imaging. *Nat Commun.* 2022 Sep 12;13(1):5363.
168. Akerboom J, Chen TW, Wardill TJ, Tian L, Marvin JS, Mutlu S, et al. Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *The Journal of Neuroscience.* 2012 Oct 3;32(40):13819–40.
169. Kostyuk AI, Tossounian MA, Panova AS, Thauvin M, Raevskii RI, Ezeriņa D, et al. Hypocrates is a genetically encoded fluorescent biosensor for (pseudo)hypohalous acids and their derivatives. *Nat Commun.* 2022 Jan 10;13(1):171.
170. Shivange A V., Borden PM, Muthusamy AK, Nichols AL, Bera K, Bao H, et al. Determining the pharmacokinetics of nicotinic drugs in the endoplasmic reticulum using biosensors. *Journal of General Physiology.* 2019 Jun 3;151(6):738–57.
171. Nichols AL, Blumenfeld Z, Fan C, Luebbert L, Blom AE, Cohen BN, et al. Correction: Fluorescence activation mechanism and imaging of drug permeation with new sensors for smoking-cessation ligands. *Elife.* 2022 Dec 15;11.
172. Chow D, Srikakulam R, Chen Y, Winkelmann DA. Folding of the Striated Muscle Myosin Motor Domain. *Journal of Biological Chemistry.* 2002 Sep;277(39):36799–807.
173. Winkelmann DA, Forgacs E, Miller MT, Stock AM. Structural basis for drug-induced allosteric changes to human β -cardiac myosin motor activity. *Nat Commun.* 2015 Aug 6;6(1):7974.
174. Ménétrey J, Isabet T, Ropars V, Mukherjea M, Pylypenko O, Liu X, et al. Processive Steps in the Reverse Direction Require Uncoupling of the Lead Head Lever Arm of Myosin VI. *Mol Cell.* 2012 Oct;48(1):75–86.
175. Schmidt H, Zalyte R, Urnavicius L, Carter AP. Structure of human cytoplasmic dynein-2 primed for its power stroke. *Nature.* 2015 Feb 1;518(7539):435–8.
176. Freudenthal BD, Brogie JE, Gakhar L, Kondratik CM, Washington MT. Crystal Structure of SUMO-Modified Proliferating Cell Nuclear Antigen. *J Mol Biol.* 2011 Feb;406(1):9–17.
177. Schalch T, Job G, Shanker S, Partridge JF, Joshua-Tor L. The Chp1–Tas3 core is a multifunctional platform critical for gene silencing by RITS. *Nat Struct Mol Biol.* 2011 Dec 13;18(12):1351–7.
178. Begley DW, Fox D, Jenner D, Juli C, Pierce PG, Abendroth J, et al. A Structural Biology Approach Enables the Development of Antimicrobials Targeting Bacterial Immunophilins. *Antimicrob Agents Chemother.* 2014 Mar;58(3):1458–67.

179. Clifton MC, Kirchdoerfer RN, Atkins K, Abendroth J, Raymond A, Grice R, et al. Structure of the Reston ebolavirus VP30 C-terminal domain. *Acta Crystallogr F Struct Biol Commun.* 2014 Apr 1;70(4):457–60.
180. Hu C, Inoue H, Sun W, Takeshita Y, Huang Y, Xu Y, et al. Structural insights into chromosome attachment to the nuclear envelope by an inner nuclear membrane protein Bqt4 in fission yeast. *Nucleic Acids Res.* 2019 Feb 20;47(3):1573–84.
181. Pan BS, Perera SA, Piesvaux JA, Presland JP, Schroeder GK, Cumming JN, et al. An orally available non-nucleotide STING agonist with antitumor activity. *Science (1979).* 2020 Aug 21;369(6506).
182. Cross AR, Roy S, Vivoli Vega M, Rejzek M, Nepogodiev SA, Cliff M, et al. Spinning sugars in antigen biosynthesis: characterization of the *Coxiella burnetii* and *Streptomyces griseus* TDP-sugar epimerases. *Journal of Biological Chemistry.* 2022 May;298(5):101903.
183. Maker A, Bolejack M, Schecterson L, Hammerson B, Abendroth J, Edwards TE, et al. Regulation of multiple dimeric states of E-cadherin by adhesion activating antibodies revealed through Cryo-EM and X-ray crystallography. *PNAS Nexus.* 2022 Sep 1;1(4).
184. Georgakopoulos ND, Talapatra SK, Gatliff J, Kozielski F, Wells G. Modified Peptide Inhibitors of the Keap1–Nrf2 Protein–Protein Interaction Incorporating Unnatural Amino Acids. *ChemBioChem.* 2018 Sep 4;19(17):1810–6.
185. Wynn RM, Li J, Brautigam CA, Chuang JL, Chuang DT. Structural and Biochemical Characterization of Human Mitochondrial Branched-chain α -Ketoacid Dehydrogenase Phosphatase. *Journal of Biological Chemistry.* 2012 Mar;287(12):9178–92.
186. Albert L, Xu J, Wan R, Srinivasan V, Dou Y, Vázquez O. Controlled inhibition of methyltransferases using photoswitchable peptidomimetics: towards an epigenetic regulation of leukemia. *Chem Sci.* 2017;8(6):4612–8.
187. Streich Jr FC, Lima CD. Capturing a substrate in an activated RING E3/E2–SUMO complex. *Nature.* 2016 Aug 18;536(7616):304–8.
188. Bentley AA, Merkulov SM, Peng Y, Rozmarynowycz R, Qi X, Pusztai-Carey M, et al. Chimeric Glutathione S-Transferases Containing Inserts of Kininogen Peptides. *Journal of Biological Chemistry.* 2012 Jun;287(26):22142–50.
189. Lim K, Ho JX, Keeling K, Gilliland GL, Ji X, Rüker F, et al. Three-Dimensional structure of *Schistosoma japonicum* glutathione S-transferase fused with a six-amino acid conserved neutralizing epitope of gp41 from hiv. *Protein Science.* 1994 Dec;3(12):2233–44.
190. Ware S, Anderson WF, Donahue JP, Hawiger J. Structure of the fibrinogen γ -chain integrin binding and factor XIIIa cross-linking sites obtained through carrier protein driven crystallization. *Protein Science.* 2008 Dec 31;8(12):2663–71.
191. Zhang Z, Devarajan P, Dorfman AL, Morrow JS. Structure of the Ankyrin-binding Domain of α -Na,K-ATPase. *Journal of Biological Chemistry.* 1998 Jul;273(30):18681–4.
192. Tang L, Guo B, Javed A, Choi JY, Hiebert S, Lian JB, et al. Crystal Structure of the Nuclear Matrix Targeting Signal of the Transcription Factor Acute Myelogenous Leukemia-1/Polyoma

- Enhancer-binding Protein 2 α B/Core Binding Factor α 2. *Journal of Biological Chemistry*. 1999 Nov;274(47):33580–6.
193. Detry S, Andries J, Bloch Y, Gabay C, Clancy DM, Savvides SN. Structural basis of human IL-18 sequestration by the decoy receptor IL-18 binding protein in inflammation and tumor immunity. *Journal of Biological Chemistry*. 2022 May;298(5):101908.
 194. Knight ML, Fan H, Bauer DL V., Grimes JM, Fodor E, Keown JR. Structure of an H3N2 influenza virus nucleoprotein. *Acta Crystallogr F Struct Biol Commun*. 2021 Jul 1;77(7):208–14.
 195. Schmidt H, Gleave ES, Carter AP. Insights into dynein motor domain function from a 3.3-Å crystal structure. *Nat Struct Mol Biol*. 2012 May 14;19(5):492–7.
 196. Pabis M, Popowicz GM, Stehle R, Fernández-Ramos D, Asami S, Warner L, et al. HuR biological function involves RRM3-mediated dimerization and RNA binding by all three RRMs. *Nucleic Acids Res*. 2019 Jan 25;47(2):1011–29.
 197. Corsini L, Hothorn M, Scheffzek K, Sattler M, Stier G. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Science*. 2008 Dec;17(12):2070–9.
 198. Wazir S, Maksimainen MM, Lehtiö L. Multiple crystal forms of human MacroD2. *Acta Crystallogr F Struct Biol Commun*. 2020 Oct 1;76(10):477–82.
 199. Wander R, Kaminski AM, Wang Z, Stancanelli E, Xu Y, Pagadala V, et al. Structural and Substrate Specificity Analysis of 3- O-Sulfotransferase Isoform 5 to Synthesize Heparan Sulfate. *ACS Catal*. 2021 Dec 17;11(24):14956–66.
 200. Lee Y, Warne T, Nehmé R, Pandey S, Dwivedi-Agnihotri H, Chaturvedi M, et al. Molecular basis of β -arrestin coupling to formoterol-bound β 1-adrenoceptor. *Nature*. 2020 Jul 30;583(7818):862–6.
 201. Warne T, Edwards PC, Doré AS, Leslie AGW, Tate CG. Molecular basis for high-affinity agonist binding in GPCRs. *Science (1979)*. 2019 May 24;364(6442):775–8.
 202. Santos CR, Polo CC, Costa MCMF, Nascimento AFZ, Meza AN, Cota J, et al. Mechanistic Strategies for Catalysis Adopted by Evolutionary Distinct Family 43 Arabinanases. *Journal of Biological Chemistry*. 2014 Mar;289(11):7362–73.



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS



Instituto de Investigación en Ciencias Básicas y Aplicadas



Cuernavaca, Mor., a 12 de marzo de 2024

DRA. LINA ANDREA RIVILLAS ACEVEDO
COORDINADORA DEL POSGRADO EN CIENCIAS

PRESENTE

Atendiendo a la solicitud para emitir DICTAMEN sobre la revisión de la tesis titulada: Determinación de estructuras cristalográficas mediante el uso de proteínas de fusión, que presenta la alumna María Cristina Cardona Echavarría (10033392) para obtener el título de Doctorado en Ciencias.

Director de tesis: Dr. Enrique Rudiño Piñera

Unidad Académica: Instituto de Investigación en Ciencias Básicas y Aplicadas (IICBA)

Nos permitimos informarle que nuestro voto es:

NOMBRE	DICTAMEN	FIRMA
Dra. Carmen Nina Pastor Colón CIDC-UAEM	APROBADO	
Dr. Rodrigo Said Razo Hernández CIDC-UAEM	APROBADO	
Dra. Adela Rodríguez Romero Instituto de Química-UNAM	APROBADO	
Dr. Cesar Millán Pacheco Facultad de farmacia-UAEM	APROBADO	
Dra. Lina Andrea Rivillas Acevedo CIDC-UAEM	APROBADO	
Dr. Carlos Daniel Amero Tello CIQ-UAEM	APROBADO	
Dr. Enrique Rudiño Piñera IBT-UNAM	APROBADO	

Atentamente

Por una humanidad culta



DRA. LINA ANDREA RIVILLAS ACEVEDO
Coordinadora del Posgrado en Ciencias
(FIRMA ELECTRÓNICA)

C.e.p. Archivo.





UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

Sello electrónico

CARMEN NINA PASTOR COLON | Fecha:2024-04-25 20:06:14 | FIRMANTE

KjmejfzMNwyxW9qbQffhpK5x87sHcxNmAZ1z+RYQoUJPrILrfiHb4hN2pYkg0AHIQ4rcnrot4Rm6/Oc6V7MIMxTDoVYERyh1wQDnpg5zivGZAp5fSG2b5AQRlqH4qeMuF+t65+9083wesJJ3MOaMitnlxigq7GtMhHJHu66SJOJIPckT0WINV4dGz5MCRwEvvw62BY75H4RWnaY755ZRrh1vNw0qXrFepBQ+wRYUNWX5CfcCu+ujWHki3LV6/9DDyBbII9CIttMVebwOGyWSExtoryRACkeWRqOmmlQcPuLI/kGoOOckRw4PxxC44ry+njinOzgVBrXPufdiFMqkxzQ==

ENRIQUE RUDIÑO PIÑERA | Fecha:2024-04-25 20:06:25 | FIRMANTE

YvDNCNU2EdOEWgJ5CMZzmo2R/lv83GgHhJsrUFFFijXw5ohybSDkN/ZRCQV+0qi86yQLeibQzt148p90TqKXOsa0SWL2KHypMiQJY0YU0dLKSngkIS0EO3cekJ6fTp7VGNrYDAdNTIpe7QApe53muKGAj9/dyYSRyWKLyJFW0viiKcvGIN9XCKGQbTg/QRtmXeWqysuiHj3mFLncP/7vpg2bvxI1SwF5OpC1fz9le71mh9YHGDoYlocdBTPU6qRNhTdoLo4cM0vHscYqDxe5RbLB57KQNguyimtkW/bZTC1y3O4hSx+NC3+yaY9DZDNDATWgRhs2ZoGqfX51+m3crw==

RODRIGO SAID RAZO HERNANDEZ | Fecha:2024-04-25 21:46:42 | FIRMANTE

S3TetVKRmfiZpjigUVE9u+X/ANgimI5qYPeU23sYeAed4XMTaUhnK10oeyL/pTWVZySjkMT5a3iAHF3Whx/RrA1dUdv2B3VaB87na6aucgNLZJs/2bWqggHsF0gy+swlK2VBUeZhaJpcvVHZKhH7AFupD/2FJD15VL3e8ntsUkTciLXOFJu/wVOsACSSY6EQTLdZTbFvRxUJCfSE82rb6H4PEXJ047LSGvR6Kujr3sOkjff+NcRPg11FhTdz9X2Z+GbtZPdPOq3oYS+T6EY3E8znU3PjGyXZp0SP5j5odo6yEBslhkBEQ/u27CmH3KNEqTKlo99dPvCoHeCN6A==

CARLOS DANIEL AMERO TELLO | Fecha:2024-04-26 10:08:00 | FIRMANTE

JzqOIRdrCdCzHUp+pNEIO8Zij5pYwMeEA1p1ncJ5/RwFZxxMgfvu/XTfppS5zN8QAi5quQbmDLPFMTTRKHn0ZZ0suEvP8A+94DIuINA519FF+DWx3cp3wlSLQXpZ/JmKlfiHuZe1o1vtcD5pBdExge0pssgQ4uiD77GepqU84zU547OVHTNh97QQ8rvxg8gmryJfZ7I/IA+Vj4ASpuNZ7xeSnZHBcK4Np44gNLPP8GzhONKDEZbYhZRUQ0YkuTO/jBS0XaxZ0rLFB8J2UPQrurwpTsh1EFiQzACGR2XCQad7DOO6VTK1yf9+NuBILaMx76e2AmrNPIXr1EuW7EsAJNg==

LINA ANDREA RIVILLAS ACEVEDO | Fecha:2024-04-26 10:52:48 | FIRMANTE

epPURKNMxEoG9Uq1Bs5vXSS0WskGLafbq+ZclZ/ulhn11cKrRa3nCyrbvM7UpW6C8hA6fRgNXRjhm/mZl4PVAwDe6S9nLzbMjPsm6WvX4zz0UULF/ZYWMPy59dYSadzJE9bhTh667reCH+MK/RhELP68Y1KRF2ExmJ6O9uteE5nJUn619ht5QZMc3MzH/V8dvcimu7aGQWWhb4NxcZN5N+HUo1KbUT6PpuYxkt0dFupmdnw9EaadsIHq2BIUK3Wu2op8bxbhR0Emt1v46bWS7Im0f1/OEphnSwuvPjrVacUyeO8TIAAARPPXlgi/eSkkiH8+emMwPV3+tM4fGAb2f2w==

ADELA RODRÍGUEZ ROMERO | Fecha:2024-04-26 13:31:12 | FIRMANTE

gOWn4/ROBQRzArMg/zgq4BEIDH98owpbSMIAMl1PsjylHLron3RNju8Fje7E6Qhf7iNNMrxhBbUxz89Hp1yygHPvmVfky+q2xQO4sBzoRjorelyhB9JhX6p2tNKyyhg7CxodnAlehyTpvE95wiP3BPpJ89sA2O8L3eDaXuKQaElEpGQU0T/LHgUDBqvjVbk1f0XHG5WCTgYVW8ol7MurtuK3EQXE3u4LRRbdSGZOpwv5mUYSZ/luZAWWhQ9QsszP9JfzDQM9QldUF8x6/Kdgpae0zdH1WhIkvn0CJt8qTgnf+kZDDR3QRWCNbcfnJw7boKA9Isr6lLvbqFlaHePjg==

CESAR MILLAN PACHECO | Fecha:2024-04-26 14:53:11 | FIRMANTE

O9gxj2KdMBJRLHVBn/w1GajAHSOpB8N75hgMs9rnMm6gEvj7Mkjm6uglXdqoPqFmrBQQ7VZaxYndIUx8KaBfhwgdGxFYUfh/e+ttjd7OODaZTtodWzS0Zix5UiWbd257N5Dm0dc24QIPD3tyk0ZJeQJpLelHdocVcZCa7IFl6gCxr7mP+E5zxr0tWBo9UYnlwPZpw0lkc5138CkTpE0UofYOk0ZcrKq499KVRlrP67a/EMe/6iWkqLynkt4V9DBef8L+eVueVQUHPwov+8w8YEb1qSwXk/jTaSW1jiAJu5IRtR133722/U0wozV/FHn9IVT7jedengE5D0kyqGFBA==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



FCwD2pJbZ

<https://efirma.uaem.mx/noRepudio/wnnbJr2BR1LMAv32wkk2hCqItXF3A0h>



UAEM
RECTORÍA
2023-2029