



Trabajo Fin de Grado
Grado en Biología

Estudio por cristalografía de rayos X de la interacción de ligandos con la proteína de la cápside del *Virus del mosaico de la sandía*

Autora:
Madalina Alexandra Dodu
Directora:
Alicia Alonso Izquierdo
Co-Directores:
Adriana Rojas Cardona
Mikel Valle Rodríguez

ÍNDICE

1.	RESUMEN.....	3
2.	ABSTRACT.....	3
3.	INTRODUCCIÓN	4
4.	OBJETIVOS	9
5.	MATERIAL Y MÉTODOS.....	9
5.1	Clonaje	9
5.2	Transformación bacteriana por choque térmico	9
5.3	Expresión proteica.....	10
5.3.1	Sobre-expresión de la CP nativa del WMV	10
5.3.2	Sobre-expresión de la CP del WMV marcada con SeMet.....	10
5.4	Purificación de la CP del WMV.....	10
5.4.1	Lisis celular	10
5.4.2	Cromatografía de afinidad.....	11
5.4.3	Digestión de la CP del WMV con proteasa TEV	11
5.4.4	Cromatografía de intercambio aniónico	11
5.4.5	Cromatografía de tamizado molecular	12
5.5	Electroforesis en geles de poliacrilamida.....	12
5.6	Cuantificación de la CP del WMV.....	12
5.6.1	Determinación de parámetros físico-químicos.....	12
5.6.2	Determinación de la concentración de la CP del WMV.....	12
5.7	Identificación de la CP del WMV	13
5.8	Cristalización de la CP del WMV con diferentes compuestos.....	13
5.8.1	Ligandos	13
5.8.2	Cristalización.....	13
5.9	Refinamiento de condiciones de cristalización	14
5.10	Recogida y procesamiento de los datos de difracción.....	14

6.	RESULTADOS.....	15
6.1	Parámetros físico-químicos de la CP del WMV.....	15
6.2	Sobre-expresión de la CP del WMV	15
6.3	Purificación de la CP del WMV.....	15
6.4	Identificación de la CP del WMV	16
6.5	Cristalización y refinamiento de la CP del WMV con los diferentes compuestos.....	17
6.6	Recogida y procesamiento de los datos de difracción.....	17
7.	DISCUSIÓN	18
8.	CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	20
9.	BIBLIOGRAFÍA	21
	ANEXO 1	25
	ANEXO 2	26
	ANEXO 3	27
	ANEXO 4.....	29
	ANEXO 5.....	30

1. RESUMEN

Los virus de plantas son unos de los principales causantes de enfermedades de cultivos. El grupo más numeroso de estos patógenos es el de los virus filamentosos y flexibles de plantas, que es el responsable de más del 50% del daño viral y, por consiguiente, de grandes pérdidas económicas. Comprende a más de 460 especies distribuidas en cuatro familias y sus proteínas de la cápside (CP) presentan un sitio de unión al RNA conservado. Esta región tiene importancia en el ensamblaje de los viriones. Por ello, se ha querido estudiar la interacción de diferentes compuestos, potencialmente antivirales, con este sitio de unión por cristalografía de rayos X, en concreto, en la CP del *Virus del mosaico de la sandía*. En este trabajo se purificó la proteína nativa y la marcada con selenometionina para su posterior cristalización con diferentes ligandos en más de 2000 condiciones. De los cristales analizados, sólo aquellos con la CP nativa difractaron y dieron lugar a patrones de difracción a una resolución máxima de 3,28 Å. Aunque los conjuntos de datos pudieron ser procesados en el grupo espacial P6, se observaron algunas reflexiones que no provenían de este tipo de organización. A pesar de ello, se hicieron diferentes intentos de resolver la estructura por reemplazamiento molecular, sin hallar aún una solución.

2. ABSTRACT

Plant viruses are one of the main causes of crop diseases. The largest group of these pathogens is formed by filamentous and flexible plant viruses, which are responsible for more than 50% of the viral damage and, consequently, for significant economic losses. The group comprises more than 460 species distributed in four families whose capsid proteins (CPs) have a conserved RNA binding site. This region is important in the assembly of virions. For this reason, we wanted to study the interaction of different potentially antiviral compounds with this binding site by X-ray crystallography, specifically, using the CP of the *Watermelon mosaic virus*. In this work, native protein and protein marked with selenomethionine were purified for their subsequent crystallization with different ligands in more than 2000 conditions. Of the analyzed crystals, only those of native CP diffracted and gave diffraction patterns at a maximum resolution of 3.28 Å. Although the datasets could be processed in the space group P6, some reflections that did not come from this type of organization were observed. Despite this, different attempts were made to solve the structure by molecular replacement, but no solution has been found yet.

3. INTRODUCCIÓN

Se calcula que para 2050 la población humana mundial habrá aumentado en un 30% desde el nivel actual y que se requerirá duplicar la producción de alimentos para poder satisfacer la demanda (Trebicki, 2020). Actualmente, los patógenos y las plagas son los responsables de la pérdida de un 10-30% de cultivos (Savary *et al.*, 2019) y, por consiguiente, de pérdidas económicas, siendo los virus uno de los principales causantes de enfermedades de las plantas (Jones & Naidu, 2019). Se estima que el cambio climático puede afectar a la virulencia y patogenicidad de estos virus, aumentando en consecuencia la frecuencia y escala de los brotes de enfermedades (Trebicki, 2020). Por lo tanto, el estudio de los virus que afectan a las plantas ayudará a desarrollar las estrategias necesarias para poder asegurar la producción mundial futura de alimentos.

Antes de proseguir, es importante dejar clara la definición de virus, dado el gran debate que ha supuesto y, muy posiblemente, sigue suponiendo actualmente en la comunidad científica. Teniendo en cuenta que los virus ya no se confunden con las partículas virales infectivas o viriones, que son una etapa del ciclo de replicación viral (Van Regenmortel, 2011), se define a los virus como entidades biológicas parasitarias consideradas organismos acelulares generadores de viriones, capaces de infectar y multiplicarse en el interior de una célula hospedadora. Estas partículas virales están compuestas por el material genético y una cápside que lo empaqueta, generalmente formada por proteínas que reciben el nombre de proteínas de la cubierta o *coat proteins* (CP) (Zamora Porras, 2018).

En cuanto a los virus vegetales, el grupo más numeroso es el de los virus filamentosos y flexibles de plantas, siendo el responsable de más de la mitad del daño viral a los cultivos en el mundo (DiMaio *et al.*, 2015). Estos presentan viriones filamentosos y flexibles, es decir, con una morfología cilíndrica flexible y formados por cápsides de CP que interaccionan directamente con el genoma viral de RNA monocatenario de polaridad positiva [(+)ssRNA], y que normalmente siguen una distribución helicoidal que les confiere su aspecto filamentosos (Zamora Porras, 2018) (**Figura 1**). Este grupo comprende más de 460 especies [International Committee on

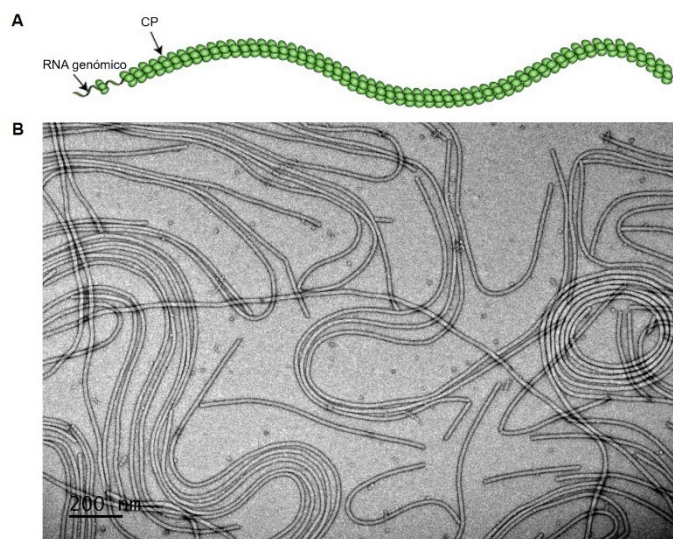


Figura 1. Viriones filamentosos y flexibles de plantas. (A) Representación esquemática de un virión flexible con proteínas de la cápside dispuestas helicoidalmente empaquetando el genoma viral. Tomada de la página ViralZone; (B) Micrografía del virión del *Virus del mosaico de la sandía* (WMV) obtenida por microscopía electrónica de transmisión en la que se observa su carácter flexible. Tomada en la Plataforma de Microscopía de CIC bioGUNE.

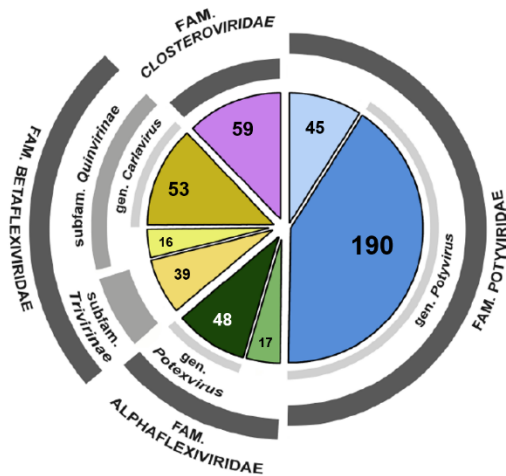


Figura 2. Diagrama con distribución de especies, géneros y familias de los virus filamentosos y flexibles de plantas. Modificado de Zamora Porras, 2018.

Taxonomy of Viruses (ICTV) versión 2020] distribuidas en cuatro familias de virus (ViralZone, consultado en junio 2021): *Alphaflexiviridae*, *Betaflexiviridae*, *Closteroviridae* y *Potyviridae* (**Figura 2**).

El género *Potyvirus*, de la familia *Potyviridae*, es el más numeroso y el más estudiado, principalmente debido a su importancia económica (Gadhavé *et al.*, 2020), con una representación de más del 40% de especies de virus filamentosos y flexibles de plantas conocidas (**Figura 2**). Una especie de virus de este género es el *Virus del mosaico de la sandía* (*Watermelon mosaic virus*, WMV), que infecta a más

de 170 especies de 26 familias de plantas, tanto monocotiledóneas como dicotiledóneas. Además de a curcubitáceas, también infecta a fabáceas, apiáceas e incluso orquídeas. Presenta una distribución mundial (Lecoq & Desbiez, 2012) y, concretamente, en los países mediterráneos es uno de los virus de cucurbitáceas más prevalentes (De Moya-Ruiz *et al.*, 2021). Los potyvirus (virus del género *Potyvirus*) se transmiten por más de 200 especies diferentes de pulgones (Gadhavé *et al.*, 2020), de los cuales más de 35 actúan como vectores de transmisión del WMV (Lecoq & Desbiez, 2012). Como indica su amplia gama de hospedadores, distribución mundial y diversidad de sus vectores, el WMV, al igual que el resto de los potyvirus (Nigam *et al.*, 2019), tiene una capacidad excepcional para adaptarse a nuevos hospedadores y entornos.

A grandes rasgos, el proceso de infección por potyvirus consiste en la entrada de los viriones a la célula huésped (que puede ser a través del estilete de los pulgones o de una célula adyacente), traducción, replicación, ensamblaje y movimiento de célula a célula. Una vez en el citoplasma, se lleva a cabo el desensamblaje y liberación del RNA genómico viral que servirá como molde para la traducción de proteínas virales, posiblemente, en la membrana del retículo endoplasmático. La replicación se cree que tiene lugar en

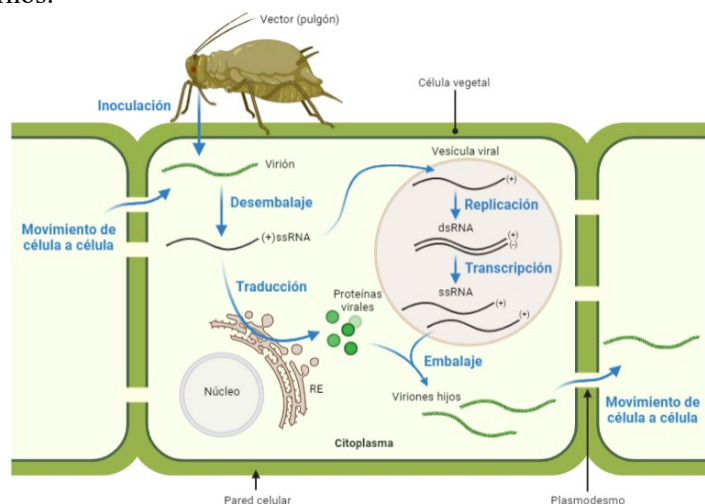


Figura 3. Proceso simplificado de la infección de viriones de *Potyvirus* a una célula vegetal. La entrada de los viriones puede darse mediante un vector de transmisión o por el movimiento de célula a célula. Tras el desensamblaje, se lleva a cabo la traducción de proteínas virales en la membrana del retículo endoplasmático (RE) y la replicación del genoma en vesículas virales, para generar nuevos viriones que puedan seguir infectando a las células adyacentes a través de los plasmodesmos. No se ha respetado la escala.

vesículas virales en el citoplasma asociadas al sistema de endomembranas y consiste en la síntesis de (-)ssRNA a partir del (+)ssRNA para poder realizar la transcripción y obtener así nuevos genomas virales. Después, se lleva a cabo el ensamblaje de los viriones hijos en el citoplasma y, finalmente, su movimiento de célula a célula a través de los plasmodesmos (Gadhavé *et al.*, 2020; Mäkinen & Hafrén, 2014; ViralZone) (**Figura 3**). El WMV induce diversos síntomas en las plantas infectadas, como mosaico, formación de ampollas y deformación en las hojas, y decoloración severa y cierta deformación en los frutos (Lecoq & Desbiez, 2012; Wang & Li, 2017) (**Figura 4**).



Figura 4. Síntomas en plantas infectadas con el *Virus del mosaico de la sandía* (WMV): mosaico en hojas (A) y decoloración del fruto (B) de una planta de sandía (De Moya-Ruiz *et al.*, 2021) y de una planta de melón (C) (Lecoq & Desbiez, 2012).

Con el fin de poder entender cualquier proceso biológico, es necesario conocer la estructura tridimensional de las biomoléculas involucradas. Para ello, se han desarrollado diferentes técnicas, siendo muy empleadas en la determinación estructural de macromoléculas biológicas la criomicroscopía electrónica (crioEM), un tipo de microscopía electrónica de transmisión en el que la muestra se examina a temperaturas criogénicas, y la cristalografía de rayos X (Dubach & Guskov, 2020; Venkataraman *et al.*, 2018). En concreto, la estructura de los viriones del WMV fue caracterizada por crioEM con una resolución de 4,0 Å. Su estructura helicoidal muestra una hélice levógira con un diámetro de 130 Å que contiene aproximadamente 8,8 subunidades de CP por vuelta y un paso de hélice de 35,2 Å (**Figura 5A**). La CP del WMV exhibe un dominio central rico en hélices α y dos brazos flexibles N- y C-terminales largos que median la polimerización de las CP en el virión y permiten movimientos entre ellas, lo que explica la naturaleza flexible de las partículas virales (Zamora *et al.*, 2017) (**Figura 5B**).

A pesar de que las secuencias genómicas de las CP de las cuatro familias de virus filamentosos y flexibles de plantas tienen una baja similitud, la arquitectura de los viriones a baja resolución parece ser similar y el plegamiento de las CP está claramente conservado, lo que indica, por tanto, un origen evolutivo común (Kendall *et al.*, 2008). A pesar de la ausencia de homología de secuencias entre las CP de distintas familias, los residuos de Ser, Arg y Asp del sitio de unión al ssRNA (**Figura 5C**) se encuentran conservados en las CP. Además, esta región es indispensable para el ensamblaje y el movimiento de célula a célula de los viriones (Zamora *et al.*, 2017). Así, el objetivo del proyecto es la caracterización a nivel atómico de la interacción de diferentes compuestos preseleccionados con

este sitio de unión de los virus vegetales filamentosos y flexibles, para la búsqueda de compuestos antivirales que puedan interferir en el ensamblaje y el empaquetamiento del genoma de un gran número de virus vegetales económicamente relevantes. Para cumplir este objetivo, estudiando la interacción en la CP del WMV, se han combinado las técnicas de crioEM y cristalografía de rayos X, siendo esta última la empleada en este trabajo.

La cristalografía de rayos X es una herramienta empleada en biología estructural que permite determinar la estructura de las moléculas a resolución atómica (0,48 Å) (Dubach & Guskov, 2020), dando la posibilidad de observar interacciones como puentes de hidrógeno, puentes salinos e

interacciones hidrofóbicas. Consiste en irradiar con rayos X cristales de la proteína de interés. En los cristales hay un ordenamiento y una repetición periódica de estas biomoléculas, es decir, una cierta distribución simétrica en el espacio formando una red cristalina (Figura 6C). El cristal difracta el haz de rayos X en muchas direcciones, dando lugar a un patrón de difracción específico (Figura 7A). Cuanto más lejos estén los puntos de reflexión del centro del detector, mayor será la información de resolución que contengan. Sin embargo, al aumentar la resolución, la señal disminuye. A cierta distancia del centro del detector, la señal será demasiado débil (Figura 7B) y, en principio, será el límite de resolución del conjunto de datos (Dubach & Guskov, 2020; Martínez-Ripoll, n.d.).

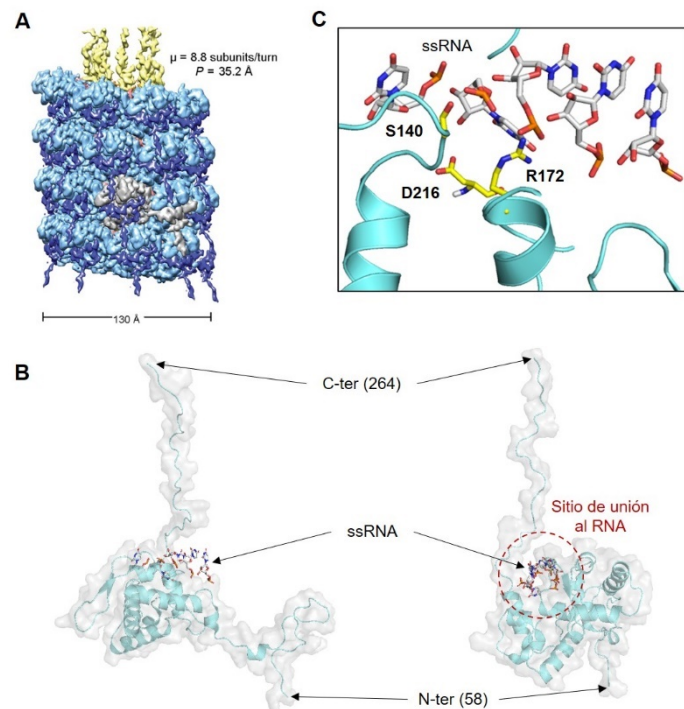


Figura 5. Estructura Cryo-EM del virión del WMV y sitio de unión al RNA en la CP. (A) Representación del mapa 3D de crioEM calculado para el WMV: región central o *core* de la CP del WMV (azul claro), brazo N-terminal (azul oscuro), brazo C-terminal (amarillo) y una subunidad de CP (gris). Tomada de Zamora *et al.*, 2017. (B) Dos vistas del modelo atómico de la CP del WMV (Protein Data Bank (PDB) código: 5ODV) donde se observa el sitio de unión del ssRNA a la proteína. (C) Detalle de la interacción entre el ssRNA y la CP del WMV centrado en los residuos Ser140, Arg172 y Asp216.

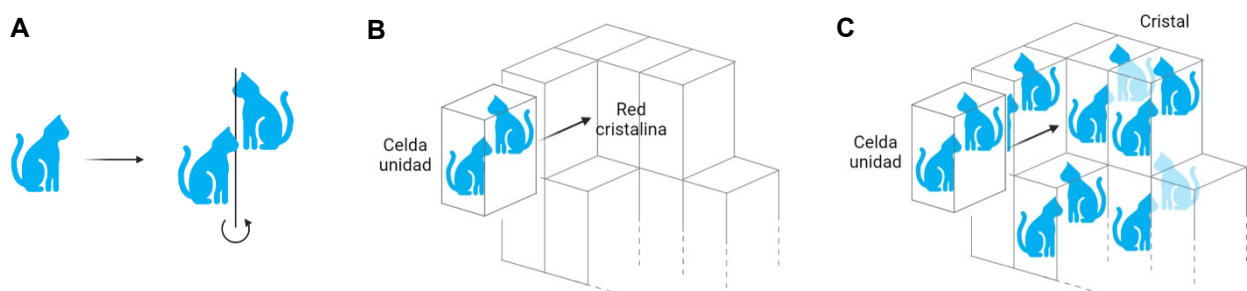


Figura 6. Ordenamiento de proteínas en un cristal. (A) Unidad asimétrica con una proteína (dibujo de gato) que, por operaciones de simetría (giro lateral de 180° en este caso), permite reproducir el contenido de (B) la celda unidad. Se observa cómo las celdas se repiten a lo largo de la red cristalina generando (C) el cristal.

La unidad que se repite y permite construir todo el cristal por operaciones de translación recibe el nombre de celda unidad (**Figura 6B**). El cómo se organizan las moléculas dentro de esta celda en el espacio se conoce como grupo espacial. La unidad mínima de la celda unidad que permite reproducir el contenido de esta celda por operaciones de simetría recibe el nombre de unidad asimétrica (**Figura 6A**). El primer paso para resolver la estructura cristalográfica de una macromolécula es determinar el grupo espacial en el cual ha cristalizado, así como los parámetros de la celda unidad. Después, para calcular la posición de los átomos dentro del cristal es necesario calcular una magnitud conocida como factor de estructura (Rupp, 2010), que es un número complejo cuya magnitud es proporcional a la intensidad de las reflexiones, pero su fase no se puede medir directamente de forma experimental. Esto se conoce como “el problema de las fases” en cristalografía y se puede calcular de forma indirecta por dos métodos conocidos como reemplazamiento molecular (RM) y reemplazamiento isomorfo (RI). En RM se emplea un modelo estructural de una proteína con una secuencia de aminoácidos homóloga. En RI se introducen átomos pesados en el cristal, que sirven de referencia para calcular las fases, o se marcan de forma específica las metioninas de la proteína, reemplazando el azufre por un átomo de selenio [Selenometionina (SeMet)] (Martínez-Ripoll, n.d.).

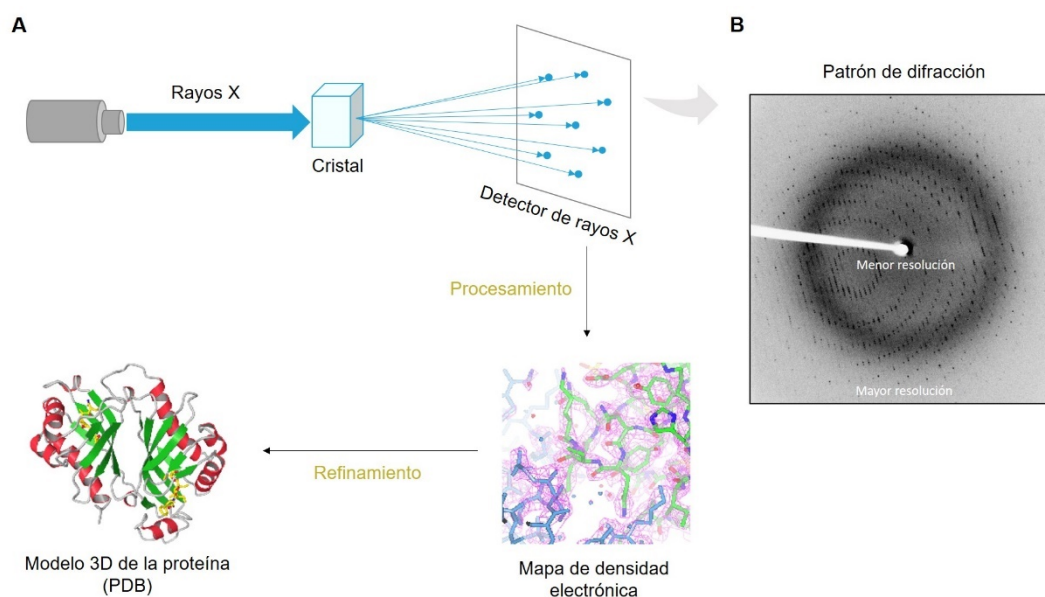


Figura 7. Diagrama esquemático para la caracterización de proteínas mediante cristalografía de rayos X. (A) El haz de rayos X incide en el cristal de la proteína y es difractado en muchas direcciones específicas que generan un patrón de difracción determinado. Este conjunto de reflexiones es después procesado hasta obtener un mapa de densidad electrónico que es refinado para construir el mapa atómico 3D de la proteína. (B) Patrón de difracción con puntos de reflexión a menor (cerca del centro del detector) y a mayor resolución.

El cristal se expone al haz de rayos X en rotación continua. Dependiendo de la simetría del cristal, esta puede variar entre 60 y 360°. En general, se toman imágenes cada 0,1° y se genera un conjunto de patrones de difracción. Estos patrones se pueden analizar con la ayuda de diferentes programas bioinformáticos que nos permiten asociar un índice a cada reflexión (h, k, l) y un valor de intensidad.

La calidad de los datos se puede estimar gracias a diferentes parámetros estadísticos, siendo importantes a tener en cuenta los valores del R-meas, el coeficiente de correlación de una mitad del conjunto de datos (escogida aleatoriamente) comparada con la otra mitad ($CC_{1/2}$), la completitud y la relación señal/ruido (I/σ) (**Tabla 1**) (Powell, 2017). Estos también se emplean para truncar los datos y así decidir qué reflexiones no son de una calidad suficiente. Finalmente, después de calcular las fases se puede generar un mapa de densidad en el cual se construye un modelo que se conoce como PDB, donde a cada átomo dentro del cristal se le asigna una posición (x, y, z) (**Figura 7A**) (Dubach & Guskov, 2020; Rupp, 2010).

Tabla 1. Valores a alta y baja resolución del R-meas, $CC_{1/2}$, completitud e I/σ a tener en cuenta durante el procesamiento del conjunto de reflexiones del patrón de difracción del cristal de proteína (Powell, 2017).

Estadístico	Valores	
	Alta resolución	Baja resolución
R-meas	-	< 10%
$CC_{1/2}$	> 40%	100%
Completitud	> 90%	-
I/σ	> 0,5	-

4. OBJETIVOS

- Purificación de la CP del WMV para su posterior cristalización con diferentes compuestos antivirales preseleccionados.
- Determinación de la estructura de la interacción de estos compuestos con la CP del WMV por cristalografía de rayos X.

5. MATERIAL Y MÉTODOS

5.1 Clonaje

La secuencia de la CP del WMV fue modificada en sus extremos flexibles C- y N-terminal para permitir la purificación y la cristalización de la proteína. De esta manera, la secuencia correspondiente a los aminoácidos 71-247 (**ANEXO 1**) fue clonada por la empresa GenScript en el vector de expresión pET-28a(+)-TEV usando los enzimas de restricción *NdeI* y *XhoI*. Esto permitió obtener la CP con el sitio de corte de la proteasa del *Virus del grabado del tabaco* (TEV) y la etiqueta de seis histidinas en el extremo N-terminal (**ANEXO 2**), empleando como marcador de selección la kanamicina.

5.2 Transformación bacteriana por choque térmico

Se transformaron con pET-28a(+)-TEV+WMVCP mediante choque térmico dos cepas competentes de *Escherichia coli*: BL21(DE3) para la sobre-expresión de la CP nativa y B834(DE3) para la sobre-expresión de la CP marcada con SeMet. Se emplearon 2 μ l de plásmido para transformar 50 μ l de células competentes. Se incubó en hielo durante 30 minutos y después, para realizar el choque térmico, se aplicó calor durante 30 segundos a 42°C seguido de una incubación en hielo durante

2 minutos. Se agregaron 250 µl de medio LB Broth (SIGMA) atemperado y se incubó durante 90 minutos a 37°C a 300 rpm. Después, se sembraron 150 µl de células de la transformación en placa con agar LB (Lennox) (CONDA) suplementado con 50 µg/ml de kanamicina y se incubaron a 37°C durante la noche.

5.3 Expresión proteica

5.3.1 Sobre-expresión de la CP nativa del WMV

Se inocularon colonias de *E. coli* BL23(DE3) transformadas en medio LB Broth con 50 µg/ml de kanamicina y se incubaron toda la noche a 37°C a 300 rpm. Al día siguiente, se inocularon con cultivo saturado 6 L de medio LB Broth fresco con 50 µg/ml de kanamicina obteniendo una densidad óptica a 600 nm (OD₆₀₀) de 0,1. Se incubó durante 2 horas a 37°C a 190 rpm hasta alcanzar una OD₆₀₀ de 0,6, momento en el cual se indujo la sobre-expresión de la proteína añadiendo 1 mM de isopropil-β-D-1-tiogalactopiranosido (IPTG) y se incubó durante la noche a 20°C a 190 rpm. El cultivo se centrifugó a 4.500 rpm a 4°C durante 30 minutos y el pellet se almacenó a -80°C. Se recogieron muestras antes y después de la inducción y se analizaron por electroforesis (apartado 5.5).

5.3.2 Sobre-expresión de la CP del WMV marcada con SeMet

Células auxótrofas de metionina *E. coli* B834(DE3) transformadas se inocularon al igual que lo descrito en el apartado anterior (5.3.1) en 4 L de medio LB Broth fresco con 50 µg/ml de kanamicina. Cuando el cultivo alcanzó una OD₆₀₀ de 0,6, se centrifugó a 4.500 rpm a 4°C durante 30 minutos. Se descartó el sobrenadante y el pellet celular se resuspendió cuidadosamente, para no romper las células, en 3 L de PBS 1X frío (a 4°C) para eliminar restos de medio LB Broth. A continuación, se volvió a centrifugar (4.500 rpm, 4°C, 30 minutos) y se descartó el sobrenadante. El pellet se resuspendió en 4 L de medio mínimo atemperado suplementado con nutrientes SeMet (Molecular Dimensions), 0,04 mg/ml SeMet y 50 µg/ml de kanamicina. Se incubó a 37°C a 190 rpm durante una hora y se indujo la sobre-expresión de la CP marcada con SeMet añadiendo 1 mM de IPTG. Se incubó a 20°C a 190 rpm durante la noche y las células se recogieron centrifugando (4.500 rpm, 4°C, 30 minutos) y se almacenaron a -80°C. Se recogieron muestras antes (tanto del cultivo LB Broth como del medio mínimo) y después de la inducción y se analizaron por electroforesis (apartado 5.5).

5.4 Purificación de la CP del WMV

Los siguientes pasos se realizaron tanto para la purificación de la CP nativa como para la CP marcada con SeMet.

5.4.1 Lisis celular

Se descongeló el pellet correspondiente a 2 L de cultivo y se resuspendió en tampón de lisis [25 mM HEPES pH 8, 0,5 M NaCl, 20% glicerol, 20 mM imidazol, 1 mM MgCl₂, ½ pastilla de inhibidores de proteasas (Roche) y DNasa]. La rotura celular se llevó a cabo, en hielo en todo momento, con el emisor de ultrasonidos *UP400S* (Hielscher) con un sonotrodo de titanio de 3 mm en la punta

(Hielscher). Se realizaron 7 ciclos de sonicación de 30 segundos a una amplitud de 60% con intervalos de 1 minuto de descanso entre pulsos. Se centrifugó a 30.000 rpm a 4°C durante 40 minutos. Se recogieron muestras del sobrenadante y del pellet y se analizaron por electroforesis (apartado 5.5) para verificar que las proteínas estaban en la fracción soluble y no en cuerpos de inclusión.

5.4.2 Cromatografía de afinidad

Para la separación de la CP del WMV de otras proteínas celulares se realizó una cromatografía de afinidad por iones metálicos inmovilizados (IMAC) empleando la columna *HisTrap Fast Flow* de 5 ml (GE Healthcare), cuya resina consta de perlas de agarosa con un ligando quelante inmovilizado cargado con iones de Ni²⁺ que aseguran la captura de proteínas con etiqueta de histidinas. La columna se equilibró en *ÄKTA Purifier* (GE Healthcare) con tampón de unión (50 mM Tris HCl pH 7,6, 150 mM NaCl, 20 mM imidazol). El sobrenadante de la centrifugación se cargó a la columna y se lavó con un flujo de 1 ml/min con tampón de unión y un gradiente hasta 100% de tampón de elución (50 mM Tris HCl pH 7,6, 150 mM NaCl, 500 mM imidazol). Después, se recogieron fracciones de la muestra que no se unieron a la resina y de las que eluyeron en el máximo de absorbancia del pico en el perfil cromatográfico. Se analizaron por electroforesis (apartado 5.5) para comprobar que no se perdió proteína en el lavado y que el pico correspondía a la CP.

Después de digerir con proteasa TEV (apartado 5.4.3), también se realizó una cromatografía de afinidad para la separación del enzima y de la CP con y sin etiqueta de histidinas.

5.4.3 Digestión de la CP del WMV con proteasa TEV

Las fracciones de CP recolectadas después de la cromatografía de afinidad se digirieron con 1,5 mg/ml de proteasa TEV para eliminar la etiqueta de histidinas de la CP mediante un corte específico. La digestión se llevó a cabo simultáneamente con una diálisis para eliminar el imidazol. Así, las fracciones se transfirieron con el enzima proteolítico a una membrana de diálisis y se incubaron en tampón de diálisis (50 mM Tris HCl pH 7,6, 150 mM NaCl, 1 mM MgCl₂) a 4°C con agitación durante la noche.

5.4.4 Cromatografía de intercambio aniónico

Se realizó después una cromatografía de intercambio aniónico empleando la columna *HiTrap Q HP* de 5 ml (GE Healthcare). Para ello, la muestra se diluyó con Tris HCl pH 7,6 hasta tener una concentración de NaCl de 100 mM para disminuir su fuerza iónica. Se equilibró la columna en *ÄKTA Purifier* (GE Healthcare) con tampón de unión (50 mM Tris HCl pH 7,6, 1 mM MgCl₂) y se inyectó la muestra. Se lavó con un flujo de 1 ml/min con tampón de unión y un gradiente hasta 100% de tampón de elución (50 mM Tris HCl pH 8,5, 1 M NaCl, 1 mM MgCl₂). Se eluyó la CP aumentando gradualmente la fuerza iónica ejercida por el NaCl. A continuación, se recogieron las fracciones del pico de la cromatografía y se analizaron por electroforesis (apartado 5.5). Las fracciones obtenidas se

concentraron utilizando un concentrador *Vivaspin 6* (Storius Stedim Biotech) con un MWCO (*Molecular weight cut-off*) de 3.000 Da hasta un volumen de 5 ml.

5.4.5 Cromatografía de tamizado molecular

Por último, la CP del WMV se purificó por cromatografía de tamizado molecular a 4°C empleando la columna *HiLoad 16/60 Superdex 75* de 120 ml (GE Healthcare) cuya resina permite la separación en función del tamaño de proteínas de hasta 75 kDa. La columna se equilibró en *ÄKTA pure 25* (GE Healthcare) con tampón (50 mM Tris HCl pH 7,7, 150 mM NaCl, 1 mM MgCl₂, 1 mM DTT). Se cargaron a la columna los 5 ml de CP concentrada, siendo este el volumen máximo que se puede cargar para que la separación sea eficiente. Después, se recogieron las fracciones del pico del cromatograma y se analizaron por electroforesis (apartado 5.5).

5.5 Electroforesis en geles de poliacrilamida

El análisis de las proteínas se realizó en geles comerciales de poliacrilamida *Mini-PROTEAN TGX Precast Gels* (Bio-Rad) mediante electroforesis vertical en *Mini-PROTEAN Tetra Vertical Electrophoresis Cell* (Bio-Rad). Las muestras se mezclaron con tampón de carga *2x Lamli Sample Buffer* (Bio-Rad) y H₂O MilliQ y se calentaron a 100°C durante 5 minutos. Como marcadores de peso molecular se emplearon *NZYColour Protein Marker II* (NZYTech) y *Precision Plus Protein Standards Dual Color* (Bio-Rad). La electroforesis se realizó a 200 mV durante 25 minutos en tampón de electroforesis *Tris-glycine-SDS buffer 10X* (Bio-Rad). Los geles se tiñeron durante 2 minutos al microondas en la solución *Problue safe stain 10X* (Giotto Biotech) y se destiñeron en varios lavados con H₂O MilliQ.

5.6 Cuantificación de la CP del WMV

5.6.1 Determinación de parámetros físico-químicos

Para el cálculo de varios parámetros físicos y químicos para la CP del WMV (peso molecular, punto isoeléctrico teórico, coeficiente de extinción molar, entre otros) se empleó la herramienta de análisis e identificación de proteínas ProtParam del servidor ExPASy.

5.6.2 Determinación de la concentración de la CP del WMV

La concentración de la proteína se determinó mediante la Ley de Lambert-Beer (**ecuación 1**) teniendo en cuenta la absorbancia obtenida a 280 nm medida en un espectrofotómetro NanoDrop One (Thermo Scientific) y el coeficiente de extinción molar de la proteína proporcionado por ProtParam (apartado 5.6.1).

$$A = \varepsilon \cdot C \cdot l \quad (\text{ecuación 1})$$

Donde A es la absorbancia, ε es el coeficiente de extinción molar (g/L), C es la concentración (g/L) y l es la longitud del camino recorrido por la luz (cm).

5.7 Identificación de la CP del WMV

Se empleó el servicio de la Plataforma de Proteómica de CIC bioGUNE para identificar la CP nativa y la CP marcada con SeMet mediante espectrometría de masas. Estas se digirieron con tripsina y cada una generó una mezcla de péptidos únicos cuya medición de masas moleculares proporcionaron unos conjuntos de datos característicos de las proteínas (Domon & Aebersold, 2006; Webster & Oxley, 2012).

5.8 Cristalización de la CP del WMV con diferentes compuestos

5.8.1 Ligandos

Se emplearon tres compuestos candidatos antivirales con los que cristalizar la CP: 25D8, 3G9 y 28C6¹. Debido a su insolubilidad en agua, se realizaron diferentes disoluciones con dimetilsulfóxido (DMSO) hasta obtener la concentración adecuada para solubilizarlos.

5.8.2 Cristalización

Para determinar la concentración adecuada de la CP para el cribado de cristalización con los diferentes ligandos, se realizó un test de pre-cristalización (PCT) (Hampton Research) con dos agentes precipitantes: sulfato de amonio y polietilenglicol 3350 (PEG 3350). Para la proteína nativa con los compuestos se emplearon 12 kits comerciales de la Plataforma de Rayos-X de CIC bioGUNE: AmSO₄ Suite (QUIAGEN), Index (Hampton Research), JCSG-*plus* (Molecular Dimensions), MIDAS (Molecular Dimensions), Morpheus (Molecular Dimensions), Natrix (Hampton Research), PACT Suite (QUIAGEN), PEG/Ion (Hampton Research), ProPlex (Molecular Dimensions), Structure Screen 1&2 (Molecular Dimensions), Stura MacroSol (Molecular Dimensions) y Wyzard Cryo 1&2 (Molecular Dimensions). Para el cribado de las condiciones de cristalización de la proteína marcada con SeMet con los compuestos, se emplearon 5 de los 12 kits comerciales: AmSO₄ Suite, Morpheus, PEG/Ion, ProPlex y Stura MacroSol. Cada kit contiene 96 condiciones diferentes.

Cada reacción de cristalización se preparó empleando un mosquito (SPT Labtech), añadiendo 1 µl de la mezcla de proteína y de compuesto, y 1 µl de cada solución tampón de cristalización de los kits comerciales de cristalización. Se empleó el método de difusión de vapor por gota sentada, que consiste en colocar una gota con solución de proteína y de agente precipitante en un recipiente sellado, y en equilibrar frente a un reservorio que contiene únicamente la solución de agente precipitante. Como la concentración de agente precipitante en la gota es menor, se produce difusión de vapor desde la gota al reservorio hasta que se equilibran las concentraciones (Casaos Aznar, 2014; McPherson & Gavira, 2014). Las mezclas resultantes se dejaron a 18°C y se observó el crecimiento de cristales al microscopio después de una semana.

¹ Se ha añadido la nomenclatura interna. Por problemas de confidencialidad y para protección de datos para su patentabilidad, se excluyen las estructuras y la nomenclatura de los compuestos antivirales.

5.9 Refinamiento de condiciones de cristalización

Se emplearon placas con 48 pocillos, una mezcla igual de proteína, ligando y tampón que en el cribado de las condiciones de cristalización (apartado 5.8.2), y diferentes soluciones tampón de refinamiento preparadas (ANEXO 3). Las soluciones de cristalización se prepararon empleando el mosquito, añadiendo diferentes proporciones de proteína+compuesto:tampón de refinamiento 1:1, 2:1 y 1:2.

5.10 Recogida y procesamiento de los datos de difracción

Los cristales obtenidos fueron pescados y enviados en nitrógeno líquido al sincrotrón ALBA (Barcelona, España) y Diamond Light Source (DLS) (Didcot, Reino Unido) para analizarlos por difracción de rayos X. Estos se mandaron con diferentes crioprotectores (Pflugrath, 2015): glicerol, parafina, silicona, PEG 600, 2-Methyl-2,4-pentanediol (MPD), xilitol y sacarosa.

Se realizaron cinco recogidas de forma remota de los datos de difracción de esos cristales, una en DLS y el resto en ALBA. Para la recogida de datos de aquellos cristales de proteína nativa se empleó una energía de 12.622 keV y, para aquellos con proteína marcada con SeMet, una de 12.6548 keV.

El procesamiento del conjunto de reflexiones obtenidas se llevó a cabo en tres etapas: indexación, integración y escalado (Powell, 2017). Para ello se emplearon, por un lado, el paquete XDS (Kabsch, 2010), con diferentes programas ejecutables que realizaron estas etapas de manera automática y, por otro, el *software* CCP4 (*Collaborative Computational Project Number 4*) (Winn *et al.*, 2011). Para algunos conjuntos de datos se realizó el proceso de indexación e integración utilizando el programa iMosflm de CCP4 (Powell *et al.*, 2017). En la primera etapa se obtuvo el grupo espacial, una estimación de las dimensiones de la celda unidad y los índices para cada reflexión (h , k , l). En la segunda etapa se definieron las intensidades de las reflexiones, sumando las intensidades de cada uno de los píxeles que forma un punto en el patrón de difracción. Posteriormente, para el proceso de escalado se empleó el programa Scala de CCP4, en el cual se normalizaron las intensidades de cada una de las imágenes de difracción.

El reemplazamiento molecular se realizó con el programa Molrep de CCP4 y el *software* Phaser (McCoy *et al.*, 2007) empleando tres versiones del PDB de la estructura del potyvirus del WMV (código 5ODV) que se generaron con el programa Coot (Emsley *et al.*, 2010): monómero, dímero y trímero.

La figura con la estructura de la CP nativa se realizó empleando PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC).

6. RESULTADOS

6.1 Parámetros físico-químicos de la CP del WMV

Se determinó que el punto isoeléctrico de la CP del WMV es de 5,83 y que el peso molecular total teórico de la proteína nativa es de 22.700,70 Da y, el de la marcada con SeMet, de 23.169,66 Da.

6.2 Sobre-expresión de la CP del WMV

El análisis por electroforesis de las muestras de antes y después de la inducción reveló una banda a la altura del peso molecular de la CP mucho más intensa después de inducir. Se observó una sobre-expresión tanto de la CP nativa del WMV (**Figura 8**) como de la marcada con SeMet (**Figura 9**), no habiendo una expresión anómala antes de inducir.

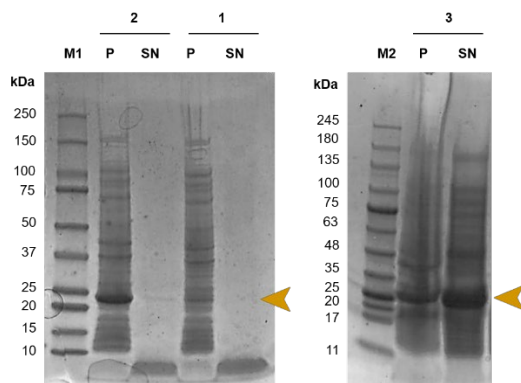


Figura 8. Sobre-expresión de la CP nativa del WMV. Electroforesis en gel de poliacrilamida. M1: marcador de peso molecular *Precision Plus Protein Standards Dual Color* (Bio-Rad), M2: marcador de peso molecular *NZYColour Protein Marker II* (NZYTech), P: muestra del pellet, SN: muestra del sobrenadante, 1: condición de pre-inducción, 2: condición de post-inducción, 3: condición de lisado celular de 2 L de cultivo.

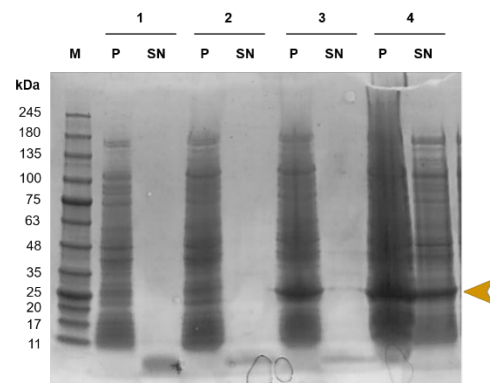


Figura 9. Sobre-expresión de la CP del WMV marcada con SeMet. Electroforesis en gel de poliacrilamida. M: marcador de peso molecular *NZYColour Protein Marker II* (NZYTech), P: muestra del pellet, SN: muestra del sobrenadante, 1: condición de pre-inducción en LB Broth, 2: condición de pre-inducción en medio mínimo, 3: condición de post-inducción, 4: condición de lisado celular de 2 L de cultivo.

6.3 Purificación de la CP del WMV

Las electroforesis en gel de las muestras lisadas indicaron que gran parte de ambas proteínas se encontraba soluble en el sobrenadante, observándose una banda más intensa en el caso de la CP nativa (**Figura 8**) que en el de la CP marcada con SeMet (**Figura 9**).

La purificación de ambas proteínas se llevó a cabo de manera eficiente (**Figura 10 y 11**), ya que toda la proteína se pegó a la resina (**Figura 10B**). Al realizar la cromatografía de afinidad después de digerir con la proteasa TEV, tanto la CP nativa como la marcada con SeMet se unieron a la columna. Por tanto, las proteínas se purificaron con las etiquetas de histidinas. En la cromatografía de intercambio aniónico, la CP nativa eluyó con una conductividad de 65,45 mS/cm (**Figura 10C**) y la marcada con SeMet con una de 50,44 mS/cm (**Figura 11C**). En la cromatografía de tamizado molecular, la proteína nativa eluyó en el ml 49,8 (**Figura 10E**) y la marcada con SeMet en el 50,6 (**Figura 11E**).

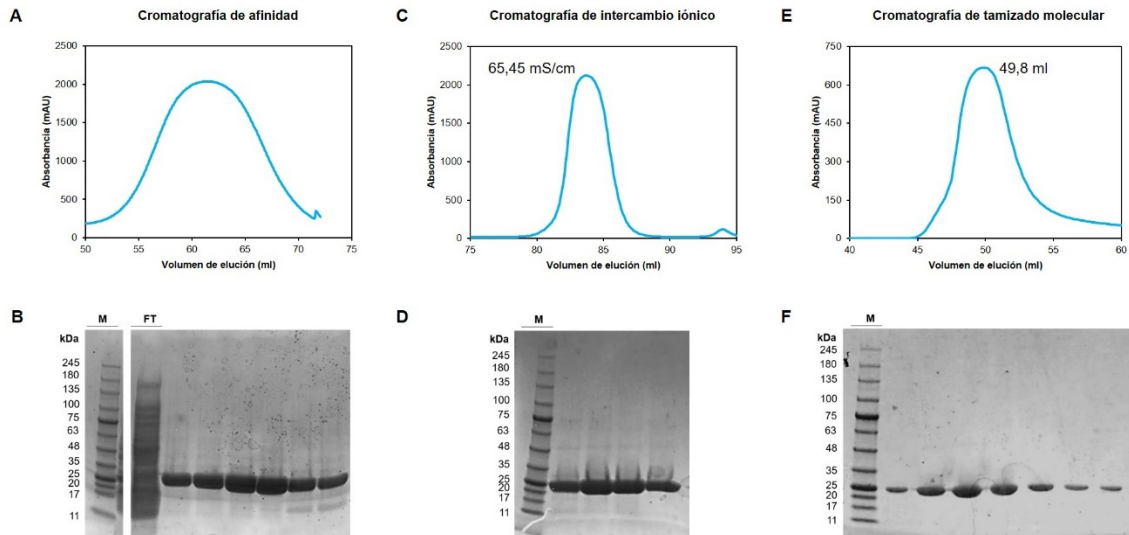


Figura 10. Detalle de los picos de absorbancia de los perfiles cromatográficos y análisis por electroforesis en geles de poliacrilamida de la CP nativa del WMV. M: marcador de peso molecular *NZYColour Protein Marker II* (NZYTech), FT: fracciones de la muestra que no se ha unido (*flow through*) a la resina en la cromatografía de afinidad. (A) Cromatografía de afinidad antes de la digestión con TEV y (B) su análisis electroforético; (C) Cromatografía de intercambio iónico y (D) su análisis electroforético; (E) Cromatografía de tamizado molecular y (F) su análisis electroforético.

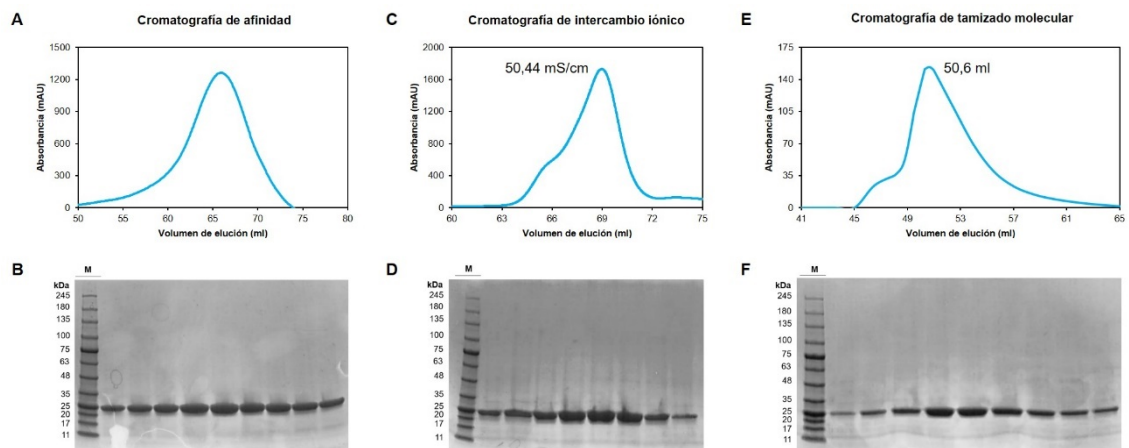


Figura 11. Detalle de los picos de absorbancia de los perfiles cromatográficos y análisis por electroforesis en geles de poliacrilamida de la CP del WMV marcada con SeMet. M: marcador de peso molecular *NZYColour Protein Marker II* (NZYTech). (A) Cromatografía de afinidad antes de la digestión con TEV y (B) su análisis electroforético; (C) Cromatografía de intercambio iónico y (D) su análisis electroforético; (E) Cromatografía de tamizado molecular y (F) su análisis electroforético.

6.4 Identificación de la CP del WMV

Tanto la CP nativa como la marcada con SeMet fueron identificadas mediante espectrometría de masas (ANEXO 4), consiguiéndose una incorporación del 100% de SeMet en la proteína marcada.

6.5 Cristalización y refinamiento de la CP del WMV con los diferentes compuestos

En el PCT se determinó una concentración proteica adecuada de 6 mg/ml. Se cristalizó con 6,15 mg/ml de CP nativa y 6,55 mg/ml de CP marcada con SeMet, con 500 μ M de compuesto.

Se testaron 1632 condiciones de los kits comerciales con proteína nativa y se obtuvieron cristales al cabo de una semana, siendo los mejores los de AmSO₄ Suite (condición B2), ProPlex [condición G6 (**Figura 12A**)] y Morpheus (condiciones G7 y H7) (**ANEXO 5**). Estas condiciones fueron refinadas en un total de 10 placas, concretamente en 480 condiciones diferentes (**ANEXO 3**). También se obtuvieron cristales en la condición G6 de ProPlex al cabo de seis meses.

6.6 Recogida y procesamiento de los datos de difracción

Se mandaron un total de 113 cristales de proteína y ligando para su análisis por difracción de rayos X. Se recogieron 3.600 imágenes por cada cristal. Casi todos los cristales no difractaron o difractaron muy poco, a excepción de los cristales de la condición G6 de ProPlex con CP nativa (**Figura 12**). Ninguno de los cristales de proteína marcada con SeMet difractó.

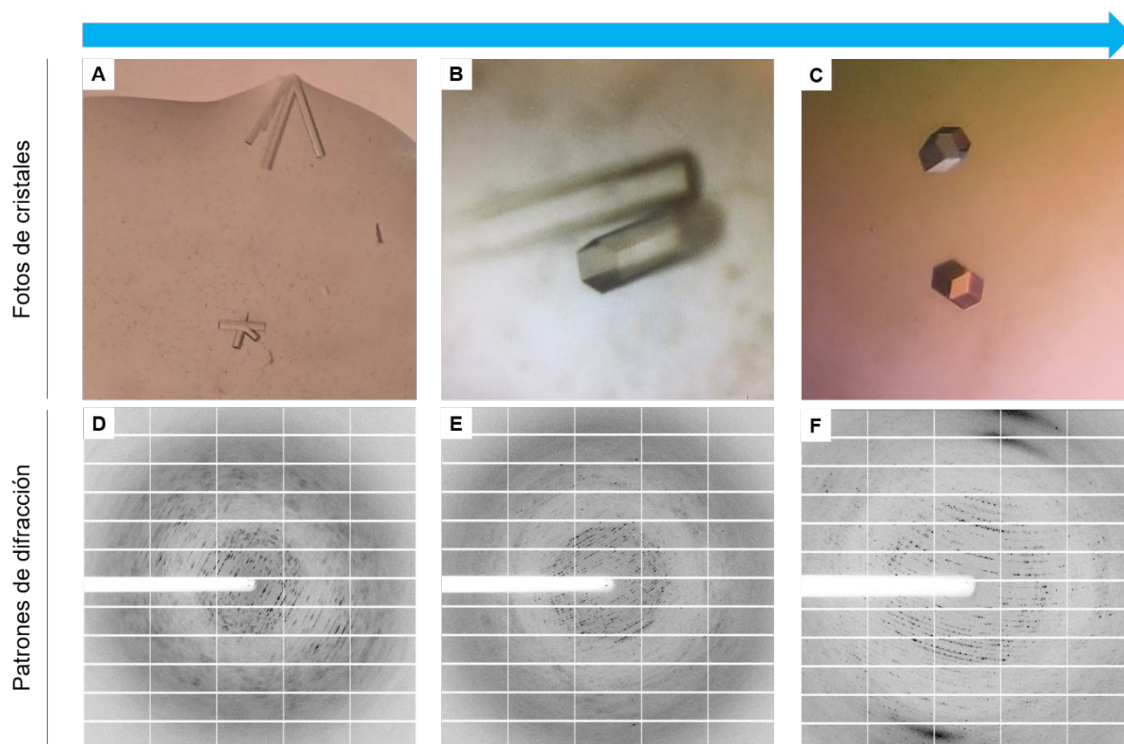


Figura 12. Evolución a lo largo del tiempo de cristales y patrones de difracción. (A, B, C) Fotos de cristales de la condición G6 del kit Proplex; (D) Patrón de difracción con anisotropía y superposición de las reflexiones; (E) Patrón de difracción con reflexiones más definidas y poca señal a alta resolución; (F) Patrón de difracción con poca señal a alta resolución.

Se obtuvo un mejor patrón de difracción de los cristales congelados después de seis meses de su aparición, de los refinados y de aquellos enviados al sincrotrón con un 21% de sacarosa como crioprotector. Sin embargo, no se obtuvo un conjunto de datos a alta resolución (**Figura 12F**).

Los conjuntos de datos fueron procesados hasta obtener valores aceptables de $CC_{1/2}$, completitud e I/σ (**Tabla 1**). Las mejores estadísticas fueron obtenidas en el procesamiento en un grupo espacial hexagonal P6 (**Tabla 2**). Sin embargo, en general, la mayoría de los conjuntos no presentan buenos valores de R-meas en la baja resolución.

Tabla 2. Estadísticas de la recopilación de datos de un cristal de la CP nativa del WMV con el compuesto 28C6. Los valores entre paréntesis corresponden a los de alta resolución.

Cristal de CP nativa + compuesto 28C6	
<i>Recopilación de datos</i>	
Longitud de onda (Å)	0.979182
Grupo espacial	P6
Resolución (Å)	8-2.9 (3.47-3.28)
Dimensiones de la celda unidad	
<i>a, b, c</i> (Å)	220.463, 220.463, 316.169
α, β, γ (°)	90.0, 90.0, 120.0
R-meas (%)	19.7
$CC_{1/2}$ (%)	97.8 (45.4)
Completitud (%)	98.8 (97.5)
I/σ	3.61 (0.78)
Número de reflexiones únicas	263285 (41839)
Redundancia	10.55 (10.54)

7. DISCUSIÓN

En este trabajo se ha logrado producir la CP nativa y la marcada con SeMet. Se ha llevado a cabo su cristalización y se han obtenido cristales en cinco condiciones de los kits comerciales AmSO₄, Morpheus y Proplex. Se ha procedido a optimizar dichas condiciones para conseguir cristales que permitan obtener conjuntos de datos de mejor calidad. Se han pescado 113 cristales, a partir de los cuales se han obtenido patrones de difracción a una resolución máxima de 3,28 Å (**Tabla 2**).

Aunque el proceso de purificación de las proteínas ha permitido conseguir proteína apta para ensayos de cristalogénesis, no se ha realizado de manera óptima. Uno de los pasos en este proceso es la eliminación de la etiqueta de histidinas tras la cromatografía de afinidad, que no se ha llevado a cabo al no poderse digerir el extremo N-terminal con la proteasa TEV. Este enzima tiene una muy alta especificidad de corte (Cesaratto *et al.*, 2016) y, además, ha mostrado actividad en otros experimentos. Por lo tanto, el que la digestión no se haya producido puede ser indicativo de que el sitio de corte no está accesible por impedimento estérico (Waugh, 2011), es decir, que una parte de la CP impide que la proteasa TEV reconozca su secuencia de corte. Aun así, la etiqueta de histidinas queda expuesta, permitiendo purificar las proteínas por cromatografía de afinidad.

Se observó una menor expresión de la CP marcada con SeMet frente a la nativa (**Figura 8; Figura 9**) lo que es frecuente cuando se emplea la cepa auxótrofa B834(DE3), ya que el rendimiento de la proteína marcada suele disminuir en un 80% con respecto al de la proteína nativa (Walden, 2010).

En la cromatografía de intercambio iónico, la conductividad en la que eluye la CP nativa fue superior a la de la marcada con SeMet (**Figura 10C**; **Figura 11C**). Teniendo en consideración que el tampón empleado fue el mismo, es posible esto sea debido a diferencias del detector de conductividad por el uso de un *ÅKTA Purifier* diferente en estas purificaciones.

A pesar de la optimización de aquellos cristales con formas poliédricas (**Figura 12A**; **ANEXO 3**), que son la mejor opción para refinar en comparación con los cristales tipo agujas (McPherson & Cudney, 2014), en la difracción falta información correspondiente a la alta resolución. Es posible que ello se deba a una gran cantidad de solvente en la red cristalina, que impide un buen ordenamiento en el cristal (McPherson & Gavira, 2014). Esto tampoco se ha podido solucionar empleando los diferentes crioprotectores, los cuales pueden entrar en los canales de solvente dentro de la red cristalina (Pflugrath, 2015) e incrementar la calidad de la difracción (Heras & Martin, 2005). Por lo tanto, dado que el patrón de difracción de un cristal se correlaciona directamente con su grado de orden interno (McPherson & Gavira, 2014), las proteínas en el cristal no tienen el orden suficiente para difractar a alta resolución. Además de la cantidad de solvente, el sitio de restricción de la proteasa TEV y la etiqueta de histidinas podrían estar también interfiriendo en la organización interna del cristal.

En el procesamiento de los conjuntos de datos, las mejores estadísticas se han obtenido en un grupo espacial hexagonal P6 (**Tabla 2**). La CP del *Virus del mosaico de la papaya* ha sido resuelta por cristalografía de rayos X en un grupo espacial hexagonal (Yang *et al.*, 2012). Esto indica que se está progresando de manera correcta, dado que también es un virus filamentosos y flexible al igual que el WMV, aunque perteneciente a la familia *Alphaflexiviridae*, y que la estructura 3D de la CP es similar a la CP del WMV.

En cuanto a la calidad del procesamiento de los datos cristalográficos, uno de los parámetros que indica que se ha elegido correctamente el grupo espacial es el R-meas (Karplus & Diederichs, 2012) y debe tener un valor menor del 10% (**Tabla 1**), que significa que las reflexiones relacionadas por simetría difieren en menos de 10% en la baja resolución. Por tanto, el valor de 19,7 obtenido (**Tabla 2**) implica que, aunque el conjunto de datos se puede procesar en el grupo espacial P6, existen reflexiones que no provienen de este tipo de organización. Esto podría ser porque el cristal tiene más de una red cristalina.

A pesar de que el procesamiento no fuese óptimo, se hicieron varios intentos de resolver la estructura por reemplazamiento molecular usando Molrep y Phaser con diferentes modelos. Desafortunadamente, aún no se ha encontrado una solución.

8. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Teniendo en cuenta los objetivos propuestos en este trabajo, se ha podido llevar a cabo la purificación de la CP del WMV para su posterior cristalización. Sin embargo, no se ha determinado la interacción de los compuestos antivirales candidatos con la proteína por cristalografía de rayos X, debido a la no obtención de una solución por reemplazamiento molecular. A pesar de ello, se pueden extraer las siguientes conclusiones:

- La proteína marcada con SeMet no tiene el orden necesario a nivel atómico en los cristales para que estos difracten, a pesar de que macroscópicamente parecen ordenados. Se puede concluir, por tanto, que no se puede determinar la estructura con estos cristales.
- La presencia de gran cantidad de solvente en los cristales de CP nativa y, posiblemente, el sitio de corte de la proteasa TEV y la etiqueta de histidinas, impiden que las proteínas se organicen y se obtengan patrones de difracción a alta resolución. A pesar de ello, se ha observado que la condición óptima para la generación de cristales es la G6 del kit ProPlex; que un mayor tiempo de equilibrio de los cristales de CP nativa permite una mejor organización de las macromoléculas en el cristal; y que un 21% de sacarosa como crioprotector mejora los patrones de difracción.

Así, con el fin de conseguir un conjunto de datos con información a alta resolución y, por consiguiente, el segundo objetivo propuesto en este trabajo, se presentan algunas ideas para proseguir con el proyecto.

- Se podría realizar un método de remojo de los cristales en una mezcla de diferentes crioprotectores (Pflugrath, 2015), con el fin de eliminar el solvente y conseguir una mejora en la calidad de la difracción. Se emplearía un 21% de sacarosa junto con otros crioprotectores.
- Dada la posibilidad de que el sitio de corte de la proteasa TEV y la etiqueta de histidinas impidan a las proteínas organizarse del todo en el cristal, habría que intentar que el enzima proteolítico realice el corte. Para ello, se podría volver a empezar clonando el inserto de la CP con residuos adicionales entre el sitio de corte de la proteasa TEV y el extremo N-terminal de la proteína (Waugh, 2011).
- Otra opción es emplear una línea de los sincrotrones que permita realizar una deshidratación *in situ* del cristal, puesto que se ha observado que la deshidratación es beneficiosa para la calidad de difracción de los cristales de proteína (Sanchez-Weatherby & Moraes, 2016).

9. BIBLIOGRAFÍA

- Casaos Aznar, J. E. (2014). *Estudio del papel de los residuos E413, R422 y R430 en el proceso de oligomerización de la enzima AIF mediante cristalografía de rayos X* [Master's thesis, University of Zaragoza]. Repository of University of Zaragoza - ZAGUAN. <https://zaguan.unizar.es/record/16286?ln=es>
- Cesaratto, F., Burrone, O. R., & Petris, G. (2016). Tobacco Etch Virus protease: A shortcut across biotechnologies. *Journal of Biotechnology*, 231, 239–249. <https://doi.org/10.1016/j.jbiotec.2016.06.012>
- De Moya-Ruiz, C., Rabadán, P., Juárez, M., & Gómez, P. (2021). Assessment of the current status of potyviruses in watermelon and pumpkin crops in Spain: Epidemiological impact of cultivated plants and mixed infections. *Plants*, 10(1), 1–14. <https://doi.org/10.3390/plants10010138>
- DiMaio, F., Chen, C. C., Yu, X., Frenz, B., Hsu, Y. H., Lin, N. S., & Egelman, E. H. (2015). The molecular basis for flexibility in the flexible filamentous plant viruses. *Nature Structural and Molecular Biology*, 22(8), 642–644. <https://doi.org/10.1038/nsmb.3054>
- Domon, B., & Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science*, 312(5771), 212–217. <https://doi.org/10.1126/science.1124619>
- Dubach, V. R. A., & Guskov, A. (2020). The resolution in X-ray crystallography and single-particle cryogenic electron microscopy. *Crystals*, 10(7), 1–13. <https://doi.org/10.3390/cryst10070580>
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), 486–501. <https://doi.org/10.1107/S0907444910007493>
- ExPASy - ProtParam tool. <https://web.expasy.org/protparam/>
- Gadhve, K. R., Gautam, S., Rasmussen, D. A., & Srinivasan, R. (2020). Aphid transmission of potyvirus: The largest plant-infecting RNA virus genus. *Viruses*, 12(7). <https://doi.org/10.3390/v12070773>
- Heras, B., & Martin, J. L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D: Biological Crystallography*, 61(9), 1173–1180. <https://doi.org/10.1107/S0907444905019451>
- International Committee on Taxonomy of Viruses (ICTV). <https://talk.ictvonline.org/>. Retrieved June 2021.
- Jones, R. A. C., & Naidu, R. A. (2019). Global Dimensions of Plant Virus Diseases: Current Status and Future Perspectives. *Annual Review of Virology*, 6, 387–409. <https://doi.org/10.1146/annurev-virology-092818-015606>

- Kabsch, W. (2010). XDS. *Acta Crystallographica Section D Biological Crystallography*, 66(2), 125–132. <https://doi.org/10.1107/S0907444909047337>
- Karplus, P. A., & Diederichs, K. (2012). Linking crystallographic model and data quality. *Science*, 336(6084), 1030–1033. <https://doi.org/10.1126/science.1218231>
- Kendall, A., McDonald, M., Bian, W., Bowles, T., Baumgarten, S. C., Shi, J., ... Stubbs, G. (2008). Structure of Flexible Filamentous Plant Viruses. *Journal of Virology*, 82(19), 9546–9554. <https://doi.org/10.1128/jvi.00895-08>
- Lecoq, H., & Desbiez, C. (2012). Viruses of cucurbit crops in the Mediterranean Region. An Ever-Changing Picture. *Advances in Virus Research*, 84, 67–126. <https://doi.org/10.1016/B978-0-12-394314-9.00003-8>
- Mäkinen, K., & Hafrén, A. (2014). Intracellular coordination of potyviral RNA functions in infection. *Frontiers in Plant Science*, 5, 110. <https://doi.org/10.3389/fpls.2014.00110>
- Martínez-Ripoll, M. (n.d.). *Cristalografía*. <https://www.xtal.iqfr.csic.es/Cristalografia/>
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4), 658–674. <https://doi.org/10.1107/S0021889807021206>
- McPherson, A., & Cudney, B. (2014). Optimization of crystallization conditions for biological macromolecules. *Acta Crystallographica Section F: Structural Biology Communications*, 70(11), 1445–1467. <https://doi.org/10.1107/S2053230X14019670>
- McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications*, 70(1), 2–20. <https://doi.org/10.1107/S2053230X13033141>
- Nigam, D., LaTourrette, K., Souza, P. F. N., & García-Ruíz, H. (2019). Genome-Wide Variation in Potyviruses. *Frontiers in Plant Science*, 10, 1439. <https://doi.org/10.3389/fpls.2019.01439>
- Pflugrath, J. W. (2015). Practical macromolecular cryocrystallography. *Acta Crystallographica Section F: Structural Biology Communications*, 71(6), 622–642. <https://doi.org/10.1107/S2053230X15008304>
- Powell, H. R. (2017). X-ray data processing. *Bioscience Reports*, 37(5). <https://doi.org/10.1042/BSR20170227>
- Powell, H. R., Battye, T. G. G., Kontogiannis, L., Johnson, O., & Leslie, A. G. W. (2017). Integrating macromolecular X-ray diffraction data with the graphical user interface iMosflm. *Nature Protocols*, 12(7), 1310–1325. <https://doi.org/10.1038/nprot.2017.037>

- Rupp, B. (2010). *Biomolecular Crystallography. Principles, Practice, and Application to Structural Biology*. New York, USA: Garland Science.
- Sanchez-Weatherby, J., & Moraes, I. (2016). Crystal dehydration in membrane protein crystallography. *Advances in Experimental Medicine and Biology*, 922, 73–89. https://doi.org/10.1007/978-3-319-35072-1_6
- Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., & Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution*, 3(3), 430–439. <https://doi.org/10.1038/s41559-018-0793-y>
- Trebicki, P. (2020). Climate change and plant virus epidemiology. *Virus Research*, 286, 198059. <https://doi.org/10.1016/j.virusres.2020.198059>
- Van Regenmortel, M. H. V. (2011). Virus Species. In M. Tibayrenc (Ed.), *Genetics and Evolution of Infectious Diseases* (pp. 3–19). Elsevier. <https://doi.org/10.1016/B978-0-12-384890-1.00001-7>
- Venkataraman, S., Savithri, H. S., & Murthy, M. R. N. (2018). Structural Aspects of Plant Viruses. In R. K. Gaur, S. Khurana, & Y. Dorokhov (Eds.), *Plant Viruses* (pp. 3–26). Boca Raton, USA: CRC Press. <https://doi.org/10.1201/b22221-1>
- ViralZone. <https://viralzone.expasy.org/>. Retrieved June 2021.
- Walden, H. (2010). Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), 352–357. <https://doi.org/10.1107/S0907444909038207>
- Wang, D., & Li, G. (2017). Host Reaction of Watermelon mosaic virus Isolates Infecting Melon from Different Geographical Origins in Xinjiang of China. *Horticultural Plant Journal*, 3(1), 23–28. <https://doi.org/10.1016/j.hpj.2017.01.010>
- Waugh, D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification*, 80(2), 283–293. <https://doi.org/10.1016/j.pep.2011.08.005>
- Webster, J., & Oxley, D. (2012). Protein identification by MALDI-TOF mass spectrometry. *Methods in Molecular Biology*, 800, 227–240. https://doi.org/10.1007/978-1-61779-349-3_15
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., ... Wilson, K. S. (2011, April). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 235–242. <https://doi.org/10.1107/S0907444910045749>
- Yang, S., Wang, T., Bohon, J., Gagné, M. È. L., Bolduc, M., Leclerc, D., & Li, H. (2012). Crystal structure of the coat protein of the flexible filamentous Papaya mosaic virus. *Journal of Molecular Biology*, 422(2). <https://doi.org/10.1016/j.jmb.2012.05.032>

Zamora, M., Méndez-López, E., Agirrezabala, X., Cuesta, R., Lavín, J. L., Amelia Sánchez-Pina, M., ... Valle, M. (2017). Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Science Advances*, 3(9), eaao2182. <https://doi.org/10.1126/sciadv.aao2182>

Zamora Porras, M. (2018). *Determinación estructural de virión del virus del mosaico de la sandia por crío-microscopía electrónica* [Doctoral thesis, Basque Country University]. Dialnet. <https://dialnet.unirioja.es/servlet/tesis?codigo=223533>

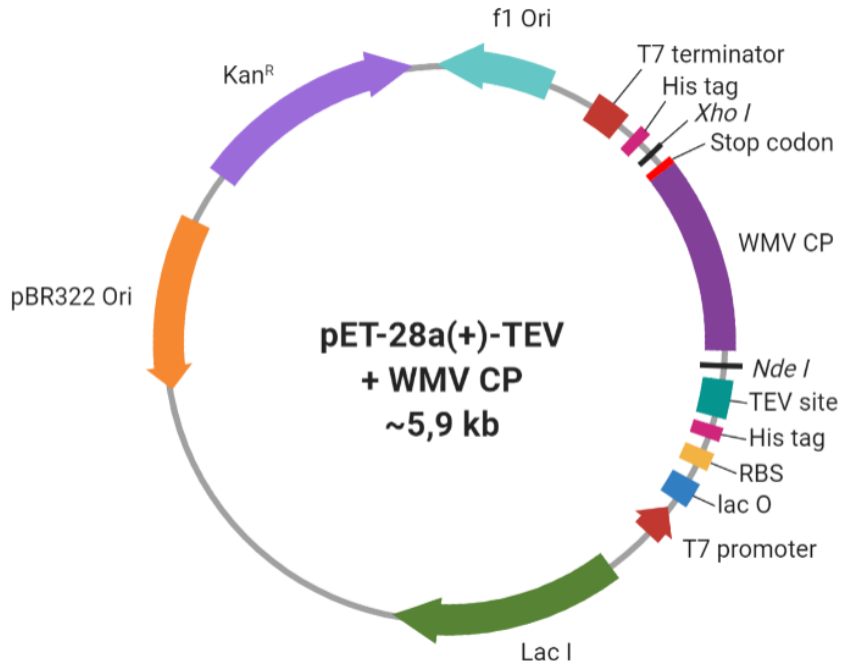
ANEXO 1. Secuencia de aminoácidos de la proteína de la cápside del WMV.

>WMV CP 71-247

MLPTVGGKIILSLDHLLEYKPNQVDLFNTRATKTQFESWYSAVKVEYDLNDEQMGVIMNGFMVWCI
DNGTSPDVNGVWVMMDGEEQVEYPLKPIVENAKPTLRQIMHHFSDAAEAYIEMRNSESPYMPRYGL
LRNLRDRELARYAFDFYEVTSKTPNRAREAIAQMKAALAGINSRL

ANEXO 2. (A) Vector pET-28a(+)-TEV con inserto de la CP del WMV. Se han omitido sitios de restricción de otros enzimas que no sean los de las empleadas en la inserción de la secuencia correspondiente a la proteína. (B) Secuencia de aminoácidos en el extremo N-terminal de la CP del WMV con la etiqueta de histidinas y el sitio de corte para la proteasa TEV.

A



B



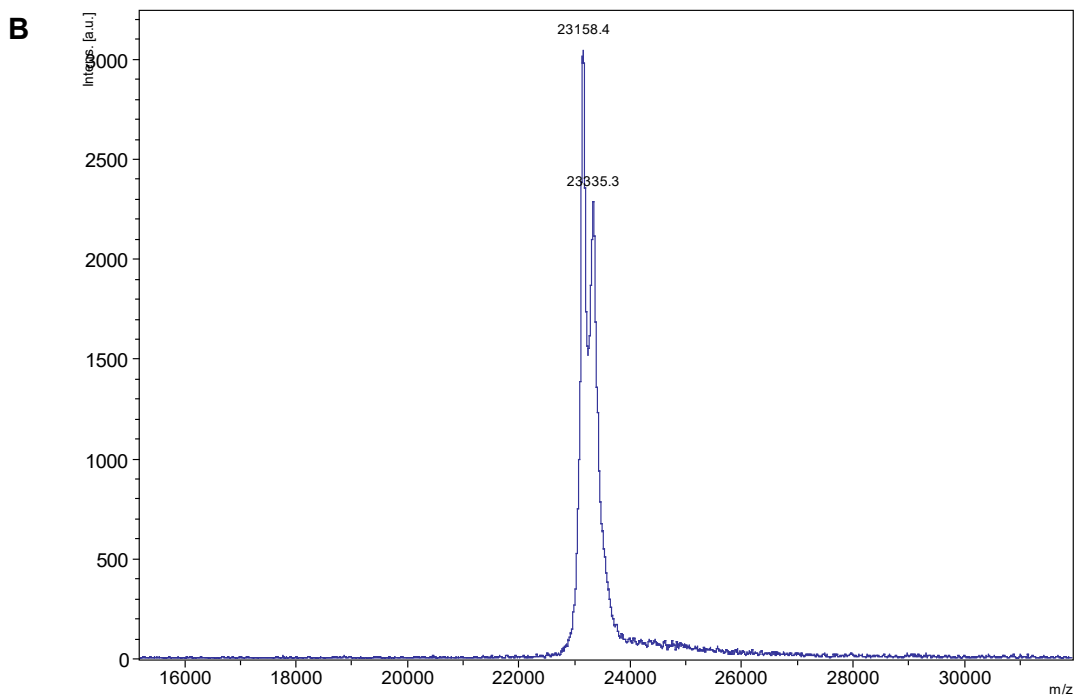
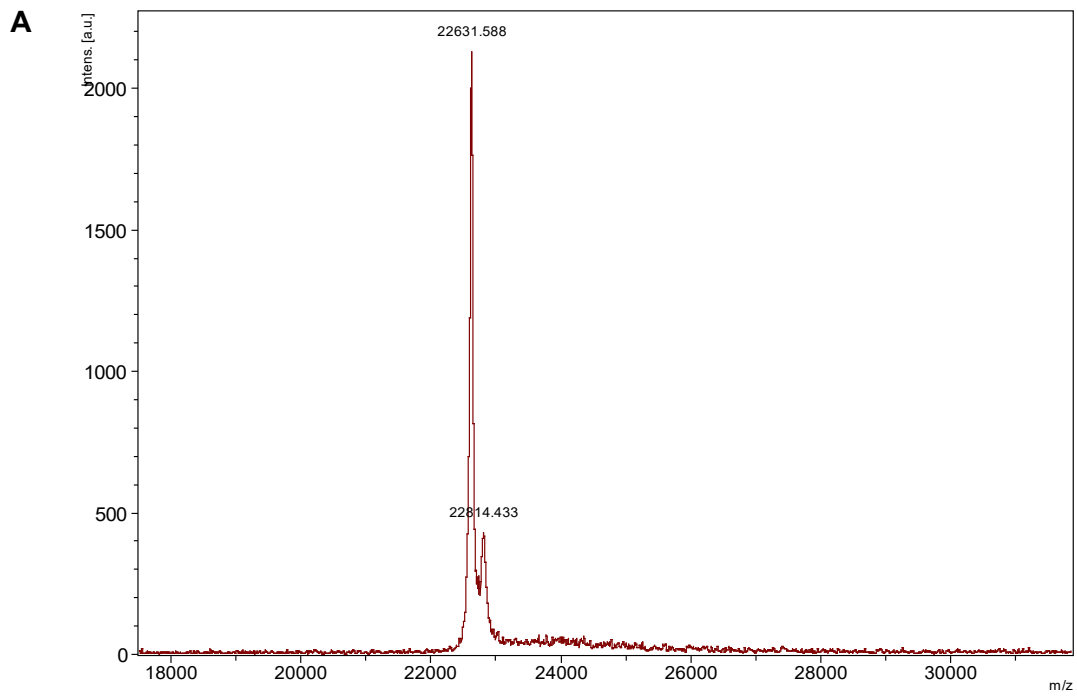
ANEXO 3. Placas de refinamiento empleadas con sus condiciones.

Placa	Fila	Tampón	Aditivo	Sal	Agente precipitante	Condición de kit para refinamiento	CP del WMV	Ligando	
1	A	0,1M Buffer 2 pH=7,5	0,1M Carboxylic acids		15% PEG + 15% Glycerol	G7 (MORPHEUS)	nativa	25D8	
	B				20% PEG + 10% Glycerol				
	C				25% PEG + 5% Glycerol				
	D				30% PEG				
	E		0,1M aminoacids		15% PEG + 15% Glycerol	H7 (MORPHEUS)			
	F				20% PEG + 10% Glycerol				
	G				25% PEG + 5% Glycerol				
	H				30% PEG				
2	A			0,20M Cesium chloride	2,00M AmSO ₄	B2 (AmSO ₄)	nativa	25D8	
	B				2,20M AmSO ₄				
	C			0,20M Cesium chloride	2,20M AmSO ₄				
	D				2,40M AmSO ₄				
	E	0,10M MES pH=6,5		1,40M Magnesium sulfate hydrate		G6 (PROPLEX)			
	F			1,60M Magnesium sulfate hydrate					
	G			1,80M Magnesium sulfate hydrate					
	H			2,00M Magnesium sulfate hydrate					
3	A	2,60M Sodium formate		0,10M Sodium acetate trihydrate pH=4,8		G6 (PROPLEX)	nativa	25D8	
	B	2,40M Sodium formate							
	C	2,20M Sodium formate							
	D								
	E	2,60M Sodium formate		0,10M Sodium citrate tribasic dihydrate pH=5,0					
	F	2,40M Sodium formate							
	G								
	H	2,20M Sodium formate							
4	A			0,20M Cesium chloride	2,80M AmSO ₄	B2 (AmSO ₄)	nativa	25D8	
	B				2,60M AmSO ₄				
	C				2,40M AmSO ₄				
	D								
	E			0,10M Cesium chloride	2,80M AmSO ₄				
	F				2,60M AmSO ₄				
	G				2,40M AmSO ₄				
	H								
5	A	2,20M Sodium formate		0,10M Sodium citrate tribasic dihydrate pH=5,0		G6 (PROPLEX)	nativa	25D8	
	B	2,00M Sodium formate							
	C	1,80M Sodium formate							
6	A	2,60M Sodium formate		0,10M Sodium acetate pH=4,6		G6 (PROPLEX)	marcada con SeMet	25D8 y 3G9	
	B	2,40M Sodium formate							
	C	2,20M Sodium formate							
	D								
	E	2,60M Sodium formate		0,10M Sodium citrate tribasic dihydrate pH=5,0					
	F	2,40M Sodium formate							
	G								
	H	2,20M Sodium formate							

ANEXO 3 (continuación). Placas de refinamiento empleadas con sus condiciones.

Placa	Fila	Buffer	Aditivo	Sal	Agente precipitante	Condición de kit para refinamiento	CP del WMV	Ligando
7	A			0,20M Potassium citrate tribasic monohydrate pH 8,3	14% PEG 3350	H11 (PEG&ION)	marcada con SeMet	25D8 y 3G9
	B				16% PEG 3350			
	C				18% PEG 3350			
	D				20% PEG 3350			
	E				22% PEG 3350			
	F				24% PEG 3350			
	G				26% PEG 3350			
	H							
8	A			0,10M Sodium acetate trihydrate pH 4,8	2,60M Sodium formate	G6 (PROPLEX)	nativa	25D8 y 3G9
	B				2,40M Sodium formate			
	C				2,20M Sodium formate			
	D				2,20M Sodium formate			
9	A	2,60M Sodium formate		0,10M Sodium acetate trihydrate pH 4,8		G6 (PROPLEX)	nativa	28C6
	B	2,40M Sodium formate						
	C							
	D	2,20M Sodium formate						
	E	2,60M Sodium formate		0,10M Sodium citrate tribasic dihydrate pH 5,0				
	F	2,40M Sodium formate						
	G							
	H	2,20M Sodium formate						
10	A	2,60M Sodium formate		0,10M Sodium acetate trihydrate pH 4,8		G6 (PROPLEX)	nativa	3G9
	B	2,40M Sodium formate						
	C							
	D	2,20M Sodium formate						
	E	2,60M Sodium formate		0,10M Sodium citrate tribasic dihydrate pH 5,0				
	F	2,40M Sodium formate						
	G							
	H	2,20M Sodium formate						

ANEXO 4. Resultados de los análisis de espectrometría de masas (A) de la CP nativa del WMV y (B) de la CP marcada con SeMet.



ANEXO 5. Condiciones en las que se han obtenido cristales de CP nativa al cabo de una semana y que han sido optimizadas.

Kit comercial	Condición	Tampón	Aditivo	Sal	Agente precipitante
AmSO₄ Suite	B2			0,20M Cesium chloride	2,20M AmSO ₄
Morpheus	G7	0,10M Buffer 2 pH 7,5 (1,0M Sodium HEPES pH 7,5 + MOPS)	0,10M Carboxylic acids		30% GOL_P4K
	H7		0,10M Amino acids		
ProPlex	G6	0,10M Sodium acetate pH 5,0		2,00M Sodium formate	