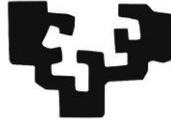


eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

**DETERMINACIÓN ESTRUCTURAL DEL VIRIÓN DEL *VIRUS*
DEL MOSAICO DE LA SANDÍA POR CRIO-MICROSCOPIA
ELECTRÓNICA**

MIGUEL ZAMORA PORRAS

TESIS DOCTORAL

JUNIO 2018



Universidad del País Vasco Euskal Herriko Unibertsitatea

CICbioGUNE
CENTER FOR COOPERATIVE RESEARCH IN BIOSCIENCES

**DETERMINACIÓN ESTRUCTURAL DEL VIRIÓN DEL *VIRUS*
DEL MOSAICO DE LA SANDÍA POR CRIO-MICROSCOPIA
ELECTRÓNICA**

Memoria presentada por

Miguel Zamora Porras

para optar al grado de Doctor

Trabajo dirigido por:

Dr. Mikel Valle

AGRADECIMIENTOS

Agradezco al Ministerio de Economía y Competitividad por la financiación aportada para la realización de esta tesis doctoral. Quiero agradecer a Mikel Valle por aceptarme en su grupo y a todos los miembros del laboratorio de microscopía electrónica por todo lo que me han enseñado y los momentos que hemos compartido. Gracias a todos.

ÍNDICE GENERAL

Capítulo 1: INTRODUCCIÓN.....	1
1.1 Características generales de los virus.....	3
1.2 Diversidad y clasificación de los virus.....	5
1.3 Los virus filamentosos.....	6
1.3.1 Las cápsides helicoidales.....	8
1.4 Virus filamentosos flexibles.....	9
1.4.1 Estructura del virión de los virus filamentosos flexibles.....	12
1.4.2 <i>Virus del mosaico de la sandía</i> y la familia <i>Potyviridae</i>	14
1.4.3 Relevancia económica y social de los virus filamentosos flexibles.....	19
1.5 Microscopía electrónica de transmisión (MET) en biología estructural.....	21
1.5.1 El microscopio electrónico de transmisión.....	22
1.5.2 Estudio de macromoléculas biológicas por MET.....	23
1.5.3 Reconstrucción tridimensional.....	23
1.5.4 Crio-ME en el estudio de los virus filamentosos flexibles.....	25
Capítulo 2: OBJETIVOS E HIPÓTESIS.....	27
Capítulo 3: MATERIALES Y MÉTODOS.....	31
3.1 Inoculación y purificación de viriones de WMV.....	33
3.2 Preparación de muestras para microscopía electrónica.....	34
3.2.1 Tinción negativa.....	34
3.2.2 Vitrificación.....	34
3.3 Obtención de imágenes en el microscopio electrónico de transmisión.....	35
3.3.1 Adquisición de imágenes por tinción negativa.....	35
3.3.2 Adquisición de imágenes en crio-ME.....	35

3.4 Procesamiento digital de imágenes.....	36
3.4.1 Corrección de defectos causados por el movimiento de la muestra durante la adquisición de imágenes.....	36
3.4.2 Estimación de parámetros de la CTF.....	36
3.4.3 Alineamiento y reconstrucción helicoidal.....	37
3.4.3.1 Selección de partículas.....	37
3.4.3.2 Ordenamiento y clasificación en 2D de segmentos helicoidales.....	38
3.4.3.3 Reconstrucción tridimensional con simetría helicoidal.....	39
3.4.3.4 Clasificación 3D.....	40
3.5 Construcción del modelo atómico de la CP de WMV.....	40
3.6 Análisis de homología estructural.....	41
3.6.1 Cálculo de distancias estructurales.....	41
3.7 Alineamiento de secuencias.....	42
3.8 Códigos de acceso.....	43
Capítulo 4: RESULTADOS.....	45
4.1 Análisis estructural de viriones de WMV por crio-ME.....	47
4.1.1 Reconstrucción del virión de WMV.....	48
4.1.2 Unidad asimétrica y ensamblaje del virión.....	55
4.2 Generación del modelo atómico de la proteína de la cápside y RNA genómico de WMV.....	58
4.3 Análisis de las interacciones CP - CP.....	61
4.4 Análisis de las interacciones CP - RNA.....	66
4.5 Análisis de la similitud estructural entre las proteínas de la cápside de los virus filamentosos flexibles.....	69
4.5.1 Mecanismo de unión a RNA común en Virus Filamentosos Flexibles.....	73
4.6 Análisis de la similitud estructural entre nucleoproteínas de virus de ssRNA.....	75
Capítulo 5: DISCUSIÓN.....	79

5.1 Simetría helicoidal en virus filamentosos flexibles.....	82
5.2 Estructura de la proteína de la cápside y ensamblaje del virión de WMV.....	83
5.3 Homología estructural en virus filamentosos flexibles.....	89
5.4 Homología estructural entre nucleoproteínas de virus ssRNA.....	92
5.4.1 Nueva familia de plegamiento para NPs de virus ssRNA.....	98
Capítulo 6: CONCLUSIONES.....	101
Anexo I.....	105
Anexo II.....	111
Anexo III.....	115
BIBLIOGRAFÍA.....	119
Publicaciones.....	133

ÍNDICE DE TABLAS Y FIGURAS

Figura 1.1: Imágenes de microscopía electrónica de partículas virales (viriones) con diferentes morfologías..... 4

Figura 1.2: Representaciones de diferentes tipos de virus filamentosos modificadas a partir de imágenes de ViralZone..... 7

Figura 1.3: Definición de la simetría helicoidal de una hélice levógira..... 9

Figura 1.4: Viriones de algunas especies representativas de las distintas familias del grupo de virus filamentosos flexibles visualizadas mediante microscopía electrónica empleando tinción negativa..... 10

Figura 1.5: Distribución de especies dentro del grupo de virus filamentosos flexibles..... 11

Figura 1.6: Estructura del virión del género *Potexvirus* en comparación con el virión de TMV..... 13

Figura 1.7: Modos de transmisión horizontal y vertical de los miembros de la familia *Potyviridae*..... 15

Figura 1.8: Esquema de la organización del genoma y modo de traducción de los miembros del género *Potyvirus*..... 16

Figura 1.9: Representación esquemática de los diferentes eventos durante la infección de una célula vegetal por un potyvirus..... 17

Figura 1.10: Ejemplos de síntomas provocados por los virus filamentosos flexibles, especialmente el género *Potyvirus*..... 18

Figura 3.1: Comparación de métodos de selección de filamentos..... 38

Figura 4.1: Micrografía de tinción de negativa de filamentos de WMV..... 47

Figura 4.2: Selección de partículas y procesamiento 2D de segmentos de WMV..... 50

Figura 4.3: Mapa de densidad 3D de un segmento del virión de WMV..... 51

Figura 4.4: Corte longitudinal del mapa de densidad 3D del virión de WMV..... 52

Figura 4.5: Curvas FSC de los volúmenes reconstruidos en los procesamientos helicoidales para comparar los métodos de selección de partículas..... 53

Figura 4.6: Simetrización del mapa de densidad 3D resultante del procesamiento helicoidal.....	54
Figura 4.7: Ensamblaje del virión de WMV.....	56
Figura 4.8: Disposición del extremo C-terminal en el interior del virión de WMV.....	57
Figura 4.9: Perfil de predicción de desorden de la proteína de la cápside de WMV calculado con DISOPRED3.....	58
Figura 4.10: Modelo atómico de proteína de la cápside de WMV unido a un fragmento de RNA genómico.....	59
Figura 4.11: Región del modelo atómico correspondiente a la zona central ordenada de la proteína de la cápside de WMV.....	60
Figura 4.12: Modelos atómicos de la proteína de la cápside de WMV mostrando la interacción del extremo N-terminal de la CP de referencia con las subunidades adyacentes.....	62
Figura 4.13: Detalle de densidades correspondientes a cadenas laterales de CP de WMV que podrían establecer interacciones con otros residuos de las CP adyacentes.....	64
Figura 4.14: Unión del RNA genómico a la proteína de la cápside de WMV.....	67
Figura 4.15: Detalle de residuos posiblemente relacionados con la interacción de la CP con el RNA.....	68
Figura 4.16: Disposición del extremo N-terminal en proteínas de la cápside de virus filamentosos flexibles.....	70
Figura 4.17: Representación del alineamiento de modelos atómicos de CPs realizado por “TM-align”.....	72
Figura 4.18: Comparación de la zona central de la proteína de la cápside de los modelos atómicos disponibles para virus filamentosos flexibles.....	73
Figura 4.19: Interacción RNA genómico (ssRNA) con la proteína de la cápside en virus filamentosos flexibles.....	74
Figura 4.20: Secuencia consenso para zonas concretas de las CPs de las diferentes familias de virus filamentosos flexibles.....	75
Figura 4.21: Comparación de nucleoproteínas de algunos virus ssRNA con la CP de WMV.....	77

Figura 4.22: Representación del alineamiento de modelos atómicos de NPs realizado por “TM-align”	78
Figura 5.1: Extremos terminales de la proteína de la cápside de WMV	84
Figura 5.2: Comparación de extremos terminales entre <i>Potyvirus</i> y <i>Potexvirus</i>	86
Figura 5.3: Comparación por alineamiento estructural de la zona central ordenada de las proteínas de la cápside de WMV y PepMV	90
Figura 5.4: Comparación de morfología y RNP de WMV, RVFV e Influenza A.....	94
Figura 5.5: Modelo evolutivo de nucleoproteínas de virus ssRNA.....	96
Figura 5.6: Árbol filogenético sin raíz basado en similitud estructural de varias nucleoproteínas de virus ssRNA.....	99
Tabla 1.1: Ejemplos de patógenos pertenecientes al grupo de virus filamentosos flexibles que infectan los cultivos más relevantes a nivel económico y social.....	20
Tabla 4.1: Resultado de la clasificación helicoidal 3D de RELION2 usando todo el conjunto de segmentos empleados para la obtención del mapa de densidad 3D inicial.....	54
Tablas 4.2 y 4.3: Valores obtenidos con el servidor “TM-align” para distintas comparaciones de modelos atómicos de proteínas de la de la cápside entre WMV y varias especies del género <i>Potexvirus</i>	71

ABREVIATURAS

2D: Bidimensional

3D: Tridimensional

Å: Angstrom

BaMV: *Virus del mosaico del Bambú (Bamboo mosaic virus)*

CCD (cámara): Dispositivo de carga acoplada (charge-coupled device)

CP: Proteína de la cápside (coat protein)

crio-ME: Crio-microscopía electrónica

C-t/C-terminal: Extremo carboxilo terminal

CTF: Función de transferencia de contraste (Contrast transfer function)

CTV: *Virus de la tristeza de los cítricos (Citrus tristeza virus)*

DNA: Ácido desoxirribonucleico

dsRNA: RNA de doble hebra

dsDNA: DNA de doble hebra

$e^{-}/\text{Å}^2$: Electrones por angstrom cuadrado

EDTA: Ácido etilendiaminotetraacético

EMDB: "Electron Microscopy Data Bank"

FSC: "Furrier Shell Correlation"

g (centrifugación):

g (masa): Gramo

°C: Grados centígrados

g/mL: Gramos por mililitro

ICTV: Comité Internacional de Taxonomía de los Virus (International Comitee on Taxonomy of Viruses)

kb: Kilobase

kDa: Kilodalton

kV: Kilovoltio

MET: Microscopía electrónica de transmisión

μL: Microlitro

μm: Micrómetro

mL: Mililitro

mM: Nanomolar

Ni: Subunidad de referencia

nm: Nanómetro

NP: Nucleoproteína

nsNSV: Virus de RNA de hebra sencilla y polaridad negativa con genoma no segmentado

N-t/N-terminal: Extremo amino terminal

ORF: Fase de lectura abierta (Open Reading Frame))

PDB: "Protein Data Bank"

PepMV: *Virus del mosaico del pepino (Pepino mosaic virus)*

PPV: *Virus de la Sharka (Plum pox virus)*

p/v: relación peso/volumen

PVX: *Virus X de la patata (Potato virus X)*

PVY: *Virus Y de la patata (Potato virus Y)*

RdRp: RNA polimerasa dependiente de RNA (RNA-dependent RNA polymerase)

RNA: Ácido ribonucleico

RNAg: RNA genómico

RNP: Ribonucleoproteína

RVFV: *Virus de la fiebre del Valle del Rift (Rift Valley fever virus)*

SMV: *Virus del mosaico de la soja (Soybean mosaic virus)*

sNSV: Virus de RNA de hebra sencilla y polaridad negativa con genoma segmentado

SPA: Análisis de partículas individuales (Single particle analysis)

(+)ssRNA: RNA de hebra sencilla y polaridad positiva

(-)ssRNA: RNA de hebra sencilla y polaridad negativa

ssDNA: DNA de hebra sencilla

ssRNA: RNA de hebra sencilla

TEV: *Virus del tabaco etch (Tobacco etch virus)*

TMV: *Virus del mosaico del tabaco (Tobacco mosaic virus)*

TSWT: *Virus del bronceado del tomate (Tomato spotted wilt tospovirus)*

VLP: Partículas similares a viriones

v/v: relación volumen/volumen

WMV: *Virus del mosaico de la sandía*

CAPITULO 1
INTRODUCCIÓN

1.1. Características generales de los virus

Los virus son entidades biológicas parasitarias consideradas organismos acelulares generadores de partículas virales infectivas o viriones, capaces de infectar y multiplicarse en el interior de una célula hospedadora (Forterre 2010). Los virus poseen un genoma propio que codifica para diferentes proteínas necesarias durante su ciclo de vida, presentando dos etapas diferenciadas: una fase extracelular y otra fase en el interior de la célula hospedadora (infección).

Los virus son parásitos celulares estrictos ya que no disponen de la maquinaria necesaria para su multiplicación e infectan todos los tipos de organismos existentes, desde bacterias hasta mamíferos. En general, la estrategia común en todos los virus es la siguiente: el virión debe reconocer a la célula hospedadora e interactuar con su superficie, y mediante diversos mecanismos el genoma del virus se introduce en el interior de la célula donde será replicado y sintetizará todas las proteínas necesarias para ensamblar nuevos viriones que saldrán de la célula hospedadora para repetir el ciclo. La infección comienza cuando el genoma del virus se introduce en la célula. Este toma el control de la misma y pone a su disposición la maquinaria celular para sintetizar todas las proteínas virales necesarias para llevar a cabo su replicación y completar su ciclo de vida. En mayor o menor medida, el genoma del virus altera el metabolismo de la célula huésped inactivando o incluso destruyendo en algunos casos el genoma de la propia célula. De este modo, tras la infección, la célula hospedadora es transformada en una “fábrica viral” encargada de sintetizar todos los componentes del virus, transformándose en un organismo distinto que constituye la fase celular del virus (Forterre 2010, Forterre and Prangishvili 2009). Dependiendo del tipo de virus, la infección puede derivar en distintas consecuencias para la célula hospedadora: i) la célula es destruida y muere, ii) la célula sobrevive pero permanece infectada persistiendo la multiplicación del virus, iii) la célula sobrevive y el genoma del virus permanece latente en el interior sin producir nuevas partículas virales, iv) la célula es inmortalizada por la acción del virus convirtiéndose en una célula tumoral.

Durante la infección, diferentes componentes del virus dentro de la célula se ensamblarán empaquetando el genoma del virus replicado para constituir viriones o partículas virales infectivas que serán liberadas para infectar nuevas células. Las partículas virales son las unidades biológicas más abundantes del planeta, están presentes en todos los entornos y presentan una morfología muy variable dependiendo del tipo de virus (Poranen, Daugelavicius, and Bamford 2002). Dichas partículas virales tienen dos componentes comunes en todos los virus: el material genético que constituye el genoma del virus y un

recipiente que lo empaqueta, generalmente de naturaleza proteica, denominado cápside. La forma y tamaño de las cápsides depende del tipo de virus. Estas pueden presentar una forma esférica o cuasi-esférica (icosaédricas), baciliforme o filamentosa y están construidas por múltiples copias de una o varias proteínas virales que se ensamblan de manera ordenada. Las cápsides icosaédricas pueden además estar envueltas por una membrana lipídica, o presentar una cola proteica como en el caso de algunos bacteriófagos. Su tamaño puede variar desde los 17 nm de diámetro en los virus más pequeños como Circovirus porcino, hasta 400 nm de diámetro en el caso de los virus más grandes como Mimivirus, y puede llegar hasta micras de longitud en el caso de los virus filamentosos (Figura 1.1). Otros virus no presentan cápsides proteicas, sino envueltas lipídicas pleomórficas sustraídas de la célula hospedadora que contienen el genoma del virus empaquetado por nucleoproteínas virales (Rossmann 2013, Reddy and Johnson 2005).

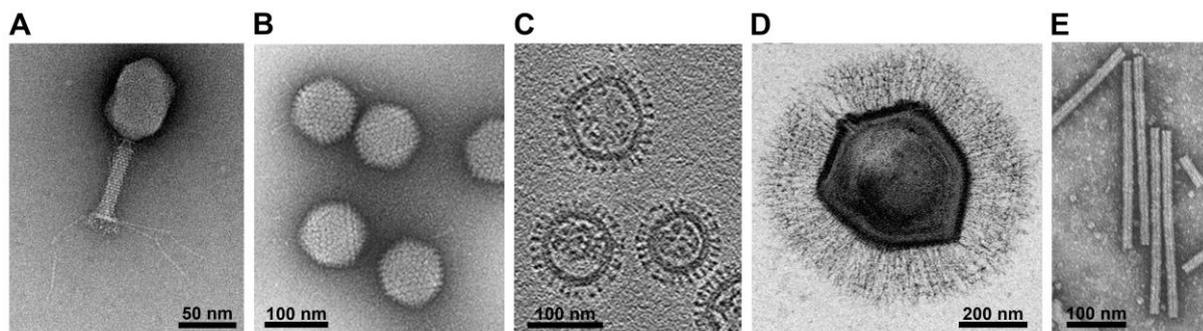


Figura 1.1: Imágenes de microscopía electrónica de partículas virales (viriones) con diferentes morfologías. **(A)** Cápside icosaédrica con cola del bacteriófago T4, (Knott and Genoud 2013). **(B)** Cápsides icosaédricas de un Adenovirus no-humano (Menéndez-Conejero et al. 2017). **(C)** Partículas virales pleomórficas de *Influenza A* (gripe) (Harris et al. 2006). **(D)** Virión de Mimivirus (Ghigo et al. 2008). **(E)** Partículas virales filamentosas del *Virus del Mosaico del Tabaco* (Gergerich and Dolja 2006).

El genoma del virus contenido en cada virión presenta sólo una naturaleza, en forma de DNA o RNA, de hebra simple o doble, circular o lineal; puede estar codificado por una sola molécula de ácido nucleico o por varias, como en el caso de los virus con genoma segmentado como el virus de la gripe. El tipo de material genético del virus determinará las peculiaridades de la replicación en la célula hospedadora; por ejemplo, algunos virus que presentan un genoma de DNA de doble hebra pueden replicar su material genético y transcribirlo usando la maquinaria de la célula hospedadora, en cambio, otros virus como aquellos con genoma de RNA de hebra simple necesitan codificar en su genoma genes para su propia maquinaria de replicación (Modrow et al. 2010).

1.2. Diversidad y clasificación de los virus

Los virus son muy diversos en cuanto a la organización y estructura de su genoma, las proteínas que codifican, sus mecanismos de replicación y la relación con los organismos que infectan. Son varias las hipótesis sobre el origen de los virus, entre ellas la sustentada por evidencias que sugieren que el origen de estos es anterior a la vida celular existente hoy en día, y polifilético, a diferencia de los organismos unicelulares cuyo origen es monofilético (Bamford 2003). Independientemente de su origen, los virus son entidades biológicas que han evolucionado a la par que los hospedadores que infectan, dando lugar a la enorme diversidad de virus actualmente presentes en la biosfera (Koonin, Dolja, and Krupovic 2015).

La necesidad de clasificar a los diferentes virus conocidos no sólo está respaldada por la necesidad humana de organización, sino que es una herramienta científica para la comparación y extrapolación de información útil de una especie a otra o de una familia de virus a otra. Sin embargo, la taxonomía de los virus difiere en aspectos fundamentales de la taxonomía de los organismos celulares, ya que los virus carecen de genes universales, como los genes del RNA ribosómico en organismos celulares, que puedan ser usados para construir una filogenia estructurada en la que todos los virus conocidos puedan encajar (Simmonds et al. 2017).

Dos esquemas en el modo de clasificación son los predominantes a la hora de identificar las agrupaciones taxonómicas de los virus conocidos. Uno de ellos es la clasificación Baltimore que separa a los virus en 7 grupos, en base a la naturaleza de su genoma y su modo de replicación (número de hebras y en caso de hebra sencilla si esta puede ser traducida directamente o no). Los grupos son DNA de doble hebra (dsDNA: grupo I), DNA hebra sencilla (ssDNA: grupo II), RNA de doble hebra (dsRNA: grupo III), RNA de hebra sencilla positiva ((+)ssRNA: grupo IV), RNA hebra sencilla negativa ((-)ssRNA: grupo V), RNA de hebra sencilla con intermediario de DNA (grupo VI) y DNA de doble hebra con intermediario de RNA (grupo VII) (Baltimore 1971). Sin embargo, esta clasificación es demasiado escueta y no nos aporta mucha información dentro de los grupos establecidos. El segundo esquema es la clasificación llevada a cabo por el Comité Internacional de Taxonomía de los Virus (ICTV), que intenta desarrollar una taxonomía general aceptada a nivel internacional que agrupe a todos los virus conocidos dentro de los taxones especie, género, familia y orden. Esta clasificación está basada en el análisis de secuencias génicas de genes equivalentes entre especies, la morfología y tamaño de los viriones, la estructura de su genoma, la presencia o ausencia de envuelta lipídica, las propiedades serológicas del virus o sus propiedades biológicas en el contexto de la infección y la interacción con el hospedador

(Fauquet 2008). El conocimiento disponible de los virus conocidos junto con la acumulación gradual de información de secuencias génicas pertenecientes a distintas especies de virus en el contexto de la clasificación Baltimore ha permitido clasificar a la mayoría de virus conocidos en 134 familias (<https://talk.ictvonline.org/taxonomy/>). A diferencia de los progresos en agrupar los diferentes virus en familias, agrupar las diferentes familias supone un reto más complicado, debido principalmente a la ausencia de similitud de secuencia entre familias de virus. Aun así, se han establecido 9 órdenes que agrupan 48 familias de virus según el último informe de ICTV. Sin embargo, esta clasificación está constantemente bajo revisión e influenciada por los avances en el conocimiento de los virus ya ubicados en ella, así como por el descubrimiento de nuevos virus.

El incremento de estructuras 3D disponibles para viriones y proteínas virales de diferentes especies ha posibilitado un tercer método de clasificación que propone definir linajes virales basados en la similitud estructural entre proteínas principales de la cápside de los diferentes virus y las rutas de ensamblaje de las mismas. Sorprendentemente, las proteínas que conforman las cápsides de los virus de familias muy distantes, con diferente tipo de genoma y que infectan hospedadores muy distintos, presentan una estructura 3D con una topología muy similar que sugiere una relación filogenética entre ellos. En base a estas similitudes se han establecido cuatro linajes principales definidos por varios dominios de plegamiento: “PRD1-like”, “Picorna-like”, “HK97-like” y “BTV-like” (Abrescia et al. 2012, Nasir and Caetano-Anollés 2017).

1.3. Los virus filamentosos

Los virus filamentosos son aquellos que presentan viriones con morfología cilíndrica de longitud variable formados por cápsides de nucleoproteínas que interaccionan directamente con el genoma del virus. La organización de estas proteínas empaquetando el genoma normalmente sigue una distribución helicoidal dando lugar a la morfología alargada de las partículas virales que les otorga su apariencia filamentososa (Figura 1.2). Los virus filamentosos están distribuidos en familias poco relacionadas aparentemente e infectan hospedadores muy distintos (Stubbs and Kendall 2012). La mayoría de estos virus son patógenos de plantas, pertenecientes a distintas familias como *Virgaviridae* (Namba and Stubbs 1986) o *Potyviridae* (Kendall et al. 2008), sin embargo, entre los bacteriófagos también se encuentran algunos miembros, como la familia *Inoviridae* (Marvin et al. 2006). También entre los virus que infectan animales se encuentran viriones filamentosos, como el caso del virus del ébola

de la familia *Filoviridae*, que presenta una nucleocápside con organización helicoidal empaquetando el genoma del virus rodeada de una envuelta lipídica (Bharat et al. 2011) (Figura 1.2).

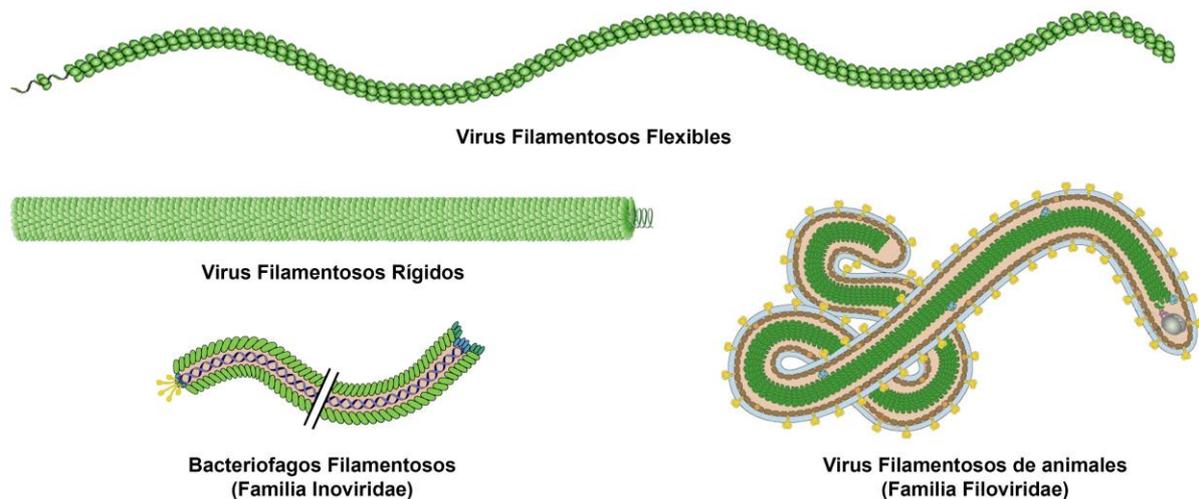


Figura 1.2: Representaciones de diferentes tipos de virus filamentosos modificadas a partir de imágenes de ViralZone (<https://viralzone.expasy.org/>) (Hulo et al. 2011). Los viriones filamentosos de cápsides con organización helicoidal presentan un elemento común básico en todos los casos, la proteína de la cápside (coloreada en verde) empaquetando el genoma del virus. Dependiendo del tipo de virus, el virión puede ser más complejo presentando otros elementos adicionales como en la familia Filoviridae, cuyos viriones están envueltos por una membrana y presentan proteínas adicionales.

La mayoría de los virus filamentosos son virus de plantas, y entre ellos el mejor caracterizado es el *Virus del mosaico del tabaco* (TMV) de la familia *Virgaviridae*. Este virus presenta un virión con forma de cilindro rígido con estructura helicoidal de aproximadamente 300 nm de longitud y 18 nm de diámetro. Su estructura fue caracterizada en los años 70 por difracción de fibra de rayos-X (Bloomer et al. 1978, Stubbs, Warren, and Holmes 1977) y también recientemente por criomicroscopía electrónica (crio-ME) (Clare and Orlova 2010). Existen otro tipo de virus filamentosos de plantas que presentan viriones flexibles, conocidos como “virus filamentosos flexibles” constituyendo un grupo formado por varias familias bastante amplias. De los viriones de los miembros del grupo de virus filamentosos flexibles se sabía muy poco a nivel estructural ya que, a diferencia de TMV, la naturaleza flexible de estos viriones imposibilitaba su estudio mediante difracción de fibra (Solovyev and Makarov 2016). En cambio, gracias al desarrollo de la criomicroscopía de alta resolución, la estructura del virión de varias especies del género *Potexvirus*, perteneciente al grupo de los virus

filamentosos flexibles, pudo ser determinada por primera vez en el año 2015 (Agirrezabala et al. 2015, DiMaio et al. 2015). Estos trabajos mostraron que existen diferencias radicales entre las proteínas de la cápside (CPs) de los virus filamentosos rígidos y los virus filamentosos flexibles pese a que a baja resolución parecían compartir una organización parecida. Es la distinta naturaleza de estas CPs la que aporta flexibilidad o rigidez a los viriones en cada caso (DiMaio et al. 2015) y sin duda se pone de manifiesto la diversidad existente entre las CPs que pueden dar lugar a viriones filamentosos de organización helicoidal.

1.3.1 Cápsides helicoidales

Las cápsides helicoidales necesitan un mayor número de proteínas que las cápsides icosaédricas para empaquetar un genoma viral de la misma longitud; por ejemplo, el genoma del virus icosaédrico poliovirus constituye un 30% de la masa total del virión, mientras que en virus helicoidal TMV, el genoma constituye un 6% de la masa total. Esto hace que sean consideradas menos eficientes que las cápsides icosaédricas (Booth, Rabb, and Beniac 2013); en cambio, este tipo de cápsides están muy presentes entre los virus; alrededor de un 50% de las especies conocidas de virus (+)ssRNA empaquetan el genoma con cápsides helicoidales (Stubbs and Kendall 2012). A diferencia de las cápsides icosaédricas, las cápsides helicoidales no imponen una limitación de espacio y pueden empaquetar genomas más largos (Solovyev and Makarov 2016).

El virión de los virus filamentosos está constituido por una cápside formada normalmente por múltiples copias o subunidades de CP. Cada CP interacciona estrechamente con el genoma del virus y lo empaqueta ensamblándose entre ellas siguiendo una organización helicoidal que resultará en la formación de una estructura en forma de tubo de longitud variable. A diferencia de los virus icosaédricos, donde el genoma no acompaña la organización simétrica de la cápside, en las cápsides helicoidales el genoma empaquetado se dispone respetando la organización helicoidal de la cápside. La simetría helicoidal viene definida por un componente rotacional y un componente translacional, que combinados dan lugar a un eje helicoidal (eje longitudinal del virión). La aplicación repetida de una rotación (ángulo azimutal: φ) a un objeto seguida de una translación (x) a lo largo de un eje resulta en la formación de una estructura con simetría helicoidal con un paso de hélice dado (la longitud de translación cuando la suma de las rotaciones alcanza 360°) (Prasad and Schmid 2012) (Figura 1.3). Las estructuras helicoidales están caracterizadas por la localización radial de las subunidades que las conforman respecto al eje helicoidal central, definiéndose así el número de subunidades por cada vuelta de hélice, equivalente a dividir 360° entre el ángulo de rotación

por subunidad (ϕ); además de la elevación de la hélice, que corresponde a la longitud de translación por cada rotación ϕ (Figura 1.3) (Prasad and Schmid 2012).

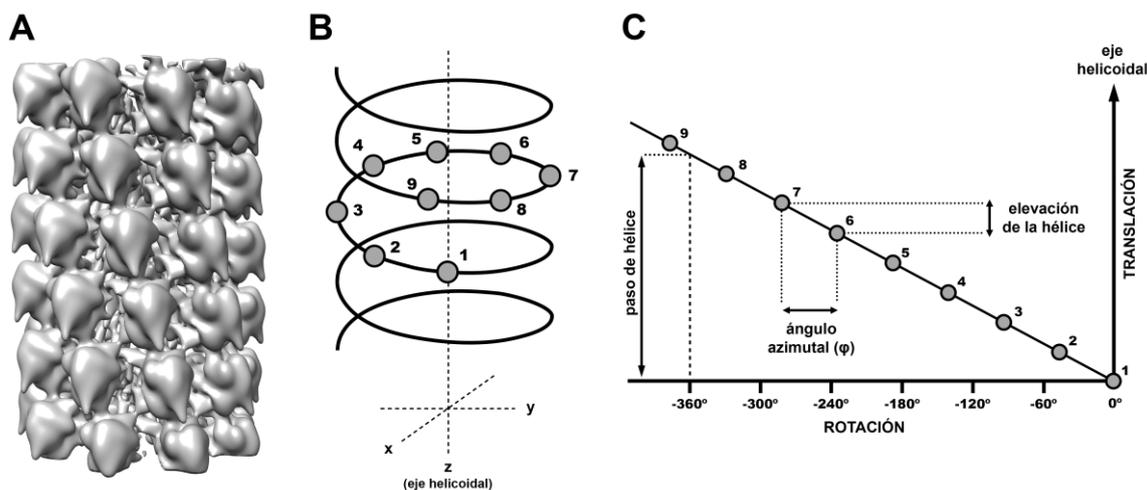


Figura 1.3: Definición de la simetría helicoidal de una hélice levógira. **(A)** Mapa 3D de una estructura con organización helicoidal levógira. **(B)** Esquema en 3D de la trayectoria de la hélice y posición de las subunidades de una vuelta de hélice de la estructura de (A). Cada vuelta de hélice está compuesta por algo menos de 9 subunidades y el eje z representa el eje helicoidal de la estructura. **(C)** Representación 2D del esquema de (B). Una vuelta de hélice está representada en dos ejes: el vertical muestra la translación de cada subunidad en la hélice (elevación de la hélice) y el horizontal muestra la rotación de cada subunidad (ángulo azimutal) respecto al eje helicoidal.

1.4. Virus filamentosos flexibles

Los virus filamentosos flexibles son un grupo de patógenos de plantas compuesto por miembros que presentan viriones filamentosos y flexibles con genoma de RNA de hebra sencilla positiva ((+)ssRNA). Este grupo engloba cuatro familias: *Potyviridae*, *Alphaflexiviridae*, *Betaflexiviridae* y *Closteroviridae*. En términos de número de especies descritas, los virus filamentosos flexibles son más abundantes que los rígidos (King et al. 2012), y sus viriones están formados por cientos de copias de la CP ensambladas entre sí para formar la cápside helicoidal empaquetando el genoma del virus. Estas partículas infectivas son delgadas (10-15 nm de diámetro) y largas (varios cientos de nm dependiendo de la familia) (Adams 2008, López-Moya and García 2008), y en imágenes de microscopía electrónica a baja resolución tienen un aspecto muy parecido en las cuatro familias (Figura 1.4). Estos virus están distribuidos por todo el mundo y en conjunto infectan una gran variedad

de plantas, herbáceas y leñosas, monocotiledóneas y dicotiledóneas. Dependiendo de la familia y la especie se transmiten por contacto directo entre plantas o a través de insectos (Adams 2008, Lecoq and Desbiez 2008). Una vez infectada la planta, el virus se distribuye a través del floema y/o de célula a célula (López-Moya and García 2008).

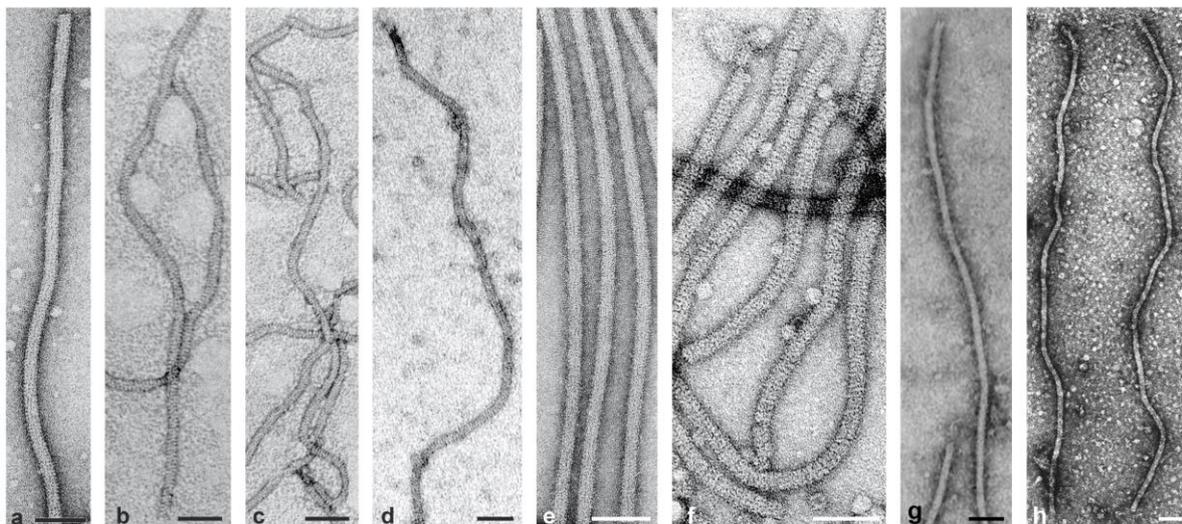


Figura 1.4: Viriones de algunas especies representativas de las distintas familias del grupo de virus filamentosos flexibles visualizadas mediante microscopía electrónica empleando tinción negativa. Los viriones son: “*Potato virus X*” (a) e “*Indian citrus ringspot virus*” (f), familia *Alphaflexiviridae*; “*Grapevine virus B*” (b), “*Apple chlorotic leaf spot virus*” (c), “*Apple stem pitting virus*” (d) y “*Carnation latent virus*” (e), familia *Betaflexiviridae*; “*Plum pox virus*” (g), familia *Potyviridae*; y “*Beet yellow virus*” (h), familia *Closteroviridae*. Barras negras/blancas = 50 nm. Figura de (Martelli et al. 2007) con modificaciones a partir de imágenes de I.M. Roberts y D.E. Lesemann para ICTV (Lefkowitz et al. 2018).

El genoma de los miembros de estas familias suele ser estar formado por una única molécula de ácido nucleico, en cambio algunos géneros dentro de estas familias presentan el genoma dividido en dos moléculas de ácido nucleico empaquetadas en filamentos distintos (Hulo et al. 2011). Dependiendo de la familia y el género, el genoma de estos virus codifica un número variable de proteínas; por ejemplo, los genomas de la familia *Alphaflexiviridae* presenta genes para 2-6 proteínas, en cambio los genomas de la familia *Closteroviridae* en algunos casos codifican hasta más de 10 proteínas (Karasev et al. 1995, Adams 2008). Todos los genomas de estos virus contienen información para la RNA polimerasa dependiente de RNA (RdRp) específica del virus, la proteína de la cápside, y otras proteínas implicadas en la replicación del genoma del virus y en su ciclo de vida, con diferentes funciones a lo largo de la infección (Adams 2008, López-Moya and García 2008, Hulo et al. 2011).

Existen otros miembros patógenos de plantas con viriones filamentosos y flexibles. Son las familias *Ophioviridae* (7 especies), cuyos viriones presentan una flexibilidad muy elevada, y los miembros del género *Tenuivirus* (7 especies) de la familia *Phenuiviridae*, de genoma circular, ambas con genoma (-)ssRNA y segmentado. Pese a ser patógenos de plantas con viriones filamentosos y flexibles, a estos miembros no se les ha relacionado por ahora con el grupo de virus filamentosos flexibles (+)ssRNA.

Actualmente, el grupo de virus filamentosos flexibles está constituido por más de 400 especies de virus distribuidos en las cuatro familias que lo componen. La familia *Potyviridae* es la más grande de las cuatro, compuesta por 205 especies agrupadas en 11 géneros, siendo el género *Potyvirus* el taxón más amplio (168 especies); la familia *Alphaflexiviridae* la componen 56 especies distribuidas en 7 géneros, siendo el mayor de ellos el género *Potexvirus* (38 especies); la familia *Betaflexiviridae* está dividida en dos subfamilias que integran 97 especies en total repartidas en 12 géneros, siendo el género *Carlavirus* el más amplio (53 especies); y por último, la familia *Closteroviridae*, compuesta por 50 especies repartidas en 5 géneros (Figura 1.5).

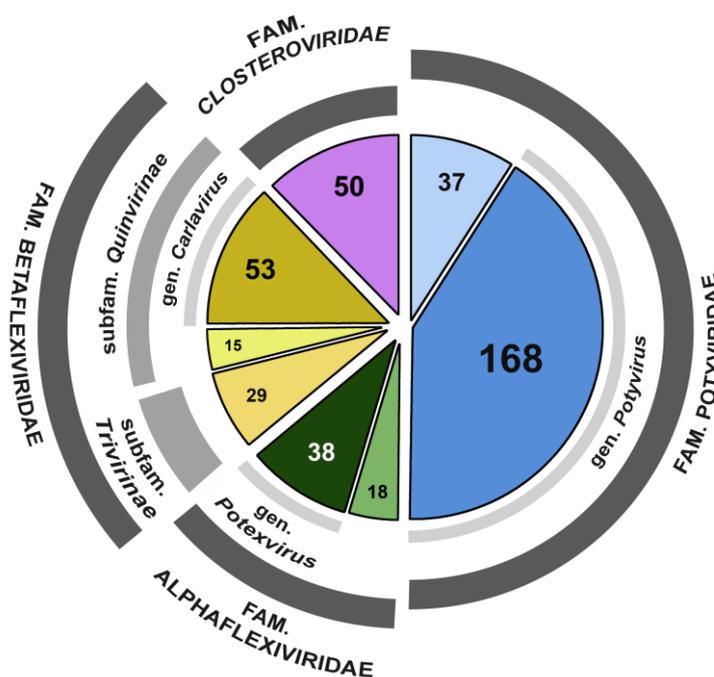


Figura 1.5: Distribución de especies dentro del grupo de virus filamentosos flexibles. Sólo el género *Potyvirus* agrupa casi la mitad de las especies de este grupo.

Según la clasificación de ICTV, las familias *Alphaflexiviridae* y *Betaflexiviridae* actualmente forman parte del orden *Tymovirales*, donde se encuentran otras familias de virus con viriones no filamentosos, y las familias *Closteroviridae* y *Potyviridae* no se encuentran asignadas en ningún orden. Por otro lado, las familias *Alphaflexiviridae*, *Betaflexiviridae* y *Closteroviridae* están incluidas en la súper-familia “alphavirus-like”, ya que sus RdRp, sus enzimas de procesamiento del genoma viral y sus helicasas son muy similares. En cambio, la familia *Potyviridae* está incluida dentro de la súper-familia “picornavirus-like” según la filogenética basada en la similitud de sus RdRp, el tipo de traducción de su material genético, y la presencia de la proteína VPg unida al genoma del virus (Koonin and Dolja 1993). Pese a esta agrupación, los miembros de la familia *Potyviridae* no encajan del todo en esta súper-familia donde predominan los virus con cápsides icosaédricas formadas por proteínas con dominios “jelly-roll” característicos del linaje “Picorna-like” (Abrescia et al. 2012), y sus cápsides filamentosas parecen tener más elementos en común con las de las otras familias de virus filamentosos flexibles (Koonin, Dolja, and Krupovic 2015, Kendall et al. 2008).

1.4.1 Estructura de viriones de los virus filamentosos flexibles

Los diferentes intentos de determinar la estructura a alta resolución de los viriones de los virus filamentosos flexibles con métodos empleados con los virus rígidos como TMV han sido un reto infructuoso debido a la alta flexibilidad que presentan estos viriones (Solovyevev and Makarov 2016). Inicialmente, mediante crio-microscopía electrónica a baja resolución se pudo confirmar la simetría helicoidal de la cápside del *Virus X de la patata* (PVX), de la familia *Alphaflexiviridae* y el *Virus del mosaico de la soja* (SMV), de la familia *Potyviridae*, lo que reveló una arquitectura del virión similar en dos familias de virus filamentosos flexibles (Kendall et al. 2008). Más adelante, se determinó por primera vez la estructura de una parte de la proteína de la cápside del *Virus del mosaico de la Papaya*, del género *Potexvirus* (familia *Alphaflexiviridae*) mediante difracción de rayos-X, lo que puso de manifiesto diferencias importantes con la CP de TMV (Yang et al. 2012). Finalmente, los avances de la crio-microscopía electrónica permitieron determinar por primera vez la estructura a alta resolución de los viriones del *Virus del mosaico del pepino* (PepMV) y del *Virus del mosaico del bambú* (BaMV), dos miembros del género *Potexvirus* por crio-ME (Agirrezabala et al. 2015, DiMaio et al. 2015). Estos resultados permitieron conocer la estructura a alta resolución del virión de miembros del género *Potexvirus*, así como el modelo atómico de la CP en el contexto del virión maduro (Agirrezabala et al. 2015) (Figura 1.6, B).

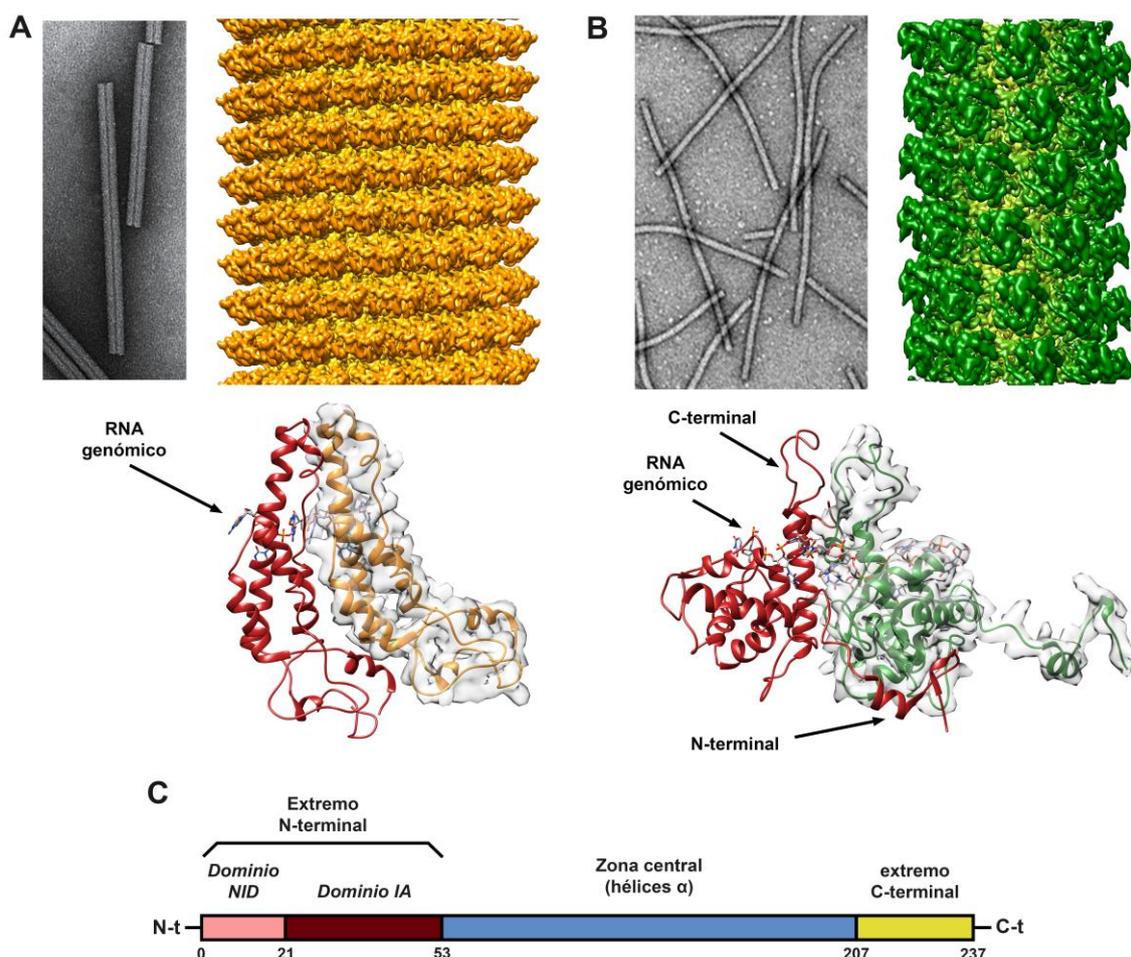


Figura 1.6: Estructura del virión del género *Potexvirus* en comparación con el virión de TMV. **(A)** Imagen de filamentos de TMV por microscopía electrónica (izquierda) (Klug 1999) y mapa 3D obtenido por criomicroscopía electrónica de una región del filamento de TMV (derecha). Abajo, modelos atómicos de dos CP de TMV adyacentes formando el virión (rojo y naranja) (códigos modelo atómico y mapa de densidad 3D: 3J06 y EMD-5185 respectivamente); la densidad 3D de una de ellas se muestra en gris semi-transparente (Ge and Zhou 2011). **(B)** Imagen de filamentos de PepMV por microscopía electrónica (izquierda) y mapa 3D obtenido por criomicroscopía electrónica de una región del filamento de PepMV (derecha). Abajo, modelos atómicos de dos CP de PepMV adyacentes en el contexto del virión ensamblado (rojo y verde) (códigos modelo atómico y mapa de densidad 3D: 5FN1 y EMD-3236 respectivamente); la densidad 3D de una de ellas se muestra en gris semi-transparente (Agirrezabala et al. 2015). **(C)** Esquema de la distribución de regiones y dominios de la CP de PepMV, muy similar al resto de miembros de *Potexvirus* descritos.

Aunque tanto los viriones de los virus filamentosos rígidos como de los virus filamentosos flexibles presentan una organización helicoidal formando un filamento, las proteínas de la cápside de los miembros descritos del género *Potexvirus* (PepMV, BaMV y PapMV), muestran una estructura muy distinta de las CP de TMV. Además, la determinación estructural de los viriones de PepMV y BaMV reveló que la organización helicoidal de sus cápsides presenta una disposición levógira, a diferencia de la disposición dextrógira que

presentan los viriones de TMV (Agirrezabala et al. 2015, DiMaio et al. 2015, Yang et al. 2012, Namba and Stubbs 1986) (Figura 1.6, A y B). Las CPs de los miembros de *Potexvirus* presentan una estructura central globular rica en hélices α y dos largos brazos terminales en ambos extremos de la proteína. El ensamblaje entre CPs está mediado principalmente por estos dos extremos N- y C-terminales flexibles, permitiendo ciertos movimientos entre CPs, lo que explicaría la flexibilidad característica de estos viriones. El extremo N-terminal es un elemento muy relevante en el ensamblaje del virión a través de su interacción con la subunidad adyacente (Figura 1.6, B); presenta un dominio denominado IA (“interaction arm”), que es la zona que establece contacto con la CP vecina, y el dominio NID, (“N-terminal intrinsically disordered”) no visible en los modelos atómicos, que se asume que se extiende hacia el exterior del virión y no se le atribuye ningún papel en el ensamblaje (Figura 1.6, C) (DiMaio et al. 2015, Agirrezabala et al. 2015, Solovyev and Makarov 2016).

Hasta la fecha sólo se conoce la estructura a alta resolución de los viriones de tres miembros del género *Potexvirus* dentro de la familia *Alphaflexiviridae*, y se desconoce la estructura de los viriones y las CPs que los forman del resto de familias de virus filamentosos flexibles. Debido a la similitud en la organización de los viriones, se piensa que estas familias tienen una relación evolutiva a nivel de sus CPs, sin embargo, no hay datos suficientes que apoyen esta idea ya que la homología de secuencia entre CPs de distintas familias de virus filamentosos flexibles es nula, y no existen datos estructurales a alta resolución de las CPs del resto de familias de virus filamentosos flexibles (Dolja et al. 1991, Kendall et al. 2008).

1.4.2 Virus del mosaico de la sandía y la familia *Potyviridae*

El Virus del mosaico de la sandía (WMV) pertenece al género *Potyvirus* dentro de la familia *Potyviridae*. Este género es el taxón más grande del grupo de virus filamentosos flexibles abarcando el 15% de los virus de plantas conocidos (Wylie et al. 2017). En general, los miembros de la familia *Potyviridae* están presentes por todo el mundo e infectan una gran variedad de familias de plantas incluyendo cultivos de especial relevancia económica como cereales, legumbres, vegetales, árboles frutales y plantas ornamentales. WMV está distribuido por todo el mundo, con gran presencia en la región del Mediterráneo. Es considerado principalmente un patógeno de cucurbitáceas, pero es uno de los miembros del género *Potyvirus* con mayor rango de hospedadores infectando hasta 170 especies de 26 familias de plantas incluyendo legumbres, orquídeas y plantas forrajeras (Lecoq and Desbiez

2008, López-Moya and García 2008). Los miembros de la familia *Potyviridae* se transmiten de forma horizontal de planta a planta a través de vectores; dependiendo del género pueden ser insectos como áfidos, ácaros o moscas, y en algunos casos hongos. Los miembros de algunos géneros también se pueden transmitir de forma vertical a través de las semillas a la descendencia (Figura 1.7) (López-Moya and García 2008, Revers and García 2015b).

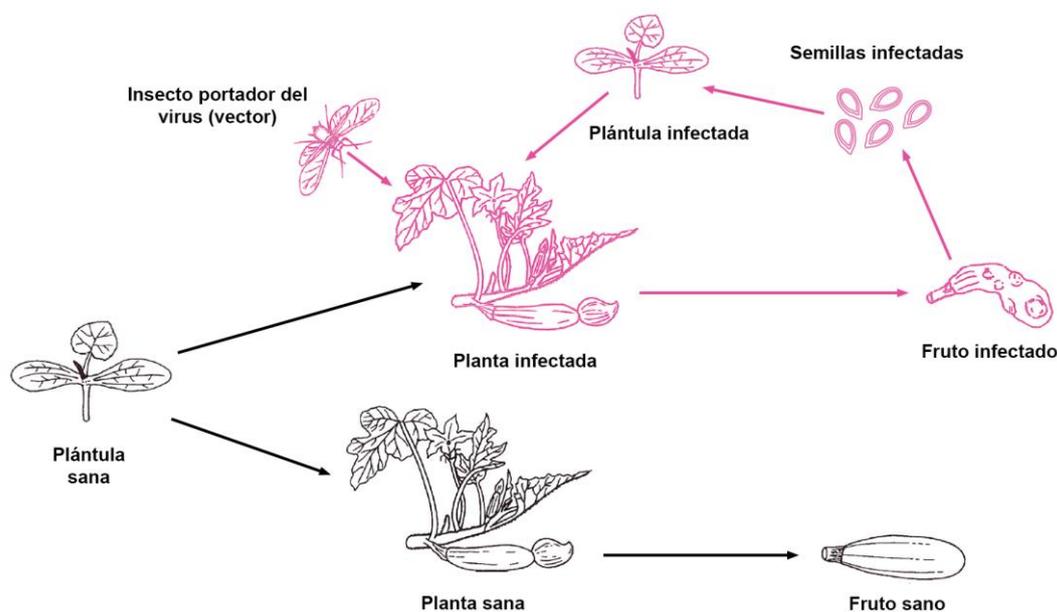


Figura 1.7: Modos de transmisión horizontal y vertical de los miembros de la familia *Potyviridae*.

El virión de los miembros del género *Potyvirus*, y de la familia *Potyviridae* en general, es muy similar al del resto de virus filamentosos flexibles. Está constituido por cientos de proteínas de la cápside con un peso molecular entre 30 y 40 kDa formando un filamento de 11-15 nm de ancho y una longitud variable entre géneros y especies comprendida entre 700 y 900 nm (López-Moya and García 2008). El genoma ssRNA de estos virus suele ser monopartito, exceptuando el género *Bymovirus* que presenta genoma segmentado, y su tamaño determina la longitud del virión, oscilando entre 8-12 kb (Brister et al. 2015). WMV presenta un genoma monopartito de unas 10 kb (Brister et al. 2015), y al igual que el resto de miembros de la familia *Potyviridae* está flanqueado por dos regiones no-codificantes (NCR): una en el extremo 5' unida covalentemente a una proteína terminal (VPg) que promueve la traducción viral entre otras funciones, y otra en el extremo 3' seguida de una cola de poli(A) de longitud variable (Revers and García 2015b, Eskelin et al. 2011). El genoma (+)ssRNA de los miembros de esta familia es usado directamente por la maquinaria de traducción de la célula hospedadora. Este contiene una única fase de lectura abierta (ORF) traducido como una única poliproteína que por la acción de proteasas virales da lugar a casi todas las proteínas

del virus (Adams, Antoniow, and Beaudoin 2005). Adicionalmente, existe una ORF alternativa producida por un mecanismo de “frameshifting”, presente en todos los miembros de la familia, que da lugar a una de las proteínas del virus (Chung et al. 2008) (Figura 1.8).

En total, el genoma de WMV, al igual que los miembros del género *Potyvirus*, codifica para 10 proteínas; 8 de ellas en común con el resto de miembros de la familia *Potyviridae*, ubicadas en la parte central y la región carboxilo terminal de la poliproteína, que presenta una organización y composición conservada en todos los miembros de la familia, siendo la región amino terminal la que más variación en composición presenta entre géneros (Revers and García 2015b) (Figura 1.8).

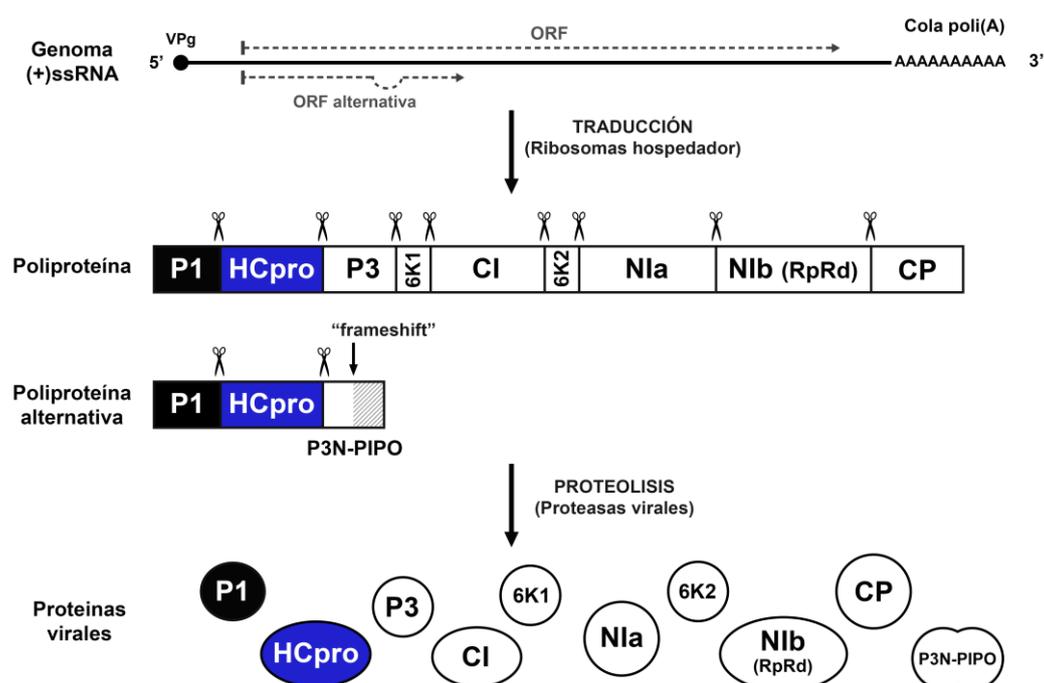


Figura 1.8: Esquema de la organización del genoma y modo de traducción de los miembros del género *Potyvirus*, muy similar al resto de géneros de la familia *Potyviridae*. El genoma (+)ssRNA está representado con una línea sólida negra con la proteína VPg representada con un círculo coloreado en negro; las dos ORFs están representadas con líneas discontinuas a ambos lados del genoma. Las poliproteínas resultantes de las dos ORFs se muestran como cajas que representan cada una de las proteínas que las forman; las regiones de la poliproteína conservadas en toda la familia *Potyviridae* se representan en blanco. La proteólisis de la poliproteína que da lugar a cada una de las proteínas del virus viene dada por la acción de las proteasas virales P1, HCpro y NIa.

Los potyvirus (miembros del género *Potyvirus*) procedentes de un vector o bien de otros tejidos o células adyacentes infectan una nueva célula de la planta. Una vez en el interior de esta, el genoma es liberado de la cápside y traducido por la maquinaria de la célula

hospedadora para dar lugar a las proteínas virales. Casi todas las proteínas del virus sintetizadas junto con diversos factores de la célula hospedadora llevarán a cabo la replicación del genoma del virus (Urququi-Inchima, Haenni, and Bernardi 2001, Kekarainen, Savilahti, and Valkonen 2002) que se cree tiene lugar en el interior de vesículas en el citoplasma aparentemente asociadas al sistema de endomembranas y los cloroplastos (Urququi-Inchima, Haenni, and Bernardi 2001, Kekarainen, Savilahti, and Valkonen 2002, Revers and García 2015b, Wei et al. 2010). Durante este proceso, el genoma del virus (+)ssRNA es usado para crear la hebra complementaria (-)ssRNA que será empleada como molde para sintetizar nuevos genomas virales, que nuevamente serán usados para la síntesis de proteínas virales o para formar nuevos viriones que podrán ser transmitidos a otra planta a través de otro vector (López-Moya and García 2008) (Figura 1.9). El genoma viral se puede transmitir a las células adyacentes a través de los plasmodesmos (movimientos célula-célula) bien formando RNPs (ribonucleoproteínas) compuestas por el RNA genómico (RNAg) unido a varias proteínas virales o bien en forma de viriones maduros (Heinlein 2015, Lucas 2006). También, a través de las células del floema, los viriones pueden moverse largas distancias para llegar a infectar otros tejidos de la planta, sin embargo, este es un sistema más complejo que requiere la interacción con otros factores de la planta (Hipper et al. 2013) (Figura 1.9).

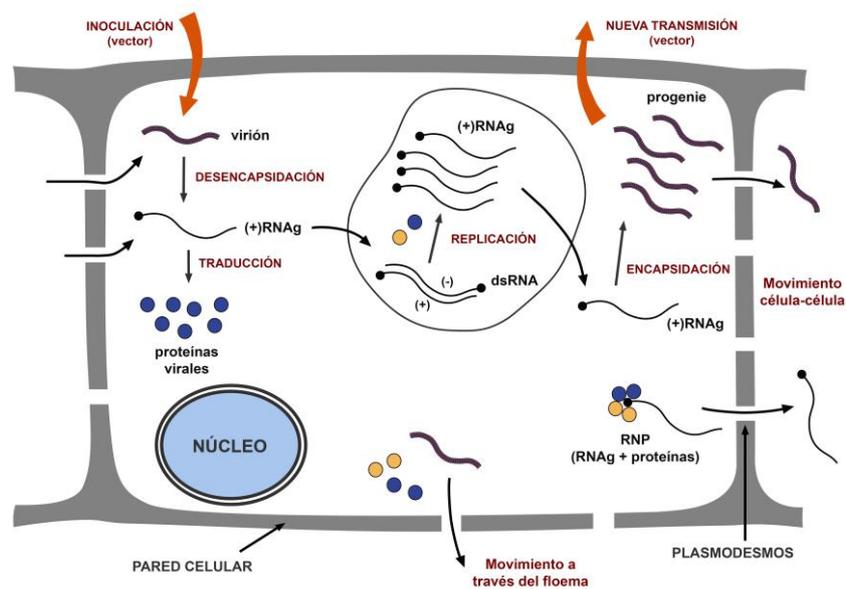


Figura 1.9: Representación esquemática de los diferentes eventos durante la infección de una célula vegetal por un potyvirus. La pared celular de la célula vegetal está coloreada en gris y el citoplasma en blanco. El ciclo del virus comienza con la entrada de partículas virales a través de un vector o desde una célula adyacente; el genoma del virus es traducido por la maquinaria de la célula sintetizando las proteínas del virus (coloreadas en azul). La replicación del genoma viral tiene lugar en el interior de vesículas y/o cloroplastos generando nuevos genomas virales que, tras ser encapsidados, darán lugar a la progenie del virus. Desde una célula infectada, los nuevos viriones o los nuevos genomas del virus formando RNPs se desplazan a células vecinas a través de los plasmodesmos, o al floema para alcanzar tejidos más lejanos, ayudados por proteínas virales y de la célula hospedadora (coloreadas en amarillo).

Las proteínas virales de los Potyvirus tienen múltiples funciones y algunas de ellas están implicadas en varios procesos durante el ciclo de vida del virus (Revers and García 2015b). Entre ellas, la proteína de la cápside que, además de su función principal como unidad estructural del virión empaquetando el RNA genómico, también está implicada en otras funciones. Esta proteína también es importante en la transmisión célula-a-célula más allá de empaquetar el genoma y junto con otras proteínas del virus también es relevante en la interacción con el vector durante su transmisión (Solovyev and Makarov 2016). Además, aunque no se sabe exactamente como ocurre la encapsidación del RNA genómico, su regulación, ni donde tiene lugar dentro de la célula, parece que los niveles de la CP en la célula hospedadora juegan un papel importante en evitar una encapsidación prematura (Mäkinen and Hafrén 2014).

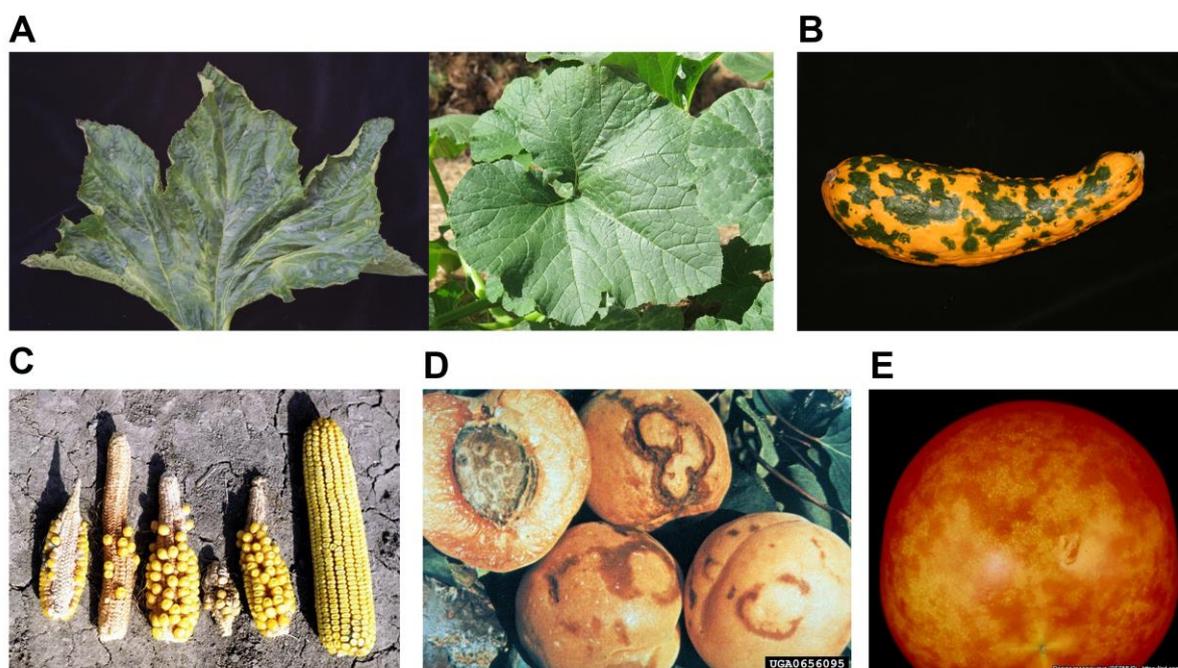


Figura 1.10: Ejemplos de síntomas provocados por los virus filamentosos flexibles, especialmente el género *Potyvirus*. (A) Hoja de planta de calabacín sana (derecha) e infectada por WMV presentando mosaicismo (izquierda). (B) Calabacín afectado por la infección de WMV presentando pérdida de color. (C) Síntomas del *Virus del mosaico enano del maíz* (gen. *Potyvirus*). (D) Síntomas del Virus de la Sharka (gen. *Potyvirus*) en albaricoques. (E) Síntomas del *Virus del mosaico del pepino* (fam. *Alphaflexiviridae*) en tomate.

Las infecciones de Potyvirus causan numerosas enfermedades de gravedad variable dependiendo del virus, el tipo de planta y la edad de la planta en el momento de la infección. La mayoría de estas infecciones provocan la aparición de síntomas primarios como mosaicismo, moteado, anillos cloróticos y decoloración en hojas, flores, frutos y otros tejidos. En muchos casos, la infección por estos virus provoca síntomas más graves como la atrofia

de las plantas jóvenes, deformaciones importantes de frutos y necrosis de varios tejidos (Figura 1.10) (Agrios 2005). WMV provoca síntomas diversos dependiendo de la cepa del virus y de la planta que infecte. Entre ellos, mosaicismo y deformaciones en las hojas, y en el caso de las cepas más agresivas disminución de crecimiento y aparición de deformaciones severas en frutos de plantas cucurbitáceas (Lecoq and Desbiez 2008).

1.4.3 Relevancia económica y social de los virus filamentosos flexibles

Los virus se encuentran entre los mayores patógenos que afectan a la agricultura a nivel mundial, y la mitad de las nuevas enfermedades emergentes identificadas en plantas son causadas por estos (Anderson et al. 2004, Oerke 2006). El género *Potyvirus* es uno de los grupos de virus más importantes entre los agentes causantes de nuevas enfermedades de plantas por todo el mundo. Es considerado uno de los grupos de patógenos de plantas más exitosos a nivel mundial, ya que a medida que se introducen nuevos cultivos en determinadas regiones aparecen nuevas especies de este género como patógenos (Rybicki and Pietersen 1999).

El origen de la gran especiación y éxito del género *Potyvirus* parece coincidir con el origen de la agricultura (Gibbs et al. 2008); además varios trabajos señalan la relación existente entre la reducción de la diversidad de plantas como consecuencia de la agricultura con un incremento y prevalencia de patógenos de las mismas, incluyendo algunos virus (Lacroix et al. 2014, Pagán et al. 2012). No existen registros anteriores al s. XX sobre de la incidencia y el impacto de los virus filamentosos flexibles en la agricultura, sin embargo, sí hay evidencias de síntomas causados por potyvirus durante el s. XVII en Holanda, como reflejan diversas pinturas de esa época (Gibbs et al. 2008, Wylie et al. 2017). Actualmente los virus filamentosos flexibles están presentes en todas las regiones agrícolas a lo largo del mundo, e infectan gran variedad de cultivos muy relevantes desde el punto de vista económico y social (Tabla 1.1). Las pérdidas ocasionadas por estos virus varían mucho dependiendo de la especie del virus y de la especie y variedad de la planta que infectan, además de la zona geográfica y la climatología, pero a menudo supone una disminución de la cosecha bastante alta, de hasta el 70% en algunos cultivos, suponiendo un impacto económico de bastante relevancia (Sastry 2013). Según los datos de La Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), los cultivos destinados a alimento más relevantes en los países en vías de desarrollo son: arroz, maíz, yuca, plátano y batata; seguidos en menor medida de vegetales frescos y frutas. En estas regiones, este tipo de cultivo está destinado

a la alimentación en las zonas locales donde se produce en lugar de exportación, de modo que las enfermedades causadas por virus en estos cultivos pueden suponer un problema mayor, ajeno a la repercusión económica (Rybicki and Pietersen 1999). Hay varias especies del grupo de virus filamentosos flexibles que infectan a los cultivos mencionados, y algunos de ellos son considerados de especial relevancia por el impacto que ocasionan (Rybicki 2015, Scholthof et al. 2011).

Cultivo	Virus
Trigo	<i>Virus del mosaico rayado del trigo (Potyviridae)</i>
Maíz*	<i>Virus del mosaico enano del maíz (Potyviridae)</i>
Soja	<i>Virus del mosaico de la soja (Potyviridae)</i>
Judía	<i>Virus del mosaico común de la judía (Potyviridae)</i>
Tomate	<i>Virus del mosaico del pepino (Alphaflexiviridae)</i> <i>Virus etch del tabaco (Potyviridae)</i>
Cucurbitáceas	<i>Virus del mosaico de la sandía (Potyviridae)</i> <i>Virus del mosaico amarillo del calabacín (Potyviridae)</i>
Patata	<i>Virus Y de la patata (Potyviridae)</i> <i>Virus X de la patata (Alphaflexiviridae)</i> <i>Virus S de la patata (Betaflexiviridae)</i>
Batata*	<i>Virus del moteado suave de la batata (Potyviridae)</i>
Yuca*	<i>Virus del manchado marrón de la yuca (Potyviridae)</i>
Cítricos	<i>Virus de la tristeza de los cítricos (Closteroviridae)</i>
Frutales de hueso	<i>Virus de la Sharka (Potyviridae)</i>
Plátanos*	<i>Virus del mosaico de las brácteas del plátano (Potyviridae)</i>
Caña de azúcar	<i>Virus del mosaico de la caña de azúcar (Potyviridae)</i>
Remolacha	<i>Virus del amarilleo de la remolacha (Closteroviridae)</i> <i>Virus del mosaico de la remolacha (Potyviridae)</i>

Tabla 1.1: Ejemplos de patógenos pertenecientes al grupo de virus filamentosos flexibles que infectan los cultivos más relevantes a nivel económico y social. Con un asterisco están marcados aquellos cultivos de especial interés por su importancia en países en vías de desarrollo (Rybicki 2015).

Se han identificado más de 40 especies de virus como patógenos de cultivos en España pertenecientes a las cuatro familias de los virus filamentosos flexibles (Melgarejo et al. 2010). Algunos de estos virus tienen una repercusión bastante relevante en la agricultura de la península; por ejemplo, el *Virus de la tristeza de los cítricos* (CTV) perteneciente a la familia *Closteroviridae*, que se encuentra en todas las zonas de cultivo de cítricos del mundo, se estima que ha matado más de 40 millones de árboles en España desde su brote en los años 50 (Cambra et al. 2000). El *Virus de la Sharka* (PPV) (familia *Potyviridae*) infecta todo tipo de frutales de hueso, es endémico en Europa y está presente en casi todas las regiones de España donde se cultivan este tipo de árboles. Se estima que el coste asociado a la presencia

de esta patología y su control en todo el mundo asciende a los 10.000 millones de dólares durante los últimos 30 años (Cambra et al. 2006). El *Virus Y de la patata* (PVY), también perteneciente a la familia *Potyviridae*, infecta plantas de patata, tomate, pimiento y tabaco. Se encuentra distribuido por las zonas donde se dan estos cultivos en España y es considerado el principal factor limitante en la producción de patata pudiendo ocasionar pérdidas de hasta el 80% de la cosecha (Kerlan and B 2008, Melgarejo et al. 2010). El *Virus del mosaico del pepino* (PepMV) (familia *Alphaflexiviridae*) infecta plantas de tomate en la cuenca mediterránea de España; también presente en América del Norte y en diferentes zonas de Europa. El impacto de este virus no parece afectar tanto al tamaño de la cosecha, sino a la calidad de los frutos (Gómez, Sempere, and Aranda 2012). Por último, WMV es un patógeno presente en numerosas regiones de España (citado en Valencia, Aragón, Madrid, Cataluña, Murcia, Andalucía y Extremadura) infectando a varios cultivos: pepino, sandía, melón, calabaza, calabacín (Melgarejo et al. 2010). WMV provoca deformaciones severas en los frutos ocasionando pérdidas de entre el 10% y el 60% de la cosecha en algunos casos (Sastry 2013).

Tanto en las zonas agrícolas de España como en el resto del mundo, la presencia de estos patógenos supone un reto. El control de estas enfermedades radica principalmente en el uso de pesticidas contra los vectores, aunque sin mucho éxito, y en el uso de variedades de cultivos resistentes a la enfermedad. Sin embargo, no existen variedades resistentes en todos los cultivos frente a todos los tipos de virus que los infectan, y las variedades resistentes pueden volverse sensibles a brotes de nuevas cepas de un mismo virus. En este contexto, es necesario desarrollar nuevas estrategias para mitigar el impacto de las enfermedades causadas por virus en plantas basadas en ampliar el conocimiento de los factores que favorecen el desarrollo de la enfermedad, así como la naturaleza de este tipo de patógenos (Aranda and Freitas-Astúa 2017, Gómez et al. 2009).

1.5 Microscopía electrónica de transmisión (MET) en biología estructural.

Para entender completamente los procesos biológicos y la organización de los componentes que constituyen cualquier organismo celular o acelular es necesario conocer la estructura tridimensional de las macromoléculas involucradas. Las técnicas empleadas para alcanzar este fin son diversas y están basadas en diferentes fenómenos físicos, pero todas ellas

suelen ser complementarias a la hora de abordar fenómenos biológicos a nivel molecular; entre ellas, la cristalografía de rayos X, la resonancia magnética nuclear y la microscopía electrónica de transmisión (MET) suelen ser habituales en el estudio de macromoléculas biológicas a nivel estructural. La MET permite estudiar muestras biológicas en un rango de tamaños muy amplio, desde orgánulos subcelulares ($>1 \mu\text{m}$) a proteínas de 100 kDa ($>10 \text{nm}$), pasando por complejos macromoleculares de gran tamaño como son los ribosomas o las partículas virales (Merk et al. 2016).

1.5.1 El microscopio electrónico de transmisión

El microscopio electrónico de transmisión produce imágenes que son proyecciones de un objeto tridimensional (muestra). Los electrones se generan a partir de una fuente de emisión, normalmente un filamento de tungsteno, wolframio o un cristal de hexaboruro de lantano (LaB_6), y son acelerados por un campo magnético. Un sistema de lentes condensadoras convierte el haz de electrones divergentes en un haz paralelo que incide sobre la muestra, ubicada en el portamuestras en el interior del microscopio. Algunos electrones atraviesan la muestra sin interactuar con ella y conservando sus ondas intactas, otros son desviados por la acción del núcleo o de los electrones de los átomos de la muestra, y otros colisionan con las partículas atómicas alterando la muestra. En definitiva, de los electrones que interactúan con la muestra, algunos son dispersados sin pérdida de energía (dispersión elástica) y otros transmiten energía a la muestra (dispersión inelástica). Aquellos electrones que sufren una dispersión elástica contribuyen a la formación de la imagen; su onda es alterada produciendo un contraste de fases con aquellos electrones que no interactúan con la muestra, transmitiendo así la información de la muestra a la imagen formada. Tras atravesar la muestra, los electrones pasan a través de un sistema de lentes involucradas en la formación de la imagen y finalmente son detectados por una cámara CCD (dispositivo de carga acoplada) o una cámara de detección directa. Debido a que los electrones interactúan fuertemente con la materia, el recorrido de los mismos en el microscopio electrónico debe mantenerse en condiciones de alto vacío, para evitar interacciones de los electrones con moléculas de gas, lo que condiciona la preparación de la muestra.

1.5.2 Estudio de macromoléculas biológicas por MET

Las condiciones de alto vacío del microscopio electrónico exigen que la muestra de estudio se encuentre en estado sólido, por lo que para el estudio de muestras biológicas por MET se emplean principalmente dos técnicas. Una de ellas, conocida como tinción negativa, implica la utilización de sales de un metal pesado; la muestra, formando una fina capa en una rejilla es embebida en una solución del mismo, típicamente acetato de uranilo o formiato de uranilo. El metal pesado se deposita como una envuelta densa que recubre y delinea las macromoléculas presentes en la muestra. Las imágenes que se obtienen con este método contienen principalmente información sobre la forma que adopta el recubrimiento del metal pesado sobre la macromolécula, proporcionando sólo información de la superficie de las macromoléculas en la muestra; sobre su forma, tamaño y posible simetría, en un rango de resolución alrededor de 20 Å. La otra técnica, conocida como crio-microscopía electrónica (crio-ME), implica estudiar los complejos biológicos inmersos en hielo vítreo, lo que permite conservar la integridad estructural de la muestra evitando su deshidratación en las condiciones de alto vacío del interior del microscopio y mantener unas características similares a su estado en solución; para ello, es necesario congelar la muestra una vez adherida a la rejilla a temperaturas de -180°C sumergiendo la rejilla rápidamente en etano líquido. A diferencia de la tinción negativa, la información obtenida con esta técnica procede directamente de la interacción de los electrones con la muestra biológica y nos aporta detalles que permiten determinar la estructura de la muestra de estudio alcanzando resolución atómica en algunos casos.

1.5.3 Reconstrucción tridimensional

En las imágenes obtenidas de un microscopio electrónico de transmisión (micrografías) obtenemos las proyecciones 2D (partículas) de las macromoléculas en solución que se encontraban en la muestra, idealmente, repartidas homogéneamente y presentando orientaciones aleatorias en todo el espacio. Mediante el método de análisis de partículas individuales (SPA), se pueden utilizar las partículas obtenidas de la micrografía para reconstruir un mapa de densidad 3D de la macromolécula presente en la muestra. Para ello, es necesario un gran número de partículas de una misma muestra y determinar qué orientación relativa presentan.

Cada partícula de la micrografía es una proyección de una molécula individual en la muestra con una determinada orientación definida por tres ángulos de Euler (ϕ , θ y ψ) y dos parámetros de translación (Δx y Δy) respecto a un sistema de referencia fijado. Se pueden diferenciar dos situaciones a la hora de determinar la orientación de las partículas, cuando la estructura de la muestra es completamente desconocida o bien cuando la estructura se conoce al menos parcialmente. Si la estructura de la muestra se desconoce existen varios métodos para determinar la orientación de las partículas sin necesidad de referencia: el método de reconstrucción cónico-aleatorio (“Random conical tilt”) y el método de líneas comunes. En el primero, las imágenes experimentales se recogen en pares inclinando el portamuestras un ángulo determinado en el interior del microscopio. Como resultado se obtienen dos proyecciones de un mismo objeto con orientaciones distintas y un ángulo que las relaciona (Radermacher et al. 1987, Radermacher 1988). El método de líneas comunes se basa en el “teorema de la sección central” que determina que un par de proyecciones 2D de un mismo objeto tiene al menos una proyección 1D (línea) en común en el espacio de Fourier. Sólo con dos proyecciones el ángulo entre ellas no puede ser determinado, ya que sólo existe una línea común; es necesaria una tercera proyección del mismo objeto, lo que incluye dos nuevas líneas comunes, que permitirá definir el ángulo entre dos líneas comunes con respecto a la tercera (Van Heel 1987, Orlova and Saibil 2011).

Cuando se dispone de un modelo 3D, bien obtenido por los métodos sin referencia, bien proveniente de estructuras disponibles similares a la muestra de estudio, o bien generado a partir de una forma prediseñada similar a la muestra de estudio (esferas en virus icosaédricos o cilindros en virus filamentosos), este se emplea como referencia para determinar los ángulos de Euler de las partículas de manera iterativa en lo que se conoce como “projection-matching” (Penczek, Grassucci, and Frank 1994, Sorzano et al. 2004). Para ello se genera una librería de proyecciones 2D del modelo 3D empleado como referencia y cada imagen experimental (partícula) es comparada con cada una de las proyecciones de la referencia. El coeficiente de correlación resultante de cada comparación permite asignar cada imagen experimental a una proyección 2D de la referencia y como resultado se asignan a cada imagen los ángulos asociados a dicha proyección. Una vez conocida la distribución angular de las imágenes, las partículas se retro-proyectan para generar un nuevo modelo 3D generado a partir de la información de las imágenes experimentales, y este nuevo modelo se usará como referencia para repetir el proceso de manera iterativa aumentando la exactitud de los ángulos asignados a cada imagen experimental. El resultado final es un mapa 3D de la muestra de estudio obtenido a partir de las imágenes experimentales, alejado de la referencia inicial.

1.5.4 Crio-ME en el estudio de los virus filamentosos flexibles

Gracias a la incorporación de las cámaras de detección directa y a la madurez de la técnica, la crio-ME ha adquirido a día de hoy un papel protagonista en el estudio estructural de macromoléculas biológicas a alta resolución (Bartesaghi et al. 2018, Merk et al. 2016).

Los viriones de los virus filamentosos flexibles son complejos proteína-RNA de gran tamaño, ideales para su estudio por crio-ME. Además, la simetría helicoidal que presenta la organización de las CPs para formar el virión supone una ventaja en el uso de esta técnica, pues esta simetría puede ser usada favorablemente para mejorar el resultado durante la reconstrucción tridimensional. Varios trabajos demuestran la utilidad de la crio-ME en la determinación de la estructura a alta resolución de viriones de virus filamentosos; se han conseguido mapas 3D de viriones de virus filamentosos rígidos como TMV a resoluciones por debajo de 4 Å (Clare and Orlova 2010, Fromm et al. 2015) y de viriones del *Virus del mosaico estriado de la cebada* (BSMV) a 4,1 Å (Clare et al. 2015). La crio-ME, a diferencia de la difracción de fibra de rayos X o la cristalografía de rayos X, ha supuesto un éxito a la hora de determinar la estructura a alta resolución de los viriones de virus filamentosos flexibles como PepMV y BaMV a 3,9 Å y 5,6 Å de resolución respectivamente (Agirrezabala et al. 2015, DiMaio et al. 2015) y se muestra como una herramienta de gran utilidad en su estudio estructural.

CAPITULO 2
OBJETIVOS E HIPÓTESIS

Se ha planteado la existencia de CPs con un plegamiento común en todas las familias que componen el grupo de virus filamentosos flexibles (Dolja et al. 1991, Kendall et al. 2008), sin embargo, hasta la fecha sólo ha sido caracterizada la estructura de las CPs pertenecientes a varios miembros de la familia *Alphaflexiviridae*, y la ausencia de datos relativos a otras familias ha hecho imposible comprobar esta hipótesis. La familia *Potyviridae* constituye un grupo de virus de plantas muy relevante, y el estudio estructural del virión de alguno de sus miembros amplía el conocimiento sobre la biología de estos virus y contribuye a identificar semejanzas con miembros de la familia *Alphaflexiviridae* con estructura descrita. Teniendo esto en cuenta, decidimos estudiar la organización estructural del virión de WMV y para ello nos propusimos los siguientes objetivos:

1. Caracterización estructural a alta resolución de viriones de WMV por crio-microscopía electrónica.
2. Obtención del modelo atómico para las proteínas de la cápsida y el ssRNA de WMV.
 - 2.1. Caracterización estructural de las interacciones proteína-proteína relevantes para el ensamblaje del virión de WMV.
 - 2.2. Caracterización estructural de las interacciones proteína-RNA relevantes para el ensamblaje del virión de WMV.
3. Análisis bioinformático de la conservación o variabilidad del plegamiento de la proteína de la cápsida de los virus filamentosos flexibles.

CAPÍTULO 3:
MATERIALES Y MÉTODOS

3.1 Inoculación y purificación de viriones de WMV

La inoculación y purificación de los viriones para su estudio estructural fue realizada por el laboratorio del profesor Miguel Aranda en el Centro de Edafología y Biología Aplicada del Segura (CBAS-CSIC).

Mediante abrasión con carborundo, los cotiledones de plántulas de calabacín (*Cucurbita pepo*) fueron inoculados con un homogeneizado compuesto por hojas secas de plantas infectadas con WMV (cepa M116) en un tampón de fosfato sódico 30 mM (pH 8.0). Las plantas inoculadas crecieron en invernadero durante 4 semanas (16 horas de fotoperiodo; 25-18°C día/noche). Las hojas jóvenes que presentaron síntomas de la infección por el virus fueron recolectadas para la purificación.

Para la purificación de los viriones a partir del material infectado, alrededor de 60 g de hojas fueron trituradas en nitrógeno líquido. El triturado de las hojas se mezcló con 320 mL de solución de homogeneización (K_2HPO_4 0,5 M, EDTA 5 mM, ácido dietilditiocarbámico 10 mM y Na_2SO_5 20 mM) en agitación durante 15 minutos a 4°C. Para retirar los restos de tejidos u otras impurezas de mayor tamaño el homogeneizado se centrifugó 10 minutos a 7500g y el sobrenadante obtenido se filtró a través de cinco capas de gasa. Al filtrado se añadió Triton X-100 a una concentración final de 1% v/v y se mezcló durante una hora en agitación a 4°C. Los viriones en solución tras la ruptura de las células fueron precipitados por ultracentrifugación a 300.000g durante 90 minutos a 4°C. El sedimento resultante fue resuspendido en solución citrato [citrato de sodio 50 mM y Na_2SO_3 20 mM (pH 7,5 ajustado con ácido cítrico)] con Triton X-100 1% mediante agitación durante 2 horas a 4°C. La solución resultante se clarificó mediante centrifugación a 9.000g durante 10 minutos a 4°C. El sobrenadante resultante de la clarificación se mezcló con cloroformo a una concentración final 10% v/v y se centrifugó a 15.000g durante 10 minutos a 4°C. La fase acuosa resultante, que contiene los viriones, se centrifugó a través de un colchón de sacarosa al 30% de 5 mL durante 110 minutos a 245.000g. El sedimento resultante se resuspendió en solución citrato durante toda la noche a 4°C y la solución resultante se clarificó mediante centrifugación a 9.000g durante 10 minutos a 4°C. Al sobrenadante resultante se le añadió Cs_2SO_4 (0,26 g/mL) y se formó un gradiente de densidad mediante centrifugación a 28.000 rpm durante 30 horas a 4°C en un rotor basculante SW 28. Los viriones de WMV quedaron concentrados en una banda turbia en el gradiente, que fue recogida con una jeringuilla. El volumen de la banda fue diluido con solución citrato y ultracentrifugado a 108.000g durante 90 minutos a 4°C. Finalmente, el sedimento resultante compuesto por los viriones de WMV purificados fue resuspendido en 200 μ L de solución citrato mediante agitación y guardado a 4°C.

3.2 Preparación de muestras para microscopía electrónica

3.2.1 Tinción negativa

Para la inspección y caracterización de los viriones de WMV a baja resolución se prepararon muestras para ser visualizadas mediante microscopía electrónica de transmisión usando el método de tinción negativa. Las muestras se prepararon sobre rejillas de cobre con una fina película de carbón que previamente fueron sometidas a descarga iónica con objeto de hidrofilar la superficie de la película de carbón para favorecer la adsorción de la muestra. A continuación, la rejilla se depositó sobre una gota de 10 μ L de muestra de viriones de WMV purificados de manera que la película de carbón estableciese contacto con la gota, se incubó durante 1 minuto y se lavó una vez con agua milli-Q para eliminar el exceso de muestra. El exceso de líquido se retiró de la rejilla mediante el contacto del borde de la misma con papel de filtro. Finalmente, las rejillas fueron teñidas con acetato de uranilo al 2% p/v durante 45 segundos y se secaron a temperatura ambiente.

3.2.2 Vitrificación

Para realizar el estudio estructural de los viriones de WMV a alta resolución por crio-ME se vitrificaron varias rejillas con la muestra purificada. Se utilizaron rejillas de cobre con capa de carbón microperforado formando agujeros dispuestos regularmente (Quantifoil R2/2) cubiertas con una fina capa de carbón. Tras someter las rejillas a descarga iónica se incubaron en presencia de 4-5 μ L de la muestra de viriones de WMV durante 1 minuto. Tras esta incubación se retiró el exceso de líquido de las rejillas con papel de filtro y rápidamente fueron sumergidas en etano líquido (-180°C) utilizando un VITROBOT (FEI). La elevada velocidad de congelación de la muestra sumergida en etano líquido evita la formación de cristales de hielo y la muestra biológica queda incluida en una capa de hielo vítreo que conserva su estado nativo. Las rejillas vitrificadas se mantuvieron en nitrógeno líquido a una temperatura aproximada de -170°C.

3.3 Obtención de imágenes en el microscopio electrónico de transmisión

3.3.1 Adquisición de imágenes por tinción negativa

Las rejillas fueron visualizadas en un microscopio electrónico de transmisión JEM-1230 (JEOL) estabilizado a 100 kV equipado con una cámara CCD Orius SC1000 (GATAN). Se obtuvieron varias micrografías a diferentes magnificaciones con el objetivo de inspeccionar la muestra y caracterizar el virión de WMV a baja resolución.

3.3.2 Adquisición de imágenes en crio-ME

Para comprobar la calidad de la vitrificación y el estado de la muestra, las rejillas fueron analizadas previamente usando un microscopio electrónico de transmisión JEM-2200FS/CR (JEOL) estabilizado a 200 kV equipado con una cámara CCD UltraScan 4000 SP (GATAN).

Para el estudio tridimensional a alta resolución del virión de WMV se visualizaron las rejillas vitrificadas usando un microscopio electrónico de transmisión Titan Krios FEI operando a 300kV y equipado con una cámara de detección directa K2 (Gatan) ubicado en NeCEN (Netherland Centre for Electron Nanoscopy). De manera automatizada gracias a la disposición regular de los agujeros de las rejillas, las imágenes se tomaron como un conjunto de 35 fotogramas cada una, durante un tiempo de exposición de 5 segundos y una dosis total de $35 \text{ e}^-/\text{Å}^2$, a una magnificación de 130.000X resultando un tamaño de pixel final de 1,1 Å.

3.4 Procesamiento digital de imágenes

3.4.1 Corrección de defectos causados por el movimiento de la muestra durante la adquisición de imágenes

Durante la adquisición de las micrografías, la incidencia del haz de electrones provoca cierto movimiento en la muestra que resulta en una pérdida de información a alta resolución que limita la calidad final del procesamiento. Para corregir esta limitación, las micrografías se adquieren como un conjunto de fotogramas que serán alineados para corregir los efectos del movimiento (Li et al. 2013). El tiempo de exposición de cada fotograma es lo suficientemente pequeño como para que el efecto del movimiento de la muestra sea despreciable (0,143 segundos por fotograma en nuestro caso). De modo que se alinearon un total de 28 fotogramas para cada micrografía (desde el fotograma 2 al 27) con el programa MotionCorr para mitigar la difuminación, resultando en micrografías corregidas compuestas por 28 fotogramas y con una dosis total de $28 \text{ e}^-/\text{Å}^2$.

3.4.2 Estimación de parámetros de la CTF

La CTF (Contrast Transfer Function) describe cómo las aberraciones del microscopio electrónico de transmisión y el mecanismo de formación la imagen que modulan el contraste en las micrografías. Durante el procesamiento es necesario corregir la CTF en las imágenes para paliar el efecto de la distorsión producida. Para llevar a cabo esta corrección, primero es necesario realizar una estimación de la CTF en cada imagen, que depende de varios parámetros: voltaje del microscopio, aberración esférica, contraste de amplitud, sub-enfoque de la imagen y astigmatismo. Varios de estos parámetros dependen de las características del microscopio, en cambio otros como el sub-enfoque y el astigmatismo son únicos de cada imagen y deben ser estimados. En nuestro caso, realizamos esta estimación con el programa CTFFIND3 (Mindell and Grigorieff 2003).

3.4.3 Alineamiento y reconstrucción helicoidal

Para conseguir la reconstrucción 3D del virión de WMV se empleó el procesamiento helicoidal, basado en el método de partículas individuales (SPA), implementado en RELION2 (He and Scheres 2017). La reconstrucción tridimensional de partículas individuales es uno de los métodos empleados en microscopía electrónica para estudiar la estructura de macromoléculas. Las imágenes obtenidas por crio-ME contienen miles de proyecciones 2D (partículas) que contienen información estructural de la macromolécula a estudiar. Estas proyecciones normalmente presentan orientaciones aleatorias y desconocidas, de modo que el conjunto de las diferentes partículas recogidas en las imágenes nos aportan información de las diferentes vistas de la macromolécula. Cuando nuestra muestra de estudio son filamentos helicoidales, estos deben ser identificados, delimitados en la micrografía y divididos en segmentos solapados que serán extraídos; en este caso cada segmento es equivalente a una partícula durante el procesamiento con el método SPA.

En primer lugar, se inspeccionaron las 466 micrografías obtenidas con el programa e2helixboxer.py en EMAN2 para descartar aquellas que presentasen viriones muy solapados o mucha contaminación con hielo. También se inspeccionó el espectro de frecuencias de las micrografías con RELION2 para descartar aquellas que presentasen un astigmatismo muy severo. Tras esta inspección, se seleccionaron 364 micrografías.

3.4.3.1 Selección de partículas

La selección de segmentos implica en primer lugar la delimitación de los filamentos que serán segmentados en la micrografía. Para la selección manual usamos el programa e2helixboxer.py de EMAN2 (Tang et al. 2007), que permite asignar manualmente un punto de origen y final en la micrografía para delimitar una región del filamento que deseamos segmentar, sin embargo, este método limita la selección a regiones rectas de los filamentos (Figura 3.1, A). La selección automática de segmentos se realizó con la herramienta “helical auto-picking” en RELION2, que permite identificar de manera automática los filamentos en la micrografía basándose en referencias aportadas por el usuario. En nuestro caso, como referencias usamos las medias 2D resultantes de los segmentos seleccionados de manera manual. La selección automática no impone la limitación de seleccionar sólo regiones perfectamente rectas (Figura 3.1, B), de modo que el número de segmentos finales es mayor al poder seleccionar filamentos que no son perfectamente rectos. En ambos casos, tras la

delimitación de los filamentos, estos son divididos en segmentos solapados, con una distancia entre segmentos mayor o igual al valor de la vuelta de hélice (en píxeles) que presente el filamento helicoidal; en nuestro caso, la distancia usada entre segmentos fue 32 píxeles. Los segmentos de los filamentos seleccionados tanto de manera manual como de manera automática fueron extraídos de la micrografía original con un tamaño de 230 píxeles x 230 píxeles con RELION2.

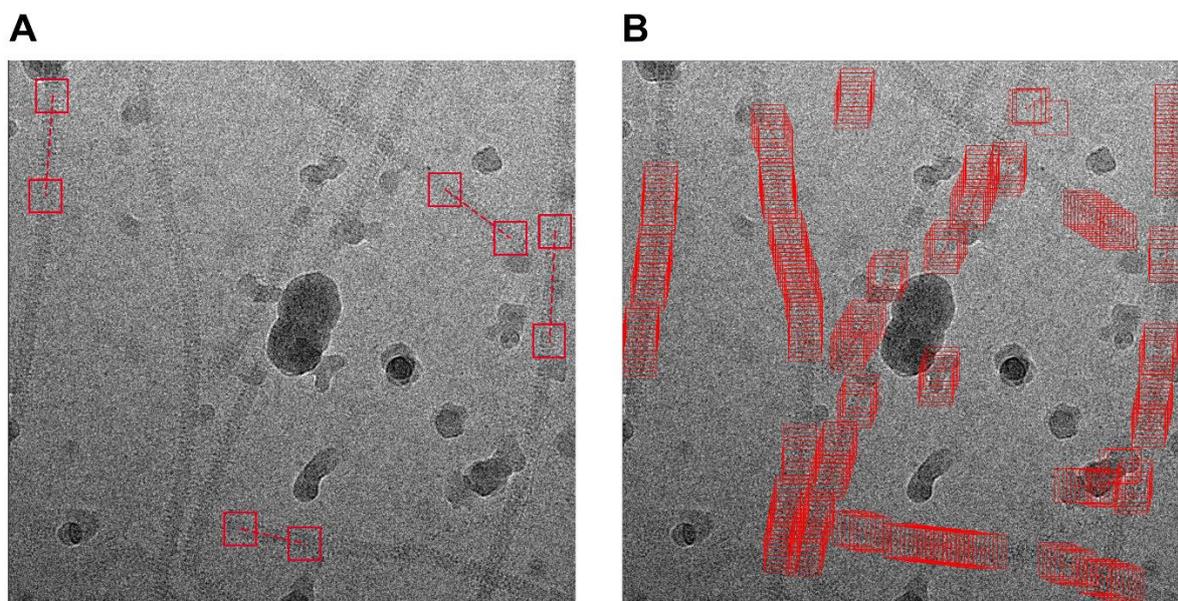


Figura 3.1: Comparación de métodos de selección de filamentos. **(A)** Método manual con EMAN2. Se delimita una región recta del filamento para posteriormente ser dividida en segmentos. **(B)** Selección automática de filamentos con RELION2. Basándose en las referencias aportadas, este método identificará filamentos en la micrografía que posteriormente dividirá en segmentos solapados con una distancia precisa entre ellos.

3.4.3.2 Ordenamiento y clasificación en 2D de segmentos helicoidales

Durante la selección automática de filamentos se incluyen falsos positivos que acaban formando parte del conjunto de partículas durante la extracción. Para eliminar segmentos extraídos que no corresponden a filamentos RELION permite ordenarlos en función de su similitud con las referencias empleadas, lo que permitió descartar los falsos positivos más evidentes.

La clasificación de las medias 2D del conjunto de segmentos basada en el método de máxima verosimilitud con el fin de diferenciar aquellas que son segmentos del virus de los falsos positivos menos evidentes se llevó a cabo con RELION2. Además esta herramienta permite

analizar la heterogeneidad en la muestra mediante alineamiento translacional y rotacional de las partículas y es capaz de agrupar las partículas en función de la orientación que representen.

3.4.3.3 Refinamiento y reconstrucción tridimensional con simetría helicoidal

Para la asignación de los parámetros de alineamiento se utilizó la herramienta de refinamiento 3D con simetría helicoidal de RELION2. Como referencia se usó un cilindro compacto sin detalles con un diámetro de 140 Å, similar al diámetro descrito para los viriones de los virus filamentosos flexibles (Kendall et al. 2008, Agirrezabala et al. 2015). Las búsquedas de los ángulos de Euler en este tipo de procesamiento implica varias particularidades: en primer lugar, dada la disposición de los filamentos en la muestra, se considera por defecto 90° el ángulo tilt (θ) de partida, y se realizan búsquedas locales en un intervalo de +/- 15°, en lugar de buscar en todo el espacio; al extraer los segmentos del filamento de la micrografía, el programa los orienta asignando un ángulo de rotación, de modo que este ángulo es el punto de partida para la búsqueda del ángulo psi (ψ), a partir del cual se realizarán búsquedas locales en un intervalo de +/- 15 °; por último, para determinar el ángulo rot (ϕ) se realizan búsquedas en todo el espacio (+/- 180°), como en los refinamientos 3D de macromoléculas asimétricas (He and Scheres 2017). La proyección de una estructura con organización helicoidal contiene información sobre varias vistas de los monómeros que la conforman debido a la simetría que presenta y los parámetros que definen esta simetría definen la orientación relativa de esas vistas. Teniendo esto en cuenta, en el procesamiento helicoidal a la hora de reconstruir el mapa de densidad 3D tras la asignación de los parámetros de alineamiento en cada iteración, cada partícula se usa varias veces aplicando una rotación y translación en el eje de simetría, o eje z, (Figura I.3) que depende de los parámetros de simetría y del número de unidades asimétricas por vuelta de hélice. Como resultado, en el mapa de densidad 3D obtenido, cada monómero que conforma la estructura helicoidal contiene información de todos los monómeros de la vuelta.

Finalmente, tras la obtención del mapa 3D, se realizó una compensación de la atenuación de las amplitudes en las altas frecuencias con RELION2. La resolución fue estimada por RELION2 durante cada iteración, en la iteración final y tras corregir el decaimiento de las altas frecuencias. Para ello, el programa divide el conjunto de datos en dos mitades y realiza el alineamiento de ambos de manera independiente resultando en la reconstrucción de dos mapas 3D. Calcula la FSC ("Fourier Shell Correlation") entre estos dos mapas que contienen la mitad del conjunto total de los datos cada uno, y se calcula la resolución como el recíproco de la frecuencia espacial donde el coeficiente de correlación es 0,143 (FSC = 0,143) (método

Gold-standard) (Scheres and Chen 2012). La resolución local de las distintas regiones del mapa de densidad 3D se estimó con el programa ResMap (Kucukelbir, Sigworth, and Tagare 2014) y RELION2.

Los parámetros de simetría empleados en el comienzo del procesamiento de datos fueron calculados previamente en el laboratorio con micrografías a baja resolución de viriones de WMV usando el programa SPRING (Desfosses et al. 2014). Estos parámetros resultaron ser similares a los descritos para los viriones del género *Potexvirus* con estructura determinada a alta resolución (Agirrezabala et al. 2015, DiMaio et al. 2015). Durante el refinamiento 3D realizado con RELION2, se llevó a cabo una optimización de estos parámetros realizando búsquedas de los valores más aptos de paso de hélice y ángulo azimutal en los intervalos de 0,2 Å y +/- 2° respectivamente (intervalos de +/-5% de los valores iniciales).

3.4.3.4 Clasificación 3D

Dado que el conjunto de partículas obtenido puede presentar heterogeneidad debido a diferentes estados o conformaciones de la macromolécula a estudiar, los métodos de clasificación se deben utilizar para obtener sub-poblaciones homogéneas de partículas. La clasificación 3D del conjunto de datos para buscar variabilidad en la organización estructural del virión de WMV se realizó mediante un método de clasificación no supervisada basado en procedimientos estadísticos de máxima verosimilitud con RELION2 (Scheres 2010).

3.5 Construcción del modelo atómico de la CP de WMV

Para la creación del modelo atómico se empleó el mapa de densidad 3D obtenido durante el refinamiento. En primer lugar, se realizó una simetrización del mapa 3D en el espacio real con RELION2 aplicando la zona central del mapa (el 30% de la longitud en eje de simetría del virión correspondiente a la zona central del mapa de densidad 3D) a los extremos (He and Scheres 2017).

La segmentación del mapa de densidad 3D total para obtener la densidad correspondiente a una subunidad o CP se realizó usando el método de Segger (Pintilie et al. 2010) en Chimera (Pettersen et al. 2004) y empleando el programa SPIDER (Shaikh et al. 2008).

Los modelos atómicos iniciales para la CP de WMV se obtuvieron con los servidores Robetta (<http://rosetta.bakerlab.org>) (Kim, Chivian, and Baker 2004) e I-TASSER (Zhang 2008). A partir del modelo atómico preliminar se modeló manualmente la estructura atómica de la CP de WMV y el fragmento de RNA correspondiente usando Coot (Emsley et al. 2010) empleando como referencia el mapa de densidad 3D de crio-ME. Con Phenix (Adams et al. 2010) se mejoró la estereoquímica del modelo atómico creado en Coot mediante refinamiento en espacio real usando como mapa de referencia la densidad segmentada correspondiente a una única CP. Tras varias rondas de modelado en Coot y optimización en Phenix, finalmente varios modelos atómicos de la CP conformando varias vueltas de hélice se refinaron conjuntamente sobre el mapa de crio-ME con Phenix empleando simetría no cristalográfica basada en la simetría del virión para mejorar las zonas de contacto entre subunidades adyacentes. El modelo atómico final que fue validado en MolProbity (Chen et al. 2010).

El cálculo del potencial electrostático superficial fue realizado con el servidor Bluues (<http://protein.bio.unipd.it/bluues>) (Fogolari et al. 2012).

3.6 Análisis de homología estructural

La búsqueda de homólogos estructurales al modelo atómico calculado para la CP de WMV se realizó en los servidores Dali (http://ekhidna.biocenter.helsinki.fi/dali_server) y Matras (Kawabata 2003, Holm and Rosenström 2010)

El alineamiento de pares de estructuras 3D, la búsqueda de residuos equivalentes en topología entre ellas y el cálculo del valor “TM-score”, que determina la similitud estructural entre ellas, se llevó a cabo con el algoritmo de TM-align (Zhang and Skolnick 2005) (<http://zhanglab.ccmb.med.umich.edu/TM-align>).

3.6.1 Cálculo de distancias estructurales

Para obtener la matriz con la medida de distancia estructural (SDM) entre cada par de estructuras se emplearon los parámetros obtenidos en la comparación de estructuras realizada con TM-align en las siguientes fórmulas (Agarwal et al. 2009, Lakshmi et al. 2015):

$$\text{SDM} = -100 * \log (w1 * \text{PFTE} + w2 * \text{SRMS})$$

$$w1 = [(1 - \text{PFTE}) + (1 - \text{SRMS})] / 2$$

$$w2 = (\text{PFTE} + \text{SRMS}) / 2$$

PFTE = número de residuos equivalentes en topología / tamaño de la proteína menor

$$\text{SRMS} = 1 - (\text{RMSD} / U)$$

RMSD es la desviación de la media cuadrática de la distancia entre los residuos equivalentes resultante del alineamiento de las estructuras. El valor U es la distancia mínima a partir de la cual dos residuos de dos estructuras alineadas se consideran equivalentes en topología; la distancia mínima considerada por TM-align son 8 Å.

3.7 Alineamiento de secuencias

En primer lugar, se obtuvieron las secuencias de aminoácidos de las CPs para todas las especies de virus filamentosos flexibles disponibles en la base de datos de secuencias de referencia (RefSeq) en NCBI (“National Center for Biotechnology Information”). Las secuencias se agruparon en función de la familia a la que pertenecen, de modo que, finalmente se obtuvieron 237 secuencias de CPs divididas en 5: *Alphaflexiviridae* (46 secuencias); *Closteroviridae* (29 secuencias); *Potyviridae* (107 secuencias); y la familia *Betaflexiviridae* dividida en dos grupos, *Betaflexiviridae* (sólo gen. Carlavirus) (44 secuencias) y *Betaflexiviridae* (resto de géneros) (11 secuencias). Las secuencias pertenecientes a la familia *Betaflexiviridae* se dividieron en dos grupos para el alineamiento ya que las secuencias del género Carlavirus presentan una distancia mayor entre los residuos conservados Ser y Arg en el presunto sitio de unión a RNA que el resto de géneros de la familia.

El alineamiento múltiple de las secuencias de cada grupo se realizó con el servidor Clustal Omega (Sievers et al. 2011) (www.ebi.ac.uk/Tools/msa/clustalo). Los archivos resultantes con las secuencias alineadas para cada grupo se usaron para generar la secuencia consenso de cada uno de ellos usando la aplicación WebLogo (Crooks et al. 2004, Schneider and Stephens 1990) (<http://weblogo.berkeley.edu>).

3.8 Códigos de acceso

El mapa de densidad 3D de crio-ME resultante de este trabajo fue depositado en la base de datos "Electron Microscopy Data Bank" (EMDB) (www.emdatabank.org) con el código de acceso EMD-3785. El modelo atómico calculado para la CP y el ssRNA de WMV a partir del mapa de densidad 3D fue depositado en la base de datos "Protein Data Bank" (PDB) (www.rcsb.org) bajo el código de acceso 5ODV.

CAPÍTULO 4
RESULTADOS

4.1 Análisis estructural de viriones de WMV por crio-ME

En este estudio se llevó a cabo la caracterización estructural de viriones de WMV por crio-ME. Para ello, fue necesario la obtención de una muestra lo más homogénea posible de partículas virales. Varias plántulas de calabacín (*Cucurbita pepo*) fueron inoculadas con el virus mediante métodos abrasivos y se dejaron crecer durante varias semanas hasta que adquirieron el tamaño adecuado. Se recogieron las hojas jóvenes de las plantas que presentaban síntomas de infección y fueron homogeneizadas para purificar viriones de WMV a partir de ellas. El homogeneizado de las hojas se clarificó mediante centrifugación y filtrado para eliminar los restos celulares más grandes. A continuación, un primer paso de ultracentrifugación permitió separar los elementos más grandes como son las partículas virales del resto de proteínas de la célula, seguido de una separación de fase acuosa y orgánica con cloroformo que eliminó restos de membranas celulares que pudieran haber precipitado junto con las partículas virales. Finalmente, varios pasos de ultracentrifugación incluyendo un colchón de sacarosa y un gradiente de cesio permitieron obtener una muestra homogénea de viriones de WMV (Figura 4.1). El proceso de inoculación de las plantas con el virus así como la purificación de las partículas se llevó a cabo en el laboratorio del profesor Miguel Aranda en el Centro de Edafología y Biología Aplicada del Segura (CBAS-CSIC).

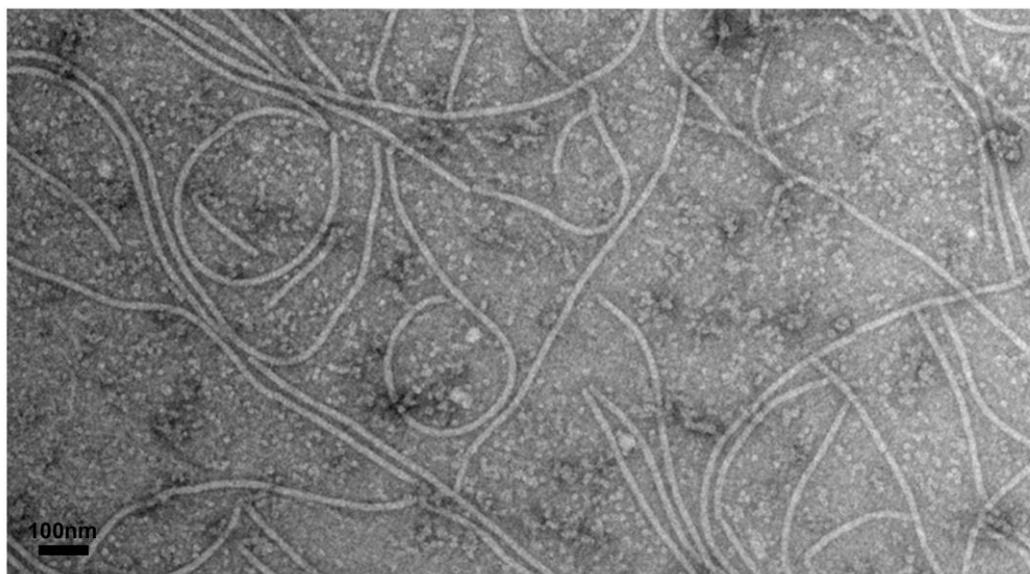


Figura 4.1: Micrografía de tinción negativa de filamentos de WMV.

Con la muestra de viriones de WMV se prepararon rejillas para ser visualizadas por tinción negativa en un microscopio electrónico estabilizado a 100 kV. En imágenes de microscopía electrónica, la muestra purificada de viriones de WMV presenta una concentración adecuada de partículas virales intactas (Figura 4.1 y 4.2, A). Aunque presenta otros elementos, que probablemente sean restos de partículas virales o intermediarios del ensamblaje del virión de WMV, la muestra presenta una homogeneidad adecuada para su estudio por microscopía electrónica. Los viriones de WMV purificados muestran características típicas de los virus filamentosos flexibles (Figura 4.1), son filamentos de aproximadamente 850 nm de longitud y 15 nm de diámetro que presentan una alta flexibilidad, muy similares al resto de especies del género *Potyvirus* (Wylie et al. 2017).

4.1.1 Reconstrucción del virión de WMV por crio-ME

Como todos los virus filamentosos, las proteínas que conforman la cápside de WMV se disponen de manera simétrica formando una hélice. La naturaleza de estos virus nos permite realizar un procesamiento de datos de crio-ME aprovechando esta marcada simetría helicoidal que presentan. El estudio de material biológico por crio-ME implica que la muestra esté inmersa en hielo vítreo sostenido en una rejilla de cobre. Muestras de 4-5 μL de viriones de WMV purificados fueron incubadas en contacto con rejillas Quantifoil R2/2 con capa de carbón agujereada y rápidamente fueron vitrificadas por inmersión en etano líquido usando un vitrobot FEI. Los agujeros presentes en la capa de carbón contienen la muestra inmersa en hielo vítreo y son las zonas de la rejilla donde se obtienen las micrografías. La disposición regular de estos agujeros en la rejilla permitió la adquisición de 466 micrografías de alta resolución de manera automática empleando el programa EPU en un microscopio Titan Krios FEI operando a 300kV y equipado con una cámara de detección directa K2 (Gatan) ubicado en NeCEN (“The Netherlands Centre for Electron Nanoscopy”). Tras la inspección de las micrografías obtenidas fueron descartadas aquellas con un espectro de frecuencias que mostraba defectos en la imagen, o las que presentaban demasiados viriones solapados por exceso de concentración en una zona, así como aquellas que no presentaban ningún virión, quedando finalmente 364 micrografías aptas para la selección de viriones.

Para llevar a cabo el procesamiento de datos de crio-ME es necesario seleccionar el mayor número de filamentos de los viriones presentes en las micrografías para posteriormente dividirlos en segmentos solapados a distancias regulares y usarlos como partículas en el método de análisis de partículas individuales. Inicialmente comenzamos la selección de

filamentos de forma manual con EMAN2 (Tang et al. 2007), que nos permitió seleccionar regiones rectas en los viriones de longitud variable, y resultó en la obtención de 915 filamentos helicoidales que fueron divididos en 13.677 segmentos de aproximadamente 25 nm de longitud. Con este conjunto de segmentos llevamos a cabo procesamientos iniciales usando protocolos de reconstrucción helicoidal en RELION2 (He and Scheres 2017), y como resultado obtuvimos mapas de densidad 3D a 4,36 Å de resolución global. A continuación, y con el objetivo de aumentar el número de partículas y poder mejorar la resolución, realizamos una nueva búsqueda en las micrografías usando el protocolo de selección automática de filamentos helicoidales de RELION2. A diferencia de EMAN2, este método no está limitado a la selección de filamentos perfectamente rectos, sino que también es capaz de seleccionar zonas de los filamentos que presentan ligeras curvaturas (Figura 3.1). De este modo podemos enriquecer nuestro conjunto de datos ya que aumentamos el número de filamentos y la longitud de los mismos, y por lo tanto el número total de segmentos. Con este método conseguimos seleccionar 50.045 segmentos que se usaron como conjunto de datos inicial para realizar un procesamiento 2D. Este procesamiento clasifica las diferentes partículas del conjunto de datos en diferentes grupos en función de la proyección del segmento de virión que represente cada una de ellas. Como resultado obtenemos diferentes grupos que contienen información de diferentes proyecciones de un segmento del virión (o el complejo de estudio) y el conjunto de segmentos de cada grupo constituye una media 2D de esa proyección (Figura 4.2). Gracias a este procesamiento pudimos descartar partículas que no eran realmente segmentos de virión, sino falsos positivos resultantes del método de selección automática, o segmentos de virión que no presentan unas características adecuadas debido a la mala calidad del hielo vítreo o a la presencia de contaminación. Como resultado del procesamiento 2D pudimos aislar 39.100 segmentos de virión que constituyen las medias 2D de las proyecciones mejor definidas (Figura 4.2, B).

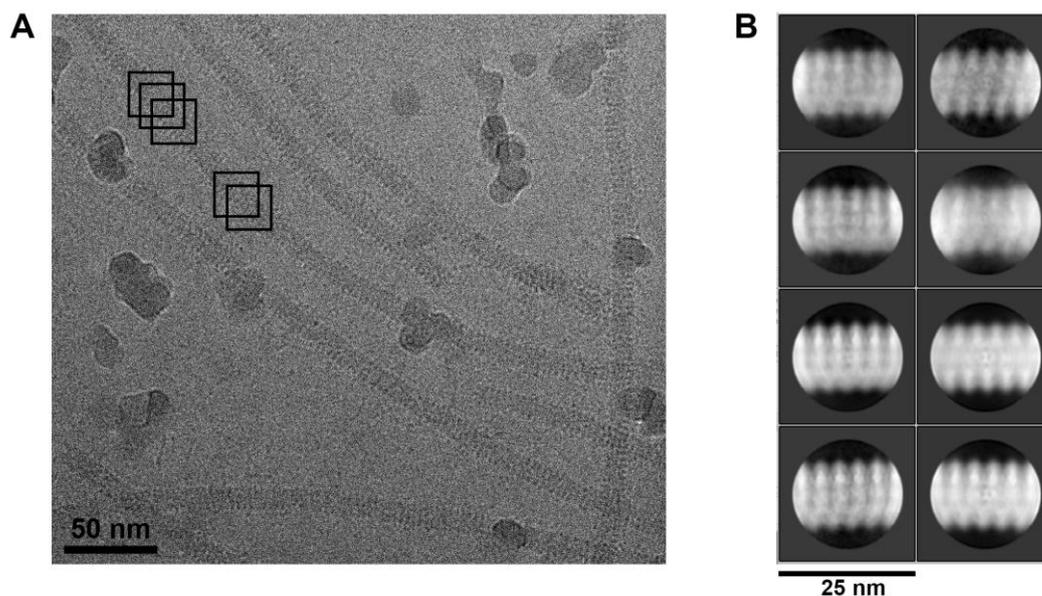


Figura 4.2: (A) Micrografía de crio-microscopía electrónica de filamentos de WMV. La selección de segmentos de filamentos en la micrografía está representada con cajas de línea negra. (B) Medias 2D de diferentes grupos de segmentos resultantes del procesamiento 2D realizado con RELION2.

Este nuevo conjunto de segmentos seleccionados tras el procesamiento helicoidal con RELION resultó en un mapa de densidad 3D a 4,36 Å de resolución global, igual que el conjunto de datos anterior. El procesamiento helicoidal está basado en la técnica de partículas individuales, pero tiene en cuenta los parámetros de simetría definidos por el ordenamiento helicoidal de las proteínas de la cápside para constituir el virión maduro. Estos parámetros fueron calculados previamente en nuestro laboratorio con segmentos de virión de WMV procedentes de micrografías a baja resolución usando el programa SPRING (Desfosses et al. 2014), que compara en espacio real las diferentes posibilidades de parámetros de simetría dentro de un rango amplio y determina cuales son los más adecuados respecto a las medias 2D de los segmentos. Los parámetros de simetría preliminares más coherentes con el conjunto de datos a baja resolución fueron 8,8 subunidades por vuelta de hélice y 35,60 Å de paso de hélice, muy similares a valores de simetría determinados para otros virus filamentosos flexibles (Kendall et al. 2008, Agirrezabala et al. 2015, DiMaio et al. 2015), y estos fueron los empleados inicialmente en el procesamiento helicoidal. Por otro lado, RELION ofrece la posibilidad de optimizar estos valores de simetría iterativamente durante el procesamiento dentro de unos intervalos relativamente pequeños. De modo que durante el procesamiento helicoidal de los datos a alta resolución con RELION los valores de simetría fueron optimizados resultando en unos valores finales de 8,81 subunidades por vuelta de hélice y 35,15 Å de paso de hélice.

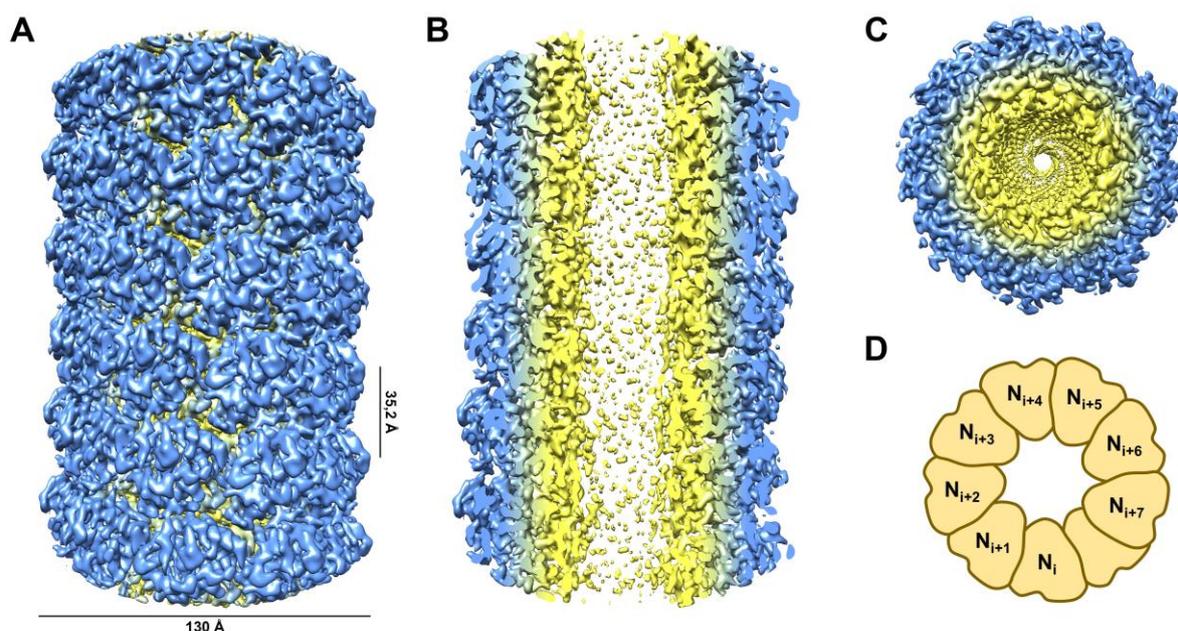


Figura 4.3: Mapa de densidad 3D de un segmento del virión de WMV. El mapa 3D presenta colores con una distribución radial respecto al eje de longitudinal del tubo; de este modo los colores azul y amarillo contrastan el exterior e interior del virión respectivamente. **(A)** Vista frontal del segmento del virión. **(B)** Vista frontal de una sección central del segmento del virión de 30 Å aproximadamente. **(C)** Vista superior del mapa de densidad 3D que muestra el interior del segmento del virión. **(D)** Esquema de la organización de las subunidades (proteína de la cápside) que conforman una vuelta de hélice en la estructura de ensamblaje del virión.

Tal y como vemos en el mapa de densidad 3D resultante del procesamiento, la estructura helicoidal del virión de WMV constituye una hélice levógira de 130 Å de diámetro exterior que contiene 8,8 unidades asimétricas (proteína de la cápside) por cada vuelta y un paso de hélice de 35.2 Å (Figura 4.3), características muy similares a las estructuras helicoidales de otros virus filamentosos flexibles (Agirrezabala et al. 2015, Kendall et al. 2008, DiMaio et al. 2015). Los extremos superior e inferior del mapa muestran peor calidad debido a las características del procesamiento helicoidal, ya que la zona central del mapa procede de la zona central del segmento, que es la parte que mejor se alinea durante el procesamiento. Igualmente, las densidades del interior del tubo en el mapa 3D, atribuibles a partes de la CP del virus, no presentan la misma calidad que el resto del mapa. Dada la heterogeneidad en la calidad observada en el mapa 3D decidimos calcular las resoluciones locales, y como resultado, pudimos identificar diferencias en su resolución local de más de 1 Å, alcanzando en las zonas centrales del mapa 3,4 Å de resolución (Figura 4.4).

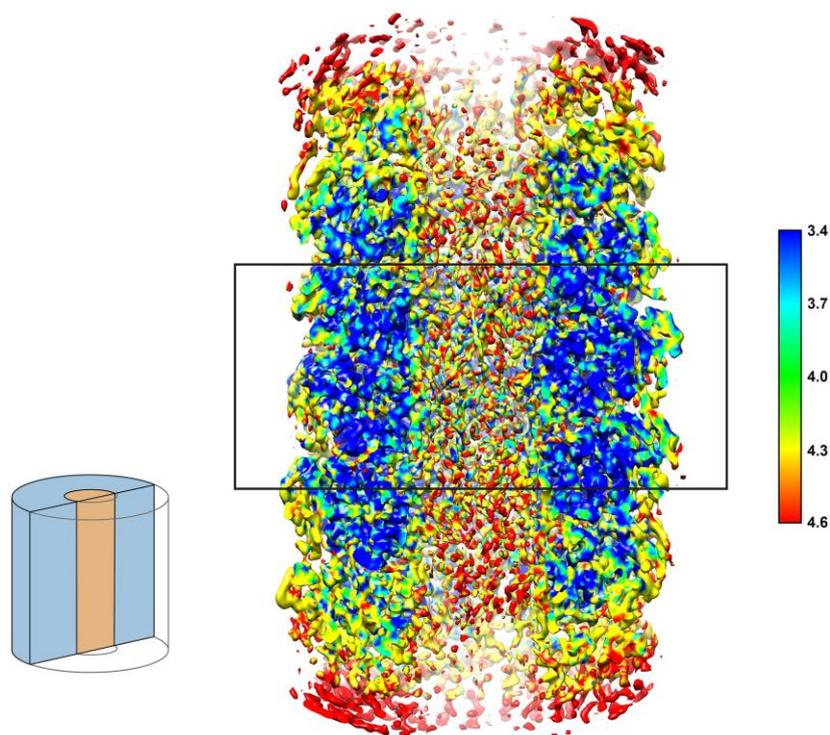


Figura 4.4: Corte longitudinal del mapa de densidad 3D del virión de WMV (esquema del corte a la izquierda). Se muestran las distintas densidades interiores del mapa de densidad 3D coloreadas en función de la resolución local que presentan calculada con RELION2. El recuadro muestra delimita la zona del mapa empleada para el cálculo de la FSC enmascarada.

El mapa de densidad 3D resultante del conjunto de segmentos procedente de la selección automática realizada con RELION2 muestra unas características muy similares al mapa de densidad 3D resultante del conjunto de segmentos procedente de la selección manual con EMAN2. Como vemos en la Figura 4.5, A los valores de resolución globales para ambos mapas de densidad 3D son idénticos (umbral FSC a 0,143) pese a las diferencias entre los conjuntos de datos. Sin embargo, al comparar los valores de la FSC calculados sólo para la zona central de los mapas de densidad 3D en los dos conjuntos de datos (FSC enmascarado; abarcando sólo el 30% central del mapa de densidad 3D como muestra el recuadro de la Figura 4.4), la resolución mejora de 4,25 Å en el conjunto de datos seleccionados de forma manual a 4,05 Å en el conjunto de datos seleccionados de forma automática (Figura 4.5 B).

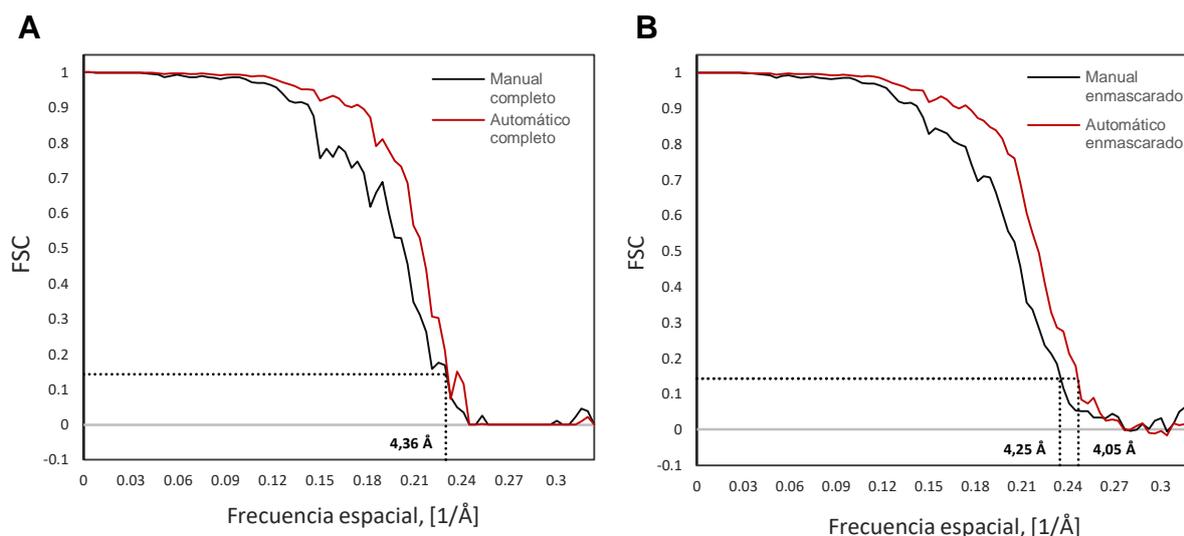


Figura 4.5: Curvas FSC (método “Gold standard”) de los volúmenes reconstruidos en los procesamientos helicoidales para comparar los métodos de selección de partículas (línea negra para partículas seleccionadas con el método manual de EMAN2 y línea roja para partículas seleccionadas con el método automático de RELION2). Curvas FSC de los volúmenes completos (**A**) y curvas FSC de los volúmenes enmascarados en la zona central, donde la reconstrucción 3D del mapa presenta mejor calidad (**B**). Las líneas de puntos señalan la resolución ($1/\text{Frecuencia espacial}$) para cada volumen a un umbral de FSC igual a 0,143.

Ya que las partículas virales de WMV, al igual que el resto de virus filamentosos flexibles, presentan bastante flexibilidad, quisimos comprobar si existe cierta heterogeneidad en la estructura del ensamblaje de las CPs de estos virus o en los valores que definen la simetría del ensamblaje. Para ello utilizamos el protocolo de clasificación helicoidal 3D de RELION2, que nos permite separar posibles diferencias estructurales e identificar posibles variaciones en los valores que definen la simetría del ensamblaje del virus. Como resultado obtuvimos una población mayoritaria (93% de los segmentos) cuyo mapa de densidad 3D presenta unos valores de simetría idénticos a los valores utilizados en el procesamiento helicoidal inicial y ninguna diferencia estructural con el mapa de densidad 3D inicial. Otras dos poblaciones minoritarias resultaron de la clasificación sin presentar tampoco diferencia alguna en la estructura del ensamblaje, sin embargo, los valores de simetría eran ligeramente distintos (Tabla 4.1). Ya que el conjunto mayoritario de partículas (93%) presenta la misma simetría, nos induce a pensar que los viriones de WMV poseen una simetría bastante estable y homogénea, y que la estructura de las proteínas de la cápside no presenta distintas conformaciones en los viriones maduros.

	Elevación de hélice	Ángulo azimutal	Número de segmentos	%
Clase 1	3,97	-40,85	1.955	5
Clase 2	3,99	-40,87	36.363	93
Clase 3	3,93	-40,85	782	2

Tabla 4.1: Resultado de la clasificación helicoidal 3D de RELION2 usando todo el conjunto de segmentos empleados para la obtención del mapa de densidad 3D inicial. Los parámetros helicoidales “elevación de hélice” y “ángulo azimutal” son usados por RELION2 y describen la simetría de la hélice; de modo que, grupos representativos con diferentes parámetros helicoidales indicarían la presencia de ligeras desviaciones de la simetría helicoidal dentro del conjunto de datos.

El mapa de densidad 3D resultante del procesamiento helicoidal presenta mejor calidad en la zona central (Figura 4.4), por lo tanto, con el objetivo de tener un mapa de densidad uniforme a la hora de caracterizar la estructura del virión, se realizó una simetrización del mapa 3D en el espacio real para aplicar las características de la zona central bien definida del mapa 3D a los extremos que presentan peor calidad (Figura 4.6). El mapa resultante se empleó para definir la unidad asimétrica (CP) y para generar el modelo atómico de la CP y el RNA genómico del virus.

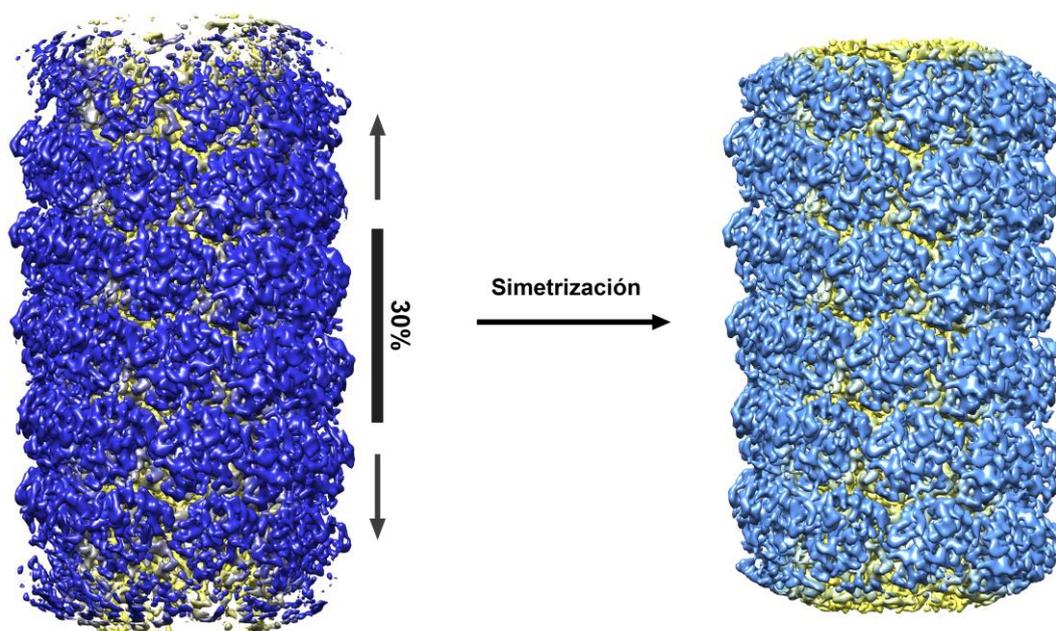


Figura 4.6: Simetrización del mapa de densidad 3D resultante del procesamiento helicoidal. En el mapa sin simetrizar (izquierda) una barra negra señala la parte central correspondiente al 30% del mapa empleado para la simetrización. En el mapa simetrizado (derecha) se aprecian los extremos mejor definidos.

4.1.2 Unidad asimétrica y ensamblaje del virión

El mapa de densidad 3D obtenido en el procesamiento corresponde a un fragmento de unos 240 Å de longitud de la cápside de WMV, empaquetando el RNA genómico, que abarca 4 vueltas de hélice completas. Aunque el complejo biológico que estamos estudiando está formado por varios elementos individuales, el mapa de crio-ME se reconstruye como una densidad conjunta, por ello, para definir la parte de la densidad global correspondiente a cada unidad asimétrica individual de la cápside (CP) y el RNA genómico fue necesario segmentar el mapa de densidad. La segmentación manual de mapas de densidad 3D puede ser un proceso subjetivo cuando el mapa de densidad no alcanza la suficiente resolución para que se distingan claramente los límites de todos los elementos que constituyen el complejo. Gracias a la alta resolución en nuestro mapa de densidad pudimos identificar correctamente los límites de cada unidad asimétrica (cada CP) y del RNA genómico en la región central, y realizar una segmentación del mapa con UCSF Chimera usando el método de Segger (Pettersen et al. 2004, Pintilie et al. 2010). El análisis visual del mapa de densidad 3D nos permitió reconocer aproximadamente la disposición de las proteínas de la cápside en el virión y el RNA genómico, sin embargo la segmentación de densidades más rigurosa va ligada a la creación del modelo atómico, y viceversa. El modelo atómico está condicionado a las densidades del mapa 3D, pero a su vez es el modelo atómico el que nos orienta para determinar la exactitud de la segmentación, y este proceso iterativo nos permitió delimitar al detalle los límites entre las proteínas de la cápside y el RNA genómico del virus.

El mapa de densidad 3D nos muestra que cada CP está formada por un dominio bien estructurado, con densidades que parecen pertenecer a hélices α en la estructura secundaria, que constituye el cuerpo principal de cada CP en la formación del virión. Los extremos terminales que parten de este cuerpo central interactúan con las subunidades vecinas mediando la polimerización entre ellas para posibilitar el ensamblaje del virión, al igual que ocurre en los virus del género *Potexvirus* (familia *Alphaflexiviridae*) (Agirrezabala et al. 2015, DiMaio et al. 2015). Del mismo modo, la naturaleza de las interacciones de estos brazos terminales con las subunidades adyacentes permitiría ciertos movimientos entre ellas, lo que explicaría la flexibilidad de estos virus. Basándonos en modelos atómicos descritos para varios miembros del género *Potexvirus* y en modelos atómicos para la proteína de la cápside (CP) de WMV obtenidos con distintos servidores de predicción como Robetta e I-TASSER (Zhang 2008, Kim, Chivian, and Baker 2004) pudimos inferir de forma preliminar la localización exterior e interior de los extremos N- y C- terminal respectivamente en nuestro mapa de densidad, ya que coincide con la disposición observada en *Potexvirus*. En el mapa de densidad 3D para el virión de WMV vemos como cada extremo N-terminal, que parte del

cuerpo central de la CP, está ubicado en el exterior del virión interaccionando con subunidades adyacentes (Figura 4.7, A y D). En cambio, el extremo C-terminal se extiende longitudinalmente al eje de simetría en sentido contrario al N-terminal. Está ubicado en el interior de la cápside y en su disposición llega a alcanzar subunidades ubicadas dos vueltas de hélice en adelante (Figura 4.8).

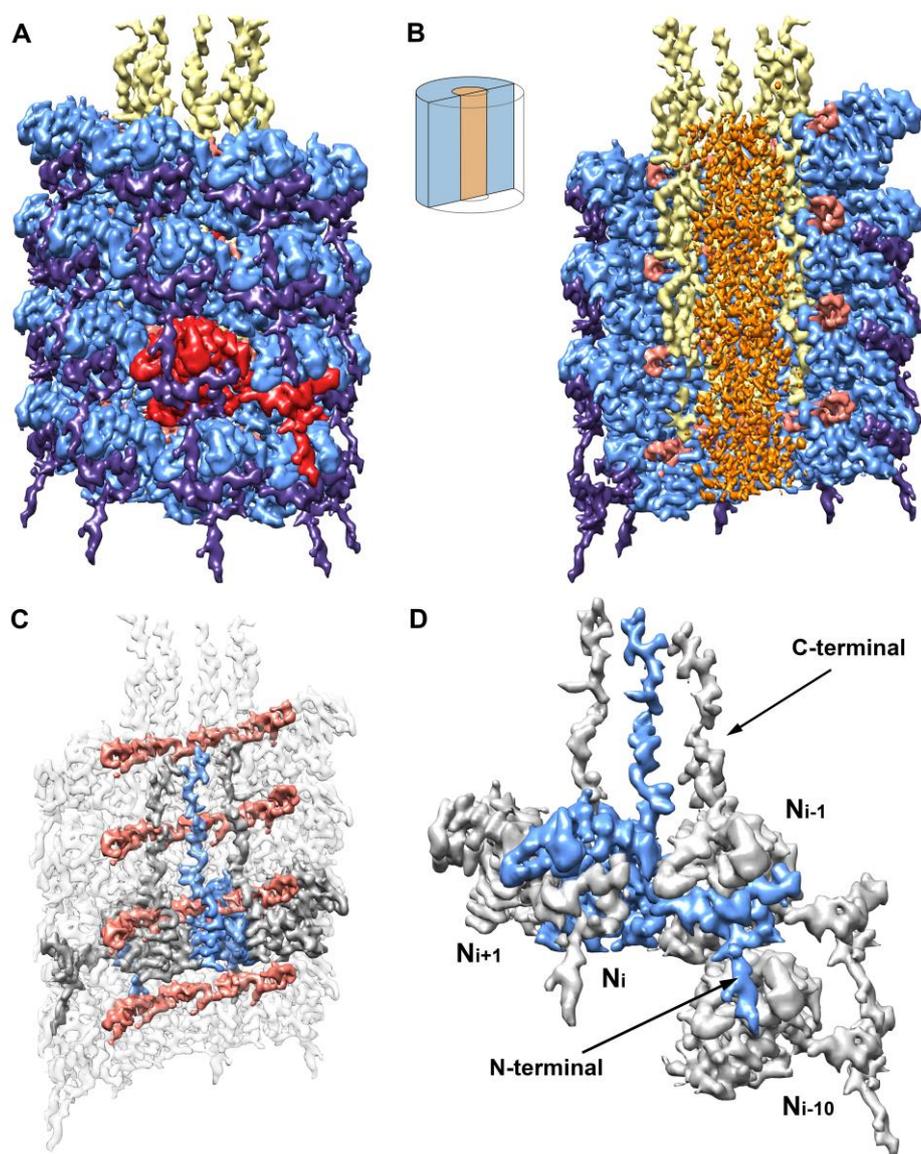


Figura 4.7: Ensamblaje del virión de WMV. Vista frontal del mapa de densidad 3D segmentado **(A)** y vista interior mediante corte longitudinal **(B)**. El código de colores en **(A)** y **(B)** delimita el cuerpo central ordenado de la CP (azul), el extremo N-terminal orientado hacia el exterior del virión (morado), el extremo C-terminal orientado hacia el interior del virión (amarillo), el RNA genómico del virus (rojo claro) y densidad interior del filamento (naranja). Además, en el panel **(A)** se muestra una subunidad (CP) completa coloreada en rojo. **(C)** Corte longitudinal, al igual que en **(B)**, del mapa de densidad que muestra las densidades correspondientes al RNA genómico (rojo claro). El resto de densidades correspondientes a las CPs se muestran semitransparentes, a excepción de tres (dos en gris y una en azul). **(D)** Detalle de ensamblaje entre subunidades. La CP de referencia (N_i), coloreada en azul, posiciona su extremo N-terminal en la subunidad adyacente (N_{i-1}) llegando a contactar con subunidades de la vuelta de hélice anterior (N_{i-10}).

Al segmentar el mapa de densidad 3D podemos identificar una densidad continua en el interior de la cápside que conserva la estructura helicoidal del virión. Esta densidad se corresponde con el RNA genómico de WMV y está íntimamente rodeada por las densidades correspondiente a las proteínas de la cápside (Figura 4.7, C), y en las zonas centrales del mapa 3D alcanza la mayor resolución presente en nuestro mapa 3D (Figura 4.4).

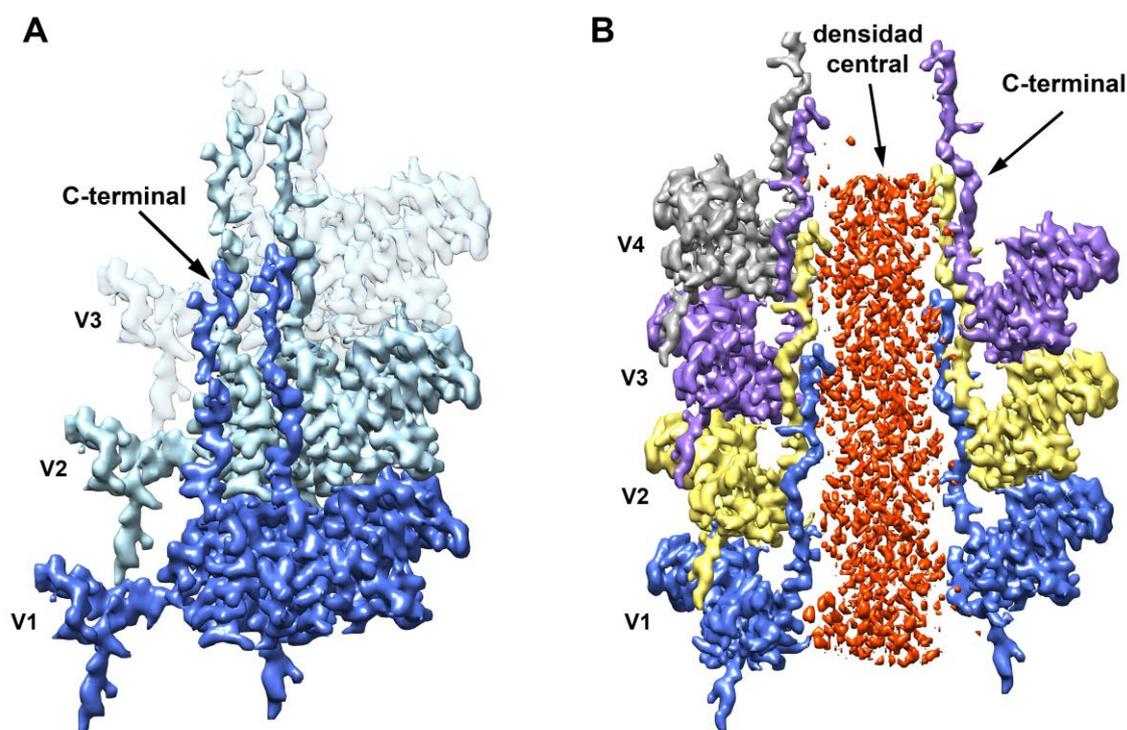


Figura 4.8: Disposición del extremo C-terminal en el interior del virión de WMV. **(A)** Dos CP (azul oscuro) de una vuelta de hélice determinada (V1) disponen sus extremos C-terminal longitudinalmente al eje de simetría del filamento estableciendo contactos con los extremos C-terminal de otras CP (azul claro) de la vuelta de hélice siguiente (V2) y llegando a alcanzar subunidades (azul claro transparente) ubicadas dos vueltas de hélice en adelante (V3). **(B)** Sección central del virión de WMV en la se muestran subunidades pertenecientes a 4 vueltas de hélice consecutivas (V1, azul; V2, amarillo; V3, violeta; y V4, gris) donde de nuevo se aprecia la disposición longitudinal del C-terminal de la CP. La densidad central (naranja) no presenta la calidad que el resto de las densidades atribuidas a las CP y el RNA genómico, sin embargo, probablemente pertenezca

4.2 Generación del modelo atómico de la proteína de la cápside y RNA genómico de WMV

Para inferir la estructura de la proteína de la cápside de WMV, se realizó una predicción del nivel de plegamiento y desorden de esta con el servidor DISOPRED3 (Jones and Cozzetto 2015). En esta predicción se identifica una región central estructurada rodeada por dos extremos desordenados (Figura 4.9). Al analizar las segmentaciones del mapa de densidad identificamos que la proteína de la cápside de WMV presenta un dominio central con estructura secundaria y terciaria claramente definida y dos largos extremos N-terminal y C-terminal, coincidiendo con las predicciones realizadas.

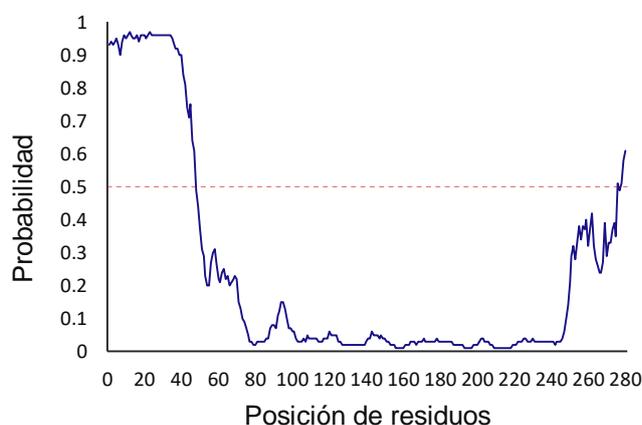


Figura 4.9: Perfil de predicción de desorden de la proteína de la cápside de WMV calculado con DISOPRED3. En la gráfica se muestra la probabilidad de constituir una zona desordenada en la proteína (eje vertical) para cada residuo (numerados de 1 a 280 en el eje horizontal). Un valor superior a 0,5 indicaría que ese residuo probablemente esté constituyendo una zona desordenada en la proteína.

Dado que el mapa de densidad 3D alcanzó una resolución por debajo de 4 Å en algunas zonas, se pueden distinguir fácilmente los elementos de estructura secundaria de la CP e identificar en algunos casos cadenas laterales de gran tamaño en algunos residuos (Figura 4.10), lo que nos permitió calcular un modelo atómico para la proteína de la cápside y el RNA genómico del virus. Para ello, usamos los servidores Robetta (Kim, Chivian, and Baker 2004) e I-TASSER (Zhang 2008) con el objetivo de generar un modelo inicial que nos sirviera para empezar a trabajar manualmente sobre el mapa de coordenadas. Los servidores de predicción de estructura secundaria como PSIPRED (McGuffin, Bryson, and Jones 2000), al igual que Robetta e iTasser, fueron capaces de ubicar en la secuencia de la proteína la zona central rica en α -hélices fácilmente identificable en nuestro mapa de densidad. A partir de este modelo inicial, continuamos manualmente con la construcción del modelo atómico que se ajustase a nuestro mapa 3D usando Coot (Emsley et al. 2010) y a continuación este modelo atómico fue refinado con PHENIX (Adams et al. 2010). Gracias a la calidad del mapa pudimos trazar la totalidad de la región central ordenada y gran parte de los extremos N- y

C-terminal que intervienen en el ensamblaje (Figura 4.10). La visualización de densidades correspondientes a cadenas laterales en muchas regiones del mapa 3D nos permitió ubicar residuos clave de la proteína de la cápside para colocar en registro la secuencia del modelo atómico respecto al mapa de densidad 3D (Figura 4.10, C). Para completar la optimización del modelo atómico, con Chimera se colocaron varias copias del modelo atómico distribuidas acorde a la simetría que presenta el virión conformando tres vueltas de hélice en un mismo archivo de coordenadas, y se refinaron conjuntamente usando Phenix. De este modo, obtuvimos un modelo atómico de la CP optimizado en contexto con las CPs adyacentes formando parte del virión maduro.

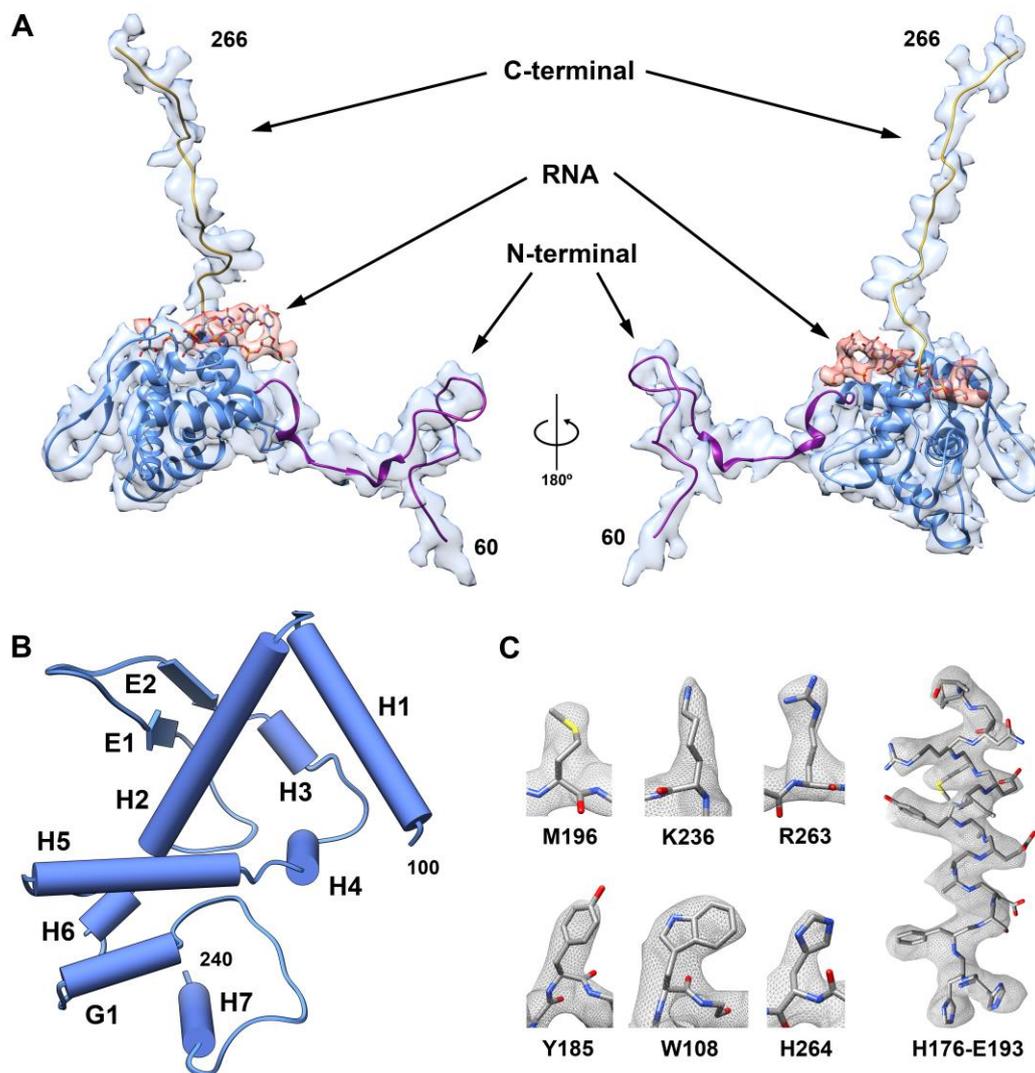


Figura 4.10: Modelo atómico de proteína de la cápside de WMV unido a un fragmento de RNA genómico. **(A)** Modelo atómico generado a partir del mapa de densidad correspondiente a la proteína de la cápside (azul semitransparente) y el fragmento de RNA genómico (rojo semitransparente). La zona central del modelo atómico está coloreada en azul, el extremo N-terminal en morado y el extremo C-terminal en amarillo. **(B)** Modelo de cilindros de la zona central organizada de la proteína de la cápside generado a partir del modelo atómico. **(C)** Detalle de densidades correspondientes a cadenas laterales de determinados residuos del modelo atómico.

Como se aprecia en la Figura 4.10, A, una región importante de unos 60 aminoácidos en el extremo N-terminal no se pudo modelar debido a la ausencia de densidad correspondiente en el mapa 3D. Esta zona se sitúa en la parte exterior de la cápside del virus y probablemente esté ausente en nuestro mapa porque es una región desordenada y presenta una alta flexibilidad, tal y como predice DISOPRED3 (Figura 4.9). Algo similar ocurre con el extremo C-terminal, en el que alrededor de 17 aminoácidos no se pudieron modelar pese a existir densidad atribuible a ellos (Figura 4.8, panel B).

Podemos acotar la zona central de la proteína de la cápside desde el residuo A100 hasta el L240. Como podemos ver en la Figura 4.10, panel B, esta región presenta una estructura globular y está compuesta por 8 hélices, de las cuales 7 poseen características de hélices α , y dos pequeñas láminas β (E1 y E2) formando una horquilla β orientada hacia el exterior del virión y ubicadas entre la hélice α H2 y la hélice α H3. La octava hélice que forma parte de esta zona globular, está situada entre la hélice α H6 y la hélice α H7 y presenta una estructura de hélice 3_{10} (G1) en el modelo atómico (Figura 4.10, B). Por un lado, las hélices α H1, H2, H3 y H4 junto con las láminas β y por otro las hélices α H2, H5 y H6 junto con la hélice 3_{10} G1, delimitan dos zonas internas de la región globular de la proteína altamente pobladas con residuos de cadenas laterales hidrofóbicas (Figura 4.11).

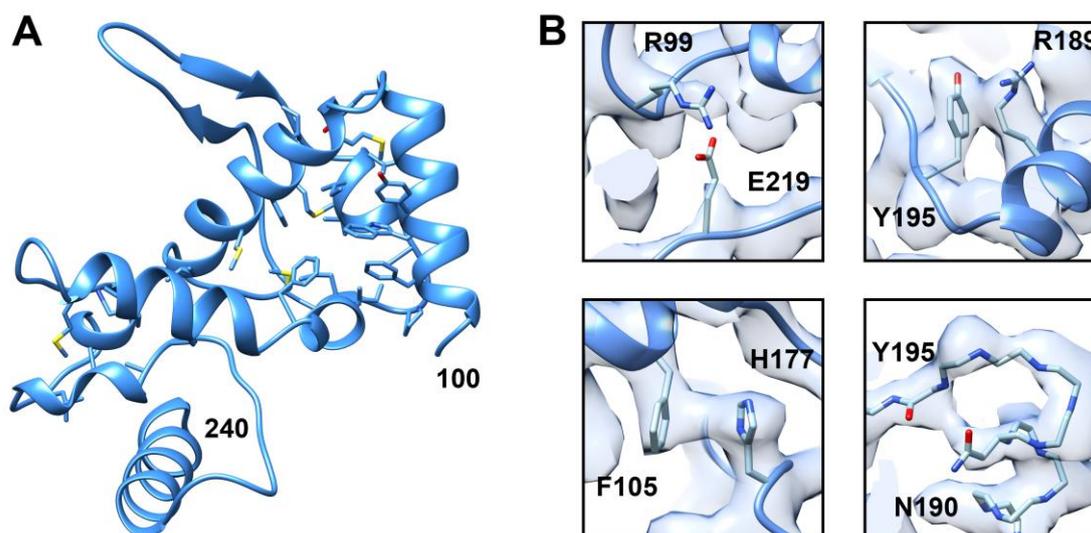


Figura 4.11: Región del modelo atómico correspondiente a la zona central ordenada de la proteína de la cápside de WMV. **(A)** Detalle de los residuos hidrofóbicos ubicados en la zona interior de la estructura terciaria de la CP. **(B)** Detalle de densidades correspondientes a cadenas laterales de diferentes residuos que podrían estar estabilizando la estructura terciaria de la proteína de la cápside de WMV.

Las cadenas laterales de estos residuos hidrofóbicos probablemente son responsables de la estructura terciaria que adoptan estas hélices en la parte central de la proteína formando entre ellas interacciones hidrofóbicas. Además de este tipo de interacción, cabe destacar posibles interacciones de otra naturaleza entre algunas cadenas laterales que podrían contribuir a la estabilización de la estructura terciaria de la zona globular de la proteína. Por ejemplo, la cadena lateral del residuo R99 ubicado pocos residuos antes de la hélice H1 parece interactuar con la cadena lateral del residuo E219 que se encuentra en el bucle que conecta la hélice 3_{10} G1 con la hélice α H7. Las densidades de las cadenas laterales de los residuos F105 en la hélice H1 e H177 entre las hélices H3 y H4 en nuestro mapa 3D también ponen de manifiesto una posible interacción entre ellos. Por otro lado, las hélices α H4 y H5 están unidas por un lazo de 6 residuos que forma un giro con un ángulo de 45° con la primera. Este giro parece estar estabilizado por la interacción de las cadenas laterales de los residuos R189 y Y195. Además también parece existir interacción entre la cadena lateral del residuo N190 y la cadena principal del residuo Y195 (Figura 4.11, B). Estas posibles interacciones se presentan en zonas peculiares de la estructura terciaria o establecen interacciones entre zonas alejadas en la secuencia de la proteína, de modo que podrían ser relevantes para la constitución de la estructura terciaria de la proteína de la cápside de WMV.

WMV posee un genoma de RNA de cadena sencilla (ssRNA). La zona segmentada del mapa de densidad correspondiente al RNA genómico de WMV reveló que cada CP abarca 5 nucleótidos del genoma. En esta densidad, ubicada aproximadamente entre la hélice α H7 y la hélice α H4, se pueden intuir la cadena de fosfatos y las bases nitrogenadas. Por efecto de la simetría helicoidal y debido a la heterogeneidad en la secuencia de nucleótidos que abarca cada CP a lo largo de todo el virión, las densidades correspondientes a las bases nitrogenadas son una mezcla de las cuatro posibilidades: Guanina, Citosina, Adenina y Uracilo. Teniendo esto en cuenta, se creó un modelo atómico para la densidad del RNA genómico compuesto por 5 ribonucleótidos de uracilo (poli-U) (Figura 4.10 y Figura 4.14).

4.3 Análisis de las interacciones CP - CP

Biológicamente, la polimerización de las proteínas de la cápside está enfocada al empaquetamiento del RNA genómico del virus y así garantizar el éxito de su transmisión. Para determinar qué partes de la CP intervienen en la interacción con otras CPs para formar la cápside del virus se calculó la superficie de contacto de los diferentes dominios de la proteína con las subunidades adyacentes. Para ello, usamos un modelo atómico de la CP de

referencia en contexto con los modelos atómicos de las CPs adyacentes que conforman el virión, y usamos PDBePISA para medir la superficie de contacto de cada parte del modelo atómico de referencia (dominio central, N-terminal y C-terminal) con los modelos adyacentes (Krissinel and Henrick 2007). Nuestro modelo atómico tiene un área superficial accesible a solventes de 16921 \AA^2 , de los cuales aproximadamente 9600 \AA^2 corresponden a la zona central ordenada y alrededor de 7300 \AA^2 a los extremos terminales. La zona central ordenada de la proteína de la cápside sin los extremos terminales (residuos 100-240; figura 4.11, A), presenta una superficie de contacto con las zonas centrales de las proteínas adyacentes de 2098 \AA^2 , en cambio, al incluir los extremos terminales la superficie de contacto total entre CPs asciende a 5142 \AA^2 . Por lo tanto, pese a presentar una menor superficie, los extremos N- y C- terminal contribuyen a la superficie de contacto entre CPs en mayor medida que la zona central, por lo que parecen ser estos extremos terminales los principales artífices de la polimerización.

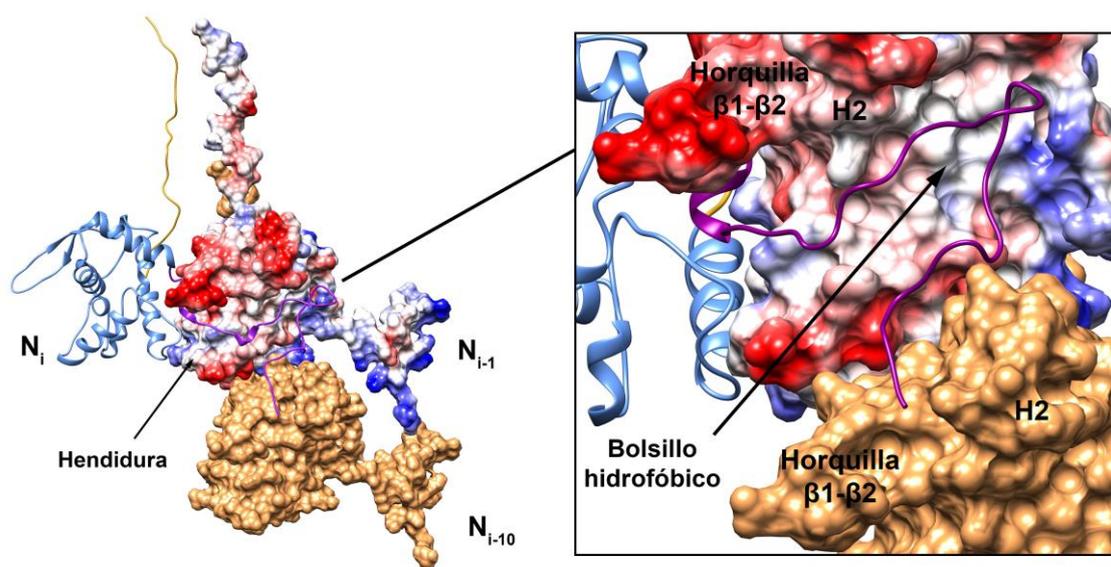


Figura 4.12: Modelos atómicos de la proteína de la cápside de WMV mostrando la interacción del extremo N-terminal de la CP de referencia (N_i) con las subunidades adyacentes (N_{i-1} y N_{i-10}). La subunidad de referencia N_i muestra el modelo atómico coloreado en azul (zona central), amarillo (C-terminal) y morado (N-terminal). La subunidad N_{i-1} muestra la superficie del modelo atómico coloreado en función del potencial electrostático calculado. La subunidad N_{i-10} muestra la superficie del modelo atómico coloreada en marrón. A la derecha del panel, se muestra en detalle la interacción del extremo N-terminal con la hendidura y el bolsillo hidrofóbico de la subunidad adyacente.

A partir de nuestro mapa de densidad 3D y el modelo atómico generado para la CP y el RNA podemos analizar la contribución de las diferentes partes de la CP en la interacción con las proteínas vecinas para conformar el virión. Nuestro mapa 3D y el modelo atómico muestran cómo el extremo N-terminal de una CP (N_i) interacciona con la subunidad adyacente N_{i-1} y

con la subunidad N_{i-10} de la vuelta de hélice anterior. Partiendo desde la región central de la CP (N_i), el brazo N-terminal se posiciona transversalmente al eje de simetría de la cápside y se acomoda en la hendidura formada entre la hélice α H2 junto con la horquilla β y la hélice α H4 de la subunidad N_{i-1} (Figura 4.12). A continuación, el extremo N-terminal forma un bucle con un giro de 90° que se coloca en un pequeño bolsillo hidrofóbico formado entre las hélices α H1, H2 y H4 de la subunidad N_{i-1} (Figura 4.12 y Figura 4.13, A). A partir de aquí, el brazo N-terminal se posiciona longitudinalmente al eje de simetría del virión hasta alcanzar la subunidad N_{i-10} donde se acomoda entre el bucle que une las láminas β y la hélice α H2 de esta subunidad (Figura 4.12 y Figura 4.7, D).

La acomodación del brazo N-terminal viene dada principalmente por la posible interacción entre residuos con cadenas laterales hidrofóbicas como es el caso del residuo L62 (N_i) con la zona cercana al residuo Y116 (N_{i-10}) o el residuo I65 (N_i) con A184 (N_{i-1}). Sin embargo, es entre el bolsillo hidrofóbico de N_{i-1} y el bucle del extremo N-terminal de N_i donde las interacciones presentan mayor protagonismo, ya que la zona del bucle que establece contacto directo con el bolsillo hidrofóbico está poblada con residuos de cadenas laterales hidrofóbicas (P72, V74, L80 y L82). Estos aminoácidos parecen interactuar con los residuos del bolsillo hidrofóbico de N_{i-1} , especialmente con Y109, G124 y F178 (Figura 4.13, A). Además de las interacciones hidrofóbicas, también hay que tener en cuenta la posible relevancia de residuos con cadenas laterales con carga en la interacción del brazo N-terminal (N_i) con las subunidades adyacentes (N_{i-1} y N_{i-10}). Por proximidad, se pueden identificar varios residuos cuyas cadenas laterales podrían establecer puentes salinos o enlaces de hidrógeno, como es el caso de K68 (N_i) con D117 (N_{i-10}), N70 (N_i) con D180 (N_{i-1}), K89 (N_i) con E192 (N_{i-1}) y Q92 (N_i) con D136 (N_{i-1}). Por otro lado el residuo F96 (N_i) podría establecer contacto con R198 (N_{i-1}) formando una interacción de tipo catión- π , lo que podría contribuir a la estabilización de la parte del N-terminal más cercana a la zona central de la CP (Figura 4.13, A).

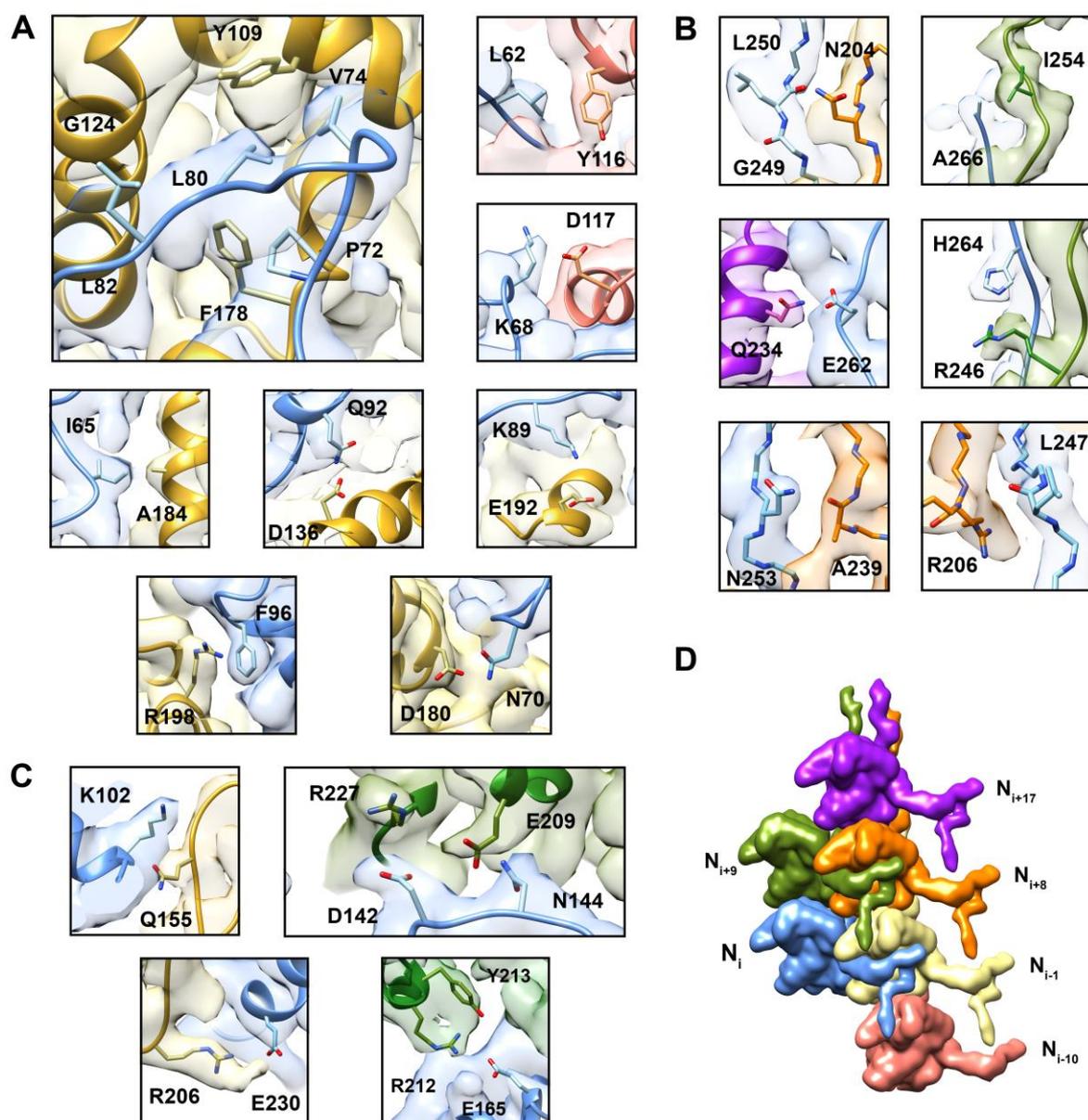


Figura 4.13: Detalle de densidades correspondientes a cadenas laterales, junto con el modelo atómico correspondiente, de determinados residuos de la proteína de la cápside de WMV que podrían establecer interacciones con otros residuos de las CP adyacentes. Los modelos atómicos y los mapas de densidad transparentes están coloreados acorde con el esquema de colores mostrado en **(D)**: En azul, el modelo atómico y densidad transparente correspondiente a la CP de referencia N_i ; en amarillo, el modelo atómico y densidad transparente correspondiente a la CP N_{i-1} ; en rojo, el modelo atómico y densidad transparente correspondiente a la CP N_{i-10} ; en verde, el modelo atómico y densidad transparente correspondiente a la CP N_{i+9} ; en naranja, el modelo atómico y densidad transparente correspondiente a la CP N_{i+8} ; y en morado, el modelo atómico y densidad transparente correspondiente a la CP N_{i+17} . **(A)** Residuos relevantes para la interacción del extremo N-terminal con las CP adyacentes. **(B)** Residuos relevantes para la interacción del extremo C-terminal con las CP adyacentes. **(C)** Residuos ubicados en la zona central de la proteína que parecen interactuar con residuos de las proteínas adyacentes.

El extremo C-terminal, aunque aparentemente de una manera más sutil en comparación con el extremo N-terminal, también parece desempeñar un papel en la polimerización de la cápside. Tomando como referencia la subunidad N_i , el extremo C-terminal nace desde la hélice α H7 y discurre longitudinalmente al eje de simetría de la cápside colocándose entre las hélices α H7 de las subunidades de la vuelta siguiente N_{i+8} y N_{i+9} . Desde este punto continúa longitudinalmente cerca de la parte inicial del extremo C-terminal de la subunidad N_{i+9} hasta alcanzar la hélice α H7 de la subunidad N_{i+17} (dos vueltas de hélice adelante) (Figura 4.8 y Figura 4.13, D). Aproximadamente en esta zona el brazo C-terminal de N_i se aproxima al brazo C-terminal de N_{i+9} , y a partir de aquí, la densidad para el extremo C-terminal de N_i deja de ser fácilmente identificable. Se pueden identificar posibles interacciones de diferente naturaleza entre las cadenas laterales de los residuos A266 (N_i) e I254 (N_{i+9}), H264 (N_i) y R246 (N_{i+9}), E262 (N_i) y Q234 (N_{i+17}). También parece existir interacción entre el grupo carbonilo del residuo L250 (N_i) y la cadena lateral de N204 (N_{i+8}), la cadena lateral del residuo N253 (N_i) y la cadena principal de A239 (N_{i+8}), y entre las cadenas principales de los residuos L247 (N_i) y R206 (N_{i+8}). En general, son estas posibles interacciones las que determinarían la disposición del extremo C-terminal y su interacción con las subunidades vecinas (Figura 4.13, B).

Si bien los extremos N-terminal y C-terminal parecen ser los responsables de la polimerización y el anclaje entre subunidades, las partes centrales de las CPs también están en contacto entre ellas. En estas zonas centrales de las CPs identificamos residuos que por proximidad podrían establecer contactos con residuos de la región central de la CP adyacente, lo que contribuiría a una mayor estabilización entre proteínas de la cápside. Interacciones de naturaleza electrostática podrían ocurrir entre las cadenas laterales de los residuos K102 (N_i) y Q155 (N_{i-1}), D142 (N_i) y R227 (N_{i+9}), N144 (N_i) y E209 (N_{i+9}), E230 (N_i) y R206 (N_{i-1}), y entre E165 (N_i) y R212 (N_{i+9}), que a su vez parece establecer contacto con Y213 (N_{i+9}) (Figura 4.13, C).

Las superficies de contacto entre CPs vienen dadas en mayor medida por los extremos N-y C-terminal. En concordancia, las posibles interacciones identificadas en nuestro modelo atómico involucran residuos ubicados principalmente en los extremos terminales de la proteína. Estos extremos terminales que no presentan una estructura secundaria definida sino quizás una mayor flexibilidad, parecen ser los principales puntos de anclaje entre CPs para la formación de la cápside, lo que explicaría, al igual que en miembros del género *Potexvirus* la característica flexibilidad de estos viriones.

4.4 Análisis de las interacciones CP - RNA

La función principal de la proteína de la cápside de WMV es proteger el RNA genómico y garantizar su transmisión con éxito. Para ello, además de una correcta polimerización entre las proteínas de la cápside, estas deben ser capaces de interactuar con el RNA del virus y empaquetarlo de manera ordenada para formar la cápside. Como vemos en la Figura 4.10, A, en nuestro modelo atómico de la CP de WMV se identifica la región de unión al RNA en la zona globular central, formando un bolsillo de unión acotado por las hélices α H4, H6 y H7, el lazo de unión entre la hélice H2 y la lámina β E1 y el lazo de unión entre la hélice 3_{10} G1 y la hélice α H7 (Figura 4.10, B). Cada subunidad de la cápside de WMV abarca 5 ribonucleótidos del RNA genómico, pese a que las densidades de las bases nitrogenadas son un promedio de señales heterogéneas y podría llevar a errores de interpretación, se puede determinar con claridad como una de las bases nitrogenadas se posiciona hacia el interior del bolsillo de unión a RNA entre los residuos R172 y K236 (nucleótido marcado como U4: Figura 4.14), mientras que se puede intuir cómo las otras cuatro se colocan orientadas hacia el interior del virión (Figura 4.14).

Como vemos en el modelo atómico generado, el bolsillo de interacción de la CP presenta varios residuos que, por su orientación y proximidad, parecen tener especial relevancia en la interacción con el RNA (Figura 4.14 y Figura 4.15, A). El aminoácido S140 en este bolsillo parece establecer contacto con el fosfato del ribonucleótido U4. En la hélice α H4 se puede identificar una densidad orientada hacia el bolsillo de unión que corresponde con la cadena lateral del residuo R172. Esta densidad está muy próxima al RNA, lo que sugiere que podría estar estableciendo contactos con el fosfato del ribonucleótido U3 y con la base nitrogenada del ribonucleótido U4 orientada hacia el interior de bolsillo (Figura 4.14, C y Figura 4.15, A), presentándose como un residuo bastante relevante en la interacción. Por otro lado, el residuo D216 está presente en el bolsillo de unión, pero no está claro que presente algún contacto con el RNA, más bien parece estar estabilizando el lazo en el que se encuentra el residuo G138 que parece establecer contacto con la pentosa del ribonucleótido U5. También formando parte del bolsillo de interacción con el RNA está el residuo Y199 que, por proximidad, parece estabilizar al residuo R203, que a su vez parece establecer contacto con el fosfato del ribonucleótido U5 (Figura 4.15, A). Sin embargo, Y199 también se ubica muy cerca de la base nitrogenada del ribonucleótido U4, por lo que podría estar estabilizándola dentro del bolsillo. Por último, el residuo K236 se ubica en la hélice α H7 muy próximo a la base nitrogenada del ribonucleótido U4, por lo que también podría contribuir a su estabilización en el interior del bolsillo (Figura 4.14, C).

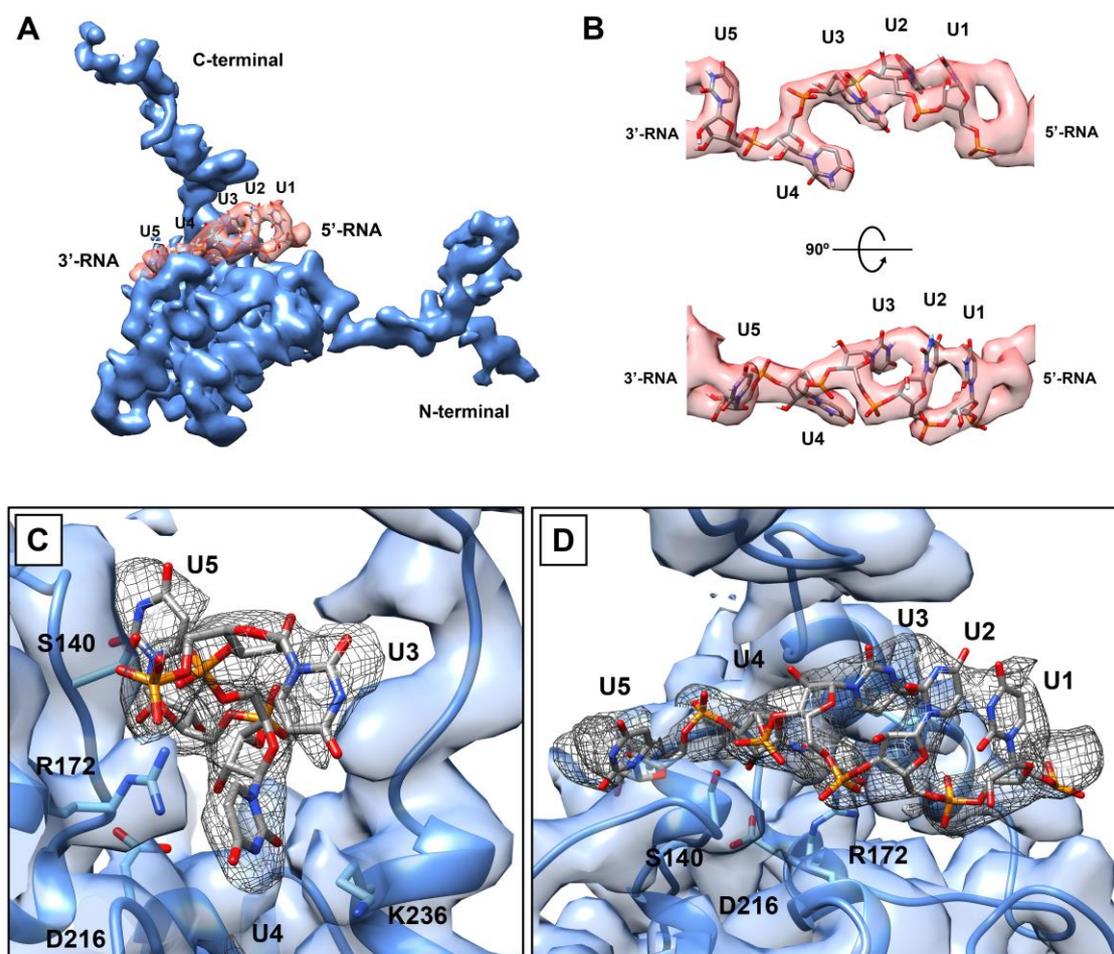


Figura 4.14: Unión del RNA genómico a la proteína de la cápside de WMV. **(A)** Modelo atómico de 5 ribonucleótidos de uracilo, dentro de su densidad correspondiente (rojo transparente) en el mapa 3D, en su unión con la CP de WMV (densidad azul). **(B)** Detalle de un fragmento de la densidad correspondiente al RNA genómico en el mapa de densidad 3D. En el interior de la densidad, se encuentra el modelo atómico generado para el fragmento de RNA abarcado por cada CP. **(C)** y **(D)** Detalles lateral y superior respectivamente del bolsillo de unión a RNA de la CP. El modelo atómico para la proteína de la cápside está coloreado en azul, dentro de la densidad para la proteína (azul semitransparente). El modelo atómico para el RNA genómico (para facilitar la visualización, se muestran 3 ribonucleótidos en **C** y 5 ribonucleótidos en **D**) se encuentra dentro de la densidad para el RNA mostrada en forma de malla transparente. En el modelo atómico de la CP se muestran residuos relevantes en el bolsillo de unión a RNA.

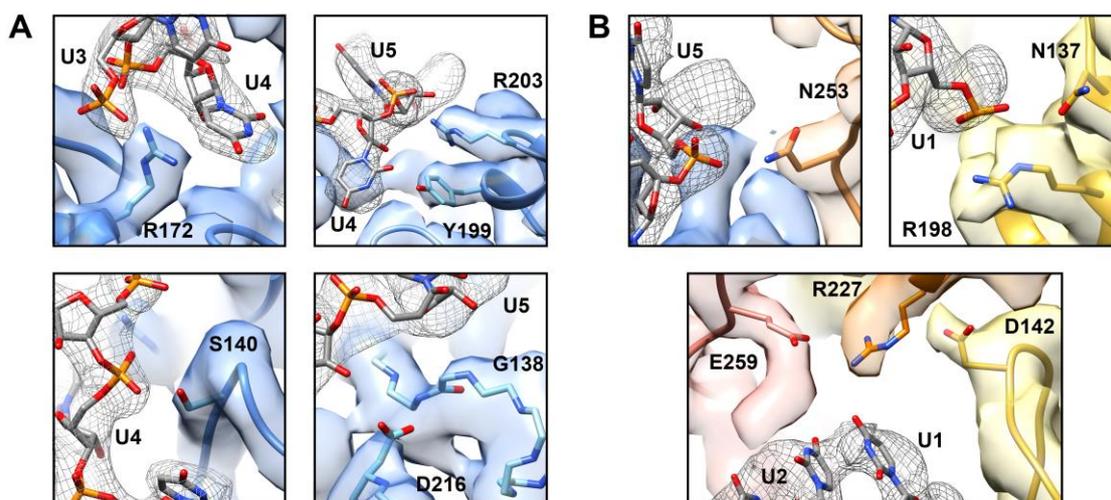


Figura 4.15: Detalle de residuos posiblemente relacionados con la interacción de la CP con el RNA. **(A)** Detalle de residuos pertenecientes a la CP de referencia (N_i) ubicados en el bolsillo de unión a RNA. El modelo atómico de la CP está coloreado en azul, dentro de la densidad correspondiente a la proteína (azul semitransparente). El modelo atómico correspondiente al fragmento de RNA se encuentra en el interior de la densidad para el RNA mostrada en forma de malla semitransparente. **(B)** Detalle de residuos de CP adyacentes que podrían estar en contacto con el fragmento de RNA de 5 ribonucleótidos abarcados por la CP de referencia (N_i). En marrón, el modelo atómico y densidad transparente correspondiente a la CP N_{i-8} ; en amarillo, el modelo atómico y densidad semitransparente correspondiente a la CP N_{i-1} ; en rosa, el modelo atómico y densidad transparente correspondiente a la CP N_{i-9} ; y en naranja, el modelo atómico y densidad transparente correspondiente a la CP N_{i+8} .

Como se ha mencionado anteriormente, cada subunidad de la cápside del virus abarca un fragmento de 5 ribonucleótidos del genoma del virus, sin embargo alguno de los contactos con el fragmento de RNA de una subunidad dada (N_i) están establecidos por residuos pertenecientes a subunidades adyacentes o ubicadas en una vuelta de hélice distinta ($N_{i+/-n}$) (Figura 4.15, B). En el mapa de densidad 3D y en el modelo atómico generado se pueden identificar otros residuos que parecen establecer contactos adicionales con el RNA o residuos relevantes fuera del bolsillo de interacción. Es el caso del residuo N137 de N_{i-1} que parece estabilizar al aminoácido R198 de N_{i-1} que a su vez parece establecer contacto con el fosfato del ribonucleótido U1 perteneciente al sección de RNA genómico unido a la CP N_i . También el residuo R227 de la subunidad N_{i+8} parece establecer contacto con la base nitrogenada del ribonucleótido U1 ubicado en la subunidad N_i . A su vez, este residuo podría estar estabilizado por los residuos E259 de la subunidad N_{i-9} y D142 de la subunidad N_{i-1} . También el residuo N253 en el C-terminal de la subunidad N_{i-8} parece establecer contacto con el fosfato del ribonucleótido U5 de la subunidad N_i .

4.5 Análisis de la similitud estructural entre las proteínas de la cápside de los virus filamentosos flexibles

Los viriones de las distintas familias de virus filamentosos flexibles presentan características parecidas y además, a baja resolución, presentan una estructura semejante (Kendall et al. 2008). Debido a esto se cree que las proteínas de la cápside de las distintas familias de virus filamentosos flexibles podrían presentar una estructura similar a pesar de la ausencia de homología de secuencia significativa entre ellas.

La arquitectura general que apreciamos en nuestro mapa 3D a alta resolución del virión de WMV (familia *Potyviridae*) es prácticamente idéntica a la arquitectura de los viriones de PepMV (*Virus del mosaico del pepino*) y BaMV (*Virus del mosaico del bambú*), pertenecientes a la familia *Alphaflexiviridae*, revelados en los mapas de densidad 3D a alta resolución recientemente publicados (Agirrezabala et al. 2015, DiMaio et al. 2015). El diámetro del virión de WMV es de 130 Å aproximadamente, al igual que el diámetro del virión de PepMV y de BaMV. Además la simetría helicoidal de estos virus presenta unos valores muy similares: 8,8 unidades asimétricas por vuelta de hélice y 35,11 Å de paso de hélice para WMV; 8,76 unidades asimétricas por vuelta de hélice y 34,6 Å de paso de hélice para PepMV (Agirrezabala et al. 2015); 8,8 unidades asimétricas por vuelta de hélice y 35,2 Å de paso de hélice para BaMV (DiMaio et al. 2015). Tanto en los miembros caracterizados del género *Potexvirus* (fam. *Alphaflexiviridae*), incluido PapMV (*Virus del mosaico de la Papaya*) (Yang et al. 2012), como en WMV del género *Potyvirus* (fam. *Potyviridae*), la proteína de la cápside presenta una zona central rica en hélices α con dos extremos terminales ubicados en zonas equivalentes en el virión. Igualmente, el extremo N-terminal de la CP de PepMV y BaMV también es relevante en la polimerización para formar el virión, y el modo de interacción de este extremo con la proteína de la cápside adyacente es muy similar al observado en WMV (Figura 4.16).

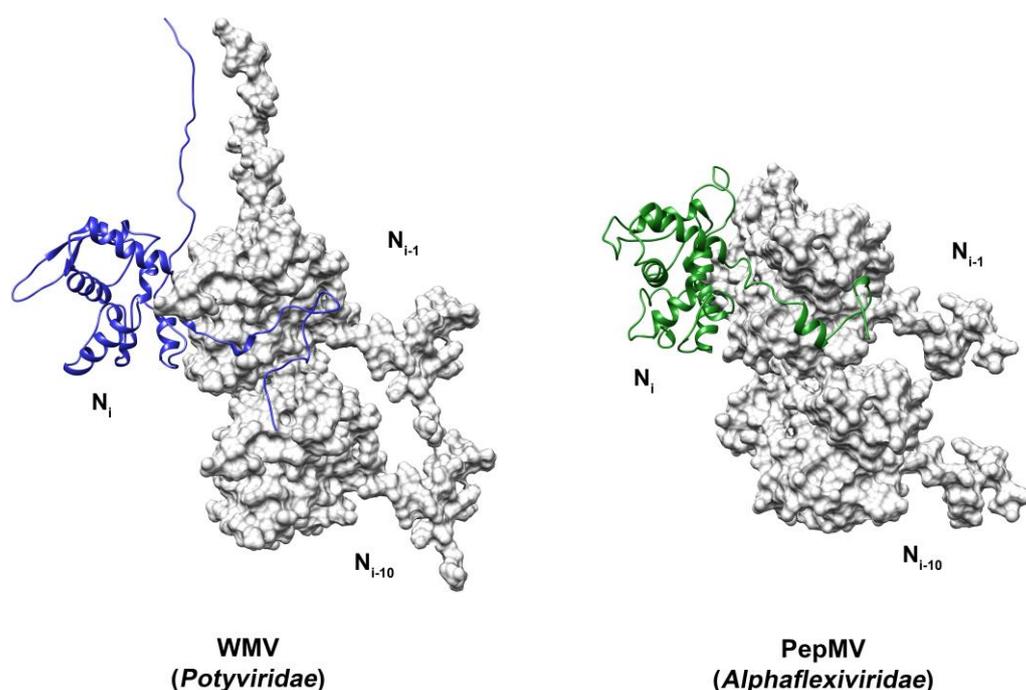


Figura 4.16: Disposición del extremo N-terminal en proteínas de la cápside de virus filamentosos flexibles. En ambos casos el modelo atómico de la subunidad de referencia (N_i) está coloreado (azul para WMV y verde para PepMV), y las subunidades adyacentes (N_{i-1} y N_{i-10}) a esta están representadas en forma de densidad gris creada a partir del modelo atómico. Código PDB para el modelo atómico de PepMV: 5fn1 (Agirrezabala et al. 2015).

Al identificar y caracterizar la zona central de la CP de WMV hemos podido observar que presenta muchas similitudes en su estructura secundaria y terciaria con la zona central de la CP de *Potexvirus*. Por ello, decidimos utilizar el servidor TM-align (Zhang and Skolnick 2005) para medir esta similitud. La comparación de dos estructuras con este servidor queda reflejada en un valor ("TM-score") comprendido entre 0 y 1. Un resultado inferior a 0,2 indicaría que el plegamiento de las dos proteínas comparadas no estaría relacionado, mientras que valores superiores a 0,5 señalarían un plegamiento común. Al comparar el modelo atómico completo generado para la CP de WMV a partir de nuestro mapa 3D, con los modelos atómicos disponibles para las CP del género *Potexvirus* (PepMV, código pdb: 5fn1; BaMV, código pdb: 5a2t; y PapMV, código pdb: 4dox) usando TM-align, obtenemos en todos los casos valores superiores a 0,5 (Tabla 4.2). Por otro lado, al comparar únicamente la parte central de la proteína de la cápside de estos virus, que es la parte que mayor similitud presenta a simple vista, prescindiendo de los extremos N- y C- terminales, el valor obtenido de TM-align es en todos los casos superior a 0,6 (Tabla 4.3 y Figura 4.17). Estos valores indican que las proteínas de la cápside de todos los virus filamentosos flexibles estructuralmente descritos presentan un plegamiento común, especialmente en la zona central ordenada.

	WMV (60-266)	PepMV (21-237)	BaMV (39-242)	PapMV (11-174)
WMV (60-266)	1	0,63	0,60	0,51
PepMV (21-237)	0,60	1	0,75	0,64
BaMV (39-242)	0,60	0,79	1	0,63
PapMV (11-174)	0,61	0,84	0,77	1

	WMV (101-239)	PepMV (54-207)	BaMV (63-214)	PapMV (35-174)
WMV (101-239)	1	0,72	0,65	0,66
PepMV (54-207)	0,66	1	0,79	0,84
BaMV (63-214)	0,60	0,79	1	0,78
PapMV (35-174)	0,66	0,92	0,84	1

Tablas 4.2 (arriba) y 4.3 (abajo): Valores obtenidos con el servidor “TM-align” para distintas comparaciones de modelos atómicos de proteínas de la cápside entre WMV y varias especies de *Potexvirus* (códigos pdb: PepMV, 5fn1; BaMV, 5a2t; PapMV, 4dox). En negrita figura el modelo atómico de referencia para cada comparación. En el caso de la tabla 4.2 (arriba) se compara la totalidad del modelo atómico disponible, sin embargo, en la tabla 4.3 (abajo) se compara únicamente la parte central de la proteína de la cápside, prescindiendo de los extremos N-terminal y C-terminal, ya que es la zona que más similitud presenta. En el servidor “TM-align”, valores por encima de 0,50 indican que las dos estructuras a comparar presentan un plegamiento común.

Conclusiones similares obtenemos usando MATRAS, otro servidor de alineamiento 3D (Kawabata 2003). Tanto por examinación visual de la estructura de la zona central de estas proteínas, como por alineamiento 3D (Figura 4.18 y Anexo I.a), se identifican elementos de estructura secundaria equivalentes entre ellas ubicados en posiciones espaciales muy similares constituyendo un plegamiento común para la zona central de estas proteínas.

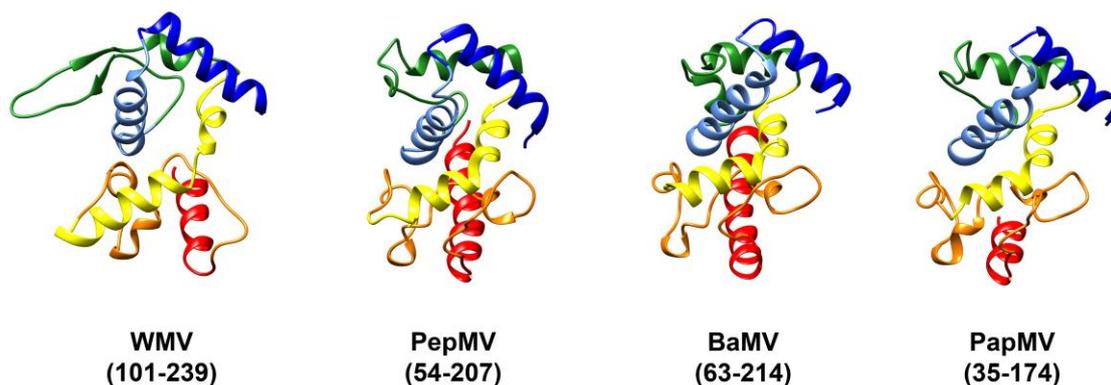


Figura 4.18: Comparación de la zona central de la proteína de la cápside de los modelos atómicos disponibles para virus filamentosos flexibles. El código de colores resalta los diferentes elementos de estructura secundaria equivalentes entre las diferentes estructuras (códigos pdb: PepMV, 5fn1; BaMV, 5a2t; PapMV, 4dox).

4.5.1 Mecanismo de unión a RNA común en virus filamentosos flexibles

El plegamiento común que observamos entre proteínas de la cápside de WMV y miembros del género Potexvirus también incluye al modo de unión de estas con el RNA genómico del virus. En todos estos virus la proteína de la cápside abarca 5 ribonucleótidos del genoma y el sitio de unión a este está ubicado en la misma posición en la estructura terciaria. En todos los casos se encuentra entre la hélice α más próxima al extremo C-terminal (hélice α en rojo en la Figura 4.18) y el resto de la zona central de la proteína. Las densidades correspondientes al RNA genómico en los mapas 3D de WMV y PepMV adoptan una disposición casi idéntica, con uno de los cinco ribonucleótidos orientando la base nitrogenada hacia el interior del bolsillo de unión (Figura 4.19, A).

Al comparar más en detalle la composición de las zonas de unión al RNA en los modelos atómicos de WMV y PepMV podemos identificar que 4 de los residuos que parecen ser relevantes para la interacción de la proteína con el RNA están conservados entre estos dos virus, ya que se encuentran ubicados en la misma posición en la estructura de las proteínas de la cápside de ambos. Estos residuos son: S140, R172, D216 y K236 en WMV; y S92, R124, D163 y K196 en PepMV (Figura 4.19, B).

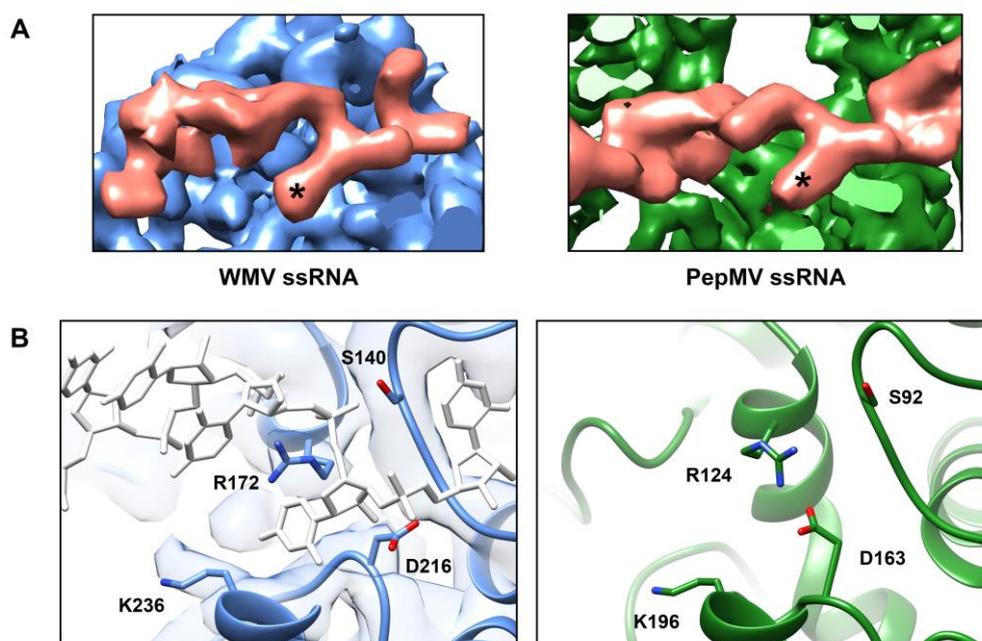


Figura 4.19: Interacción RNA genómico (ssRNA) con la proteína de la cápside en virus filamentosos flexibles. **(A)** Densidades atribuibles al fragmento de RNA genómico (5 ribonucleótidos) unido a la CP. En ambos paneles la densidad para el RNA está coloreada en rojo claro, mientras que la densidad para la proteína está coloreada en azul (WMV) y en verde (PepMV). En ambos casos un asterisco indica la densidad del ribonucleótido orientado hacia el interior del bolsillo de interacción. Código para mapa de densidad 3D de PepMV: EMD-3236 (Agirrezabala et al. 2015). **(B)** Detalle de residuos ubicados en el bolsillo de interacción con el RNA que se encuentran conservados entre WMV (familia *Potyviridae*) y PepMV (familia *Alphaflexiviridae*). El modelo atómico de la CP de WMV se presenta en azul, dentro de la densidad correspondiente a la proteína (azul transparente), mientras que el modelo atómico para la CP de PepMV se presenta en verde (código pdb: 5fn1). El modelo atómico para el RNA genómico se presenta en blanco.

Al realizar un alineamiento múltiple de secuencias de referencia de proteínas de la cápside para distintas especies de cada familia de virus filamentosos flexibles obtuvimos la secuencia consenso correspondiente para cada familia. Identificamos en esta secuencia que tres de los residuos conservados en el sitio de unión a RNA entre WMV y de PepMV estaban también conservados en sus correspondientes familias (*Potyviridae* y *Alphaflexiviridae* respectivamente). Además, estos tres residuos también aparecen conservados en las familias *Betaflexiviridae* y *Closteroviridae*, ubicados en zonas de la secuencia equivalentes al sitio de unión a RNA. Asimismo, la separación entre estos tres residuos es muy similar en las 4 familias, a excepción de algunos géneros de la familia *Betaflexiviridae* (Figura 4.20). La conservación de estos tres residuos sugiere un modo de interacción con el RNA común entre los virus filamentosos flexibles, y por ende una estructura similar de sus nucleoproteínas.

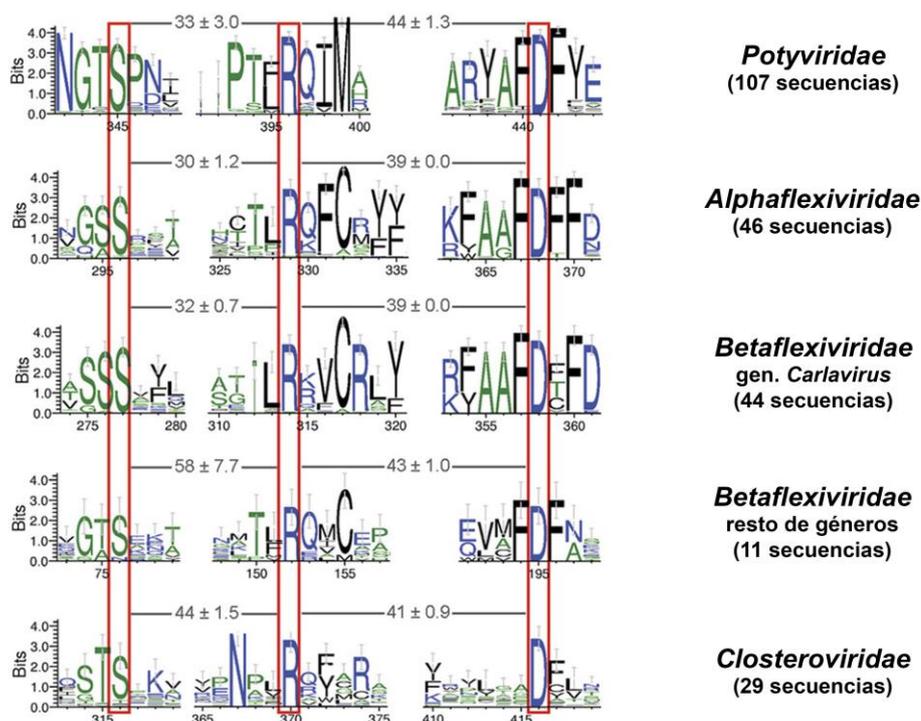


Figura 4.20: Secuencia consenso para zonas concretas de las CPs de las diferentes familias de virus filamentosos flexibles. La distancia entre los aminoácidos conservados (recuadrados en rojo) en la secuencia consenso está medida por número de residuos (media y desviación estándar). A la derecha, se indica el número de secuencias comparadas para obtener la secuencia consenso de cada familia. El código de color en el esquema de la secuencia consenso es: azul para residuos hidrofílicos, verde para residuos neutrales y negro para residuos hidrofóbicos.

4.6 Análisis de la similitud estructural entre nucleoproteínas de virus de ssRNA

Parece existir un plegamiento común entre las proteínas de la cápside de las diferentes familias de virus filamentosos flexibles. En sí mismas, estas proteínas son nucleoproteínas (NP), ya que tienen la capacidad de interactuar con ácidos nucleicos, siendo esta su función principal. Muchos virus disponen de nucleoproteínas que empaquetan o protegen su genoma, ya constituyan estas la cápside del virus (proteínas de la cápside), como es el caso de los virus filamentosos, o se encuentren incluidas dentro de otras estructuras del virus.

Dada la existencia de un plegamiento común entre las CP de los miembros descritos de virus filamentosos flexibles, decidimos analizar si este mismo plegamiento está presente en otras familias de virus, usando para ello los servidores MATRAS (Kawabata 2003) y Dali (Holm and Rosenström 2010). Con estos servidores comparamos el modelo atómico 3D de una

proteína determinada con diferentes modelos atómicos descritos en diferentes bases de datos. En concreto, comparamos la zona central (101-239) de la CP de WMV con los modelos atómicos depositados en el servidor “Protein Data Bank” (PDB). Con ambos métodos los candidatos obtenidos que presentan mayor similitud con la zona central de WMV fueron los miembros descritos del género *Potexvirus* (PepMV, BaMV y PapMV), pero también aparecen otros candidatos con cierto grado de similitud. Estos otros candidatos son nucleoproteínas pertenecientes a la familia *Orthomyxoviridae* y al orden *Bunyavirales*, incluyendo las familias *Phenuiviridae* (género *Phlebovirus*), *Peribunyaviridae* y *Tospoviridae*; todos ellos virus con genoma (-)ssRNA.

Inspeccionando los modelos atómicos 3D de las nucleoproteínas candidatas obtenidas en la búsqueda, observamos que presentan un plegamiento general con bastantes diferencias entre sí; sin embargo, reconocemos la similitud con la zona central de la CP de WMV en una zona concreta de estas proteínas al igual que ocurría entre virus filamentosos flexibles (Figura 4.21). En la nucleoproteína del virus RVFV (“*Rift Valley Fever Virus*”), miembro del género *Phlebovirus* (familia *Phenuiviridae*), esta zona similar a la CP de WMV se encuentra entre los residuos 34 y 181 aproximadamente; para el virus TSWT (*Virus del bronceado del tomate*, “*Tomato spotted wilt tospovirus*”), de la familia *Tospoviridae*, esta zona está comprendida entre los residuos 33 y 174; en el virus La Crosse, miembro de la familia *Peribunyaviridae*, la zona similar se encuentra entre los residuos 17 y 171; en cambio en la nucleoproteína del virus de la gripe (Influenza A), perteneciente a la familia *Orthomyxoviridae*, la zona similar se ubica entre los residuos 21 y 202 aproximadamente.

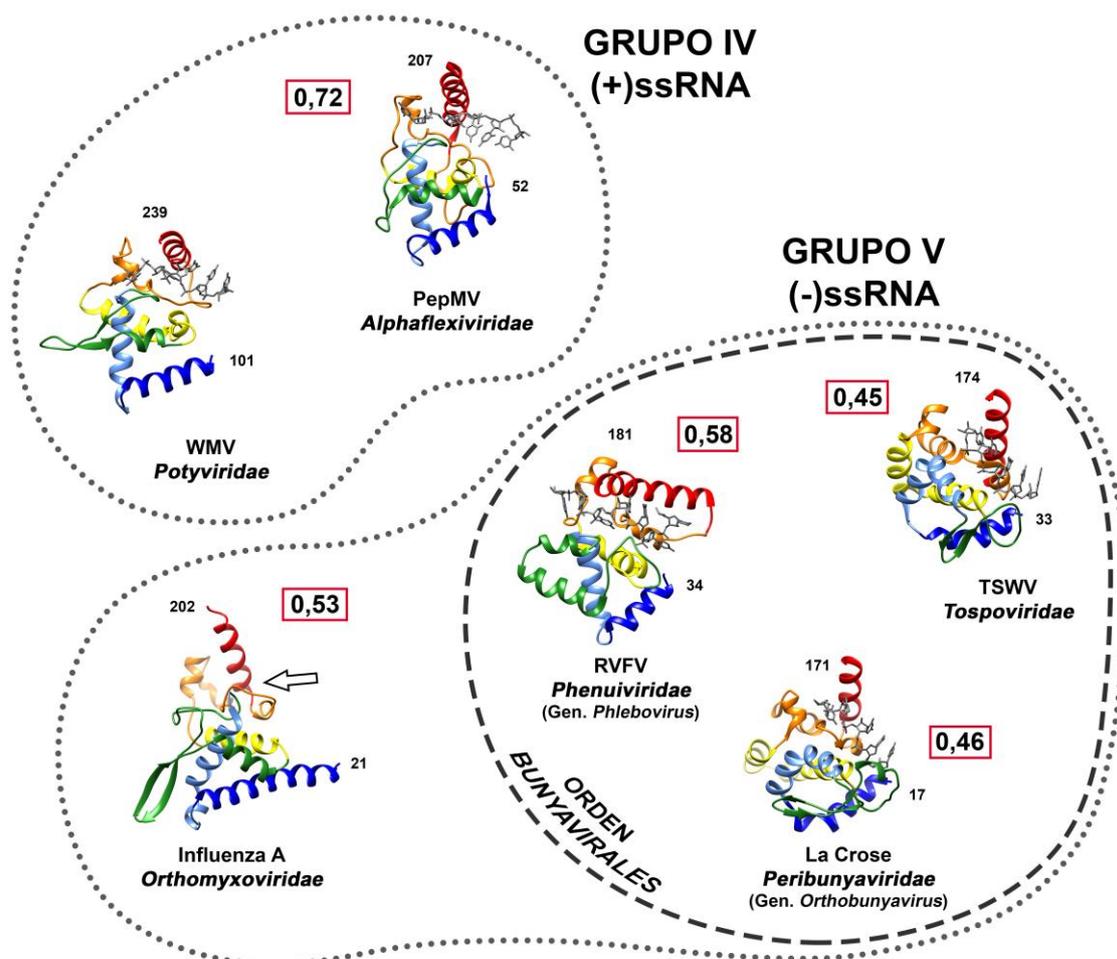


Figura 4.21: Comparación de nucleoproteínas de algunos virus ssRNA con la CP de WMV. El código de colores señala elementos de estructura secundaria comunes entre los distintos modelos atómicos. En todos ellos se muestra un fragmento del RNA genómico del virus unido a la nucleoproteína, a excepción de Influenza A, en el que el presunto sitio de unión a RNA está indicado con una flecha (Ye, Krug, and Tao 2006). En un recuadro rojo se indica valor de similitud estructural obtenido con TM-align en la comparación de cada NP con la CP de WMV (Códigos pdb: PepMV, 5fn1; Influenza A,

Para analizar la similitud de todas estas nucleoproteínas con la zona central de la CP de WMV realizamos un alineamiento 3D de estas proteínas empleando los servidores TM-align y MATRAS (Figura 4.22 y Anexo I.b). De nuevo usamos el modelo atómico generado para la zona central de la proteína de la cápside de WMV y lo comparamos con la región parecida en cada una de las nucleoproteínas de los virus candidatos. El alineamiento resultó en valores de “TM-score” por encima de 0,5 al comparar la CP de WMV con la NP de RVFV y la NP de Influenza A. En cambio este valor es ligeramente menor (por encima de 0,4) al comparar la CP de WMV con la NP de TSWV y la NP del virus La Crosse. Estos valores indican la existencia de un plegamiento común entre las regiones de unión a RNA de nucleoproteínas de miembros de distintas familias de virus con genoma (+)ssRNA y (-)ssRNA.

CAPITULO 5
DISCUSIÓN

Las familias que forman el grupo de virus filamentosos flexibles constituyen uno de los mayores conjuntos de patógenos de plantas conocidos, y su impacto en la agricultura es muy relevante a nivel mundial (Revers and García 2015a). Aumentar el conocimiento a nivel estructural de estos virus permitiría identificar posibles alternativas para evitar sus infecciones o paliar el efecto de estas. Además, avances en la comprensión estructural de estos virus, también pueden contribuir a mejorar el desarrollo de nano-partículas virales basadas en cápsides de virus filamentosos, de especial relevancia en el campo de los nano-biomateriales (Narayanan and Han 2017).

Pese a su gran relevancia, poco se conocía a nivel estructural de estos virus, y la estructura a alta resolución del virión de algún miembro de este grupo ha sido imposible de determinar hasta el desarrollo de la crio-microscopía electrónica de alta resolución. Otros virus filamentosos como el *Virus del mosaico del tabaco* (TMV), cuyos viriones no presentan flexibilidad, permitieron su estudio estructural por cristalografía de rayos X y la estructura a alta resolución de estos viriones es conocida desde hace tiempo (Bloomer et al. 1978, Stubbs, Warren, and Holmes 1977). Aunque algunos trabajos muestran la estructura helicoidal de los viriones de diferentes familias de virus filamentosos flexibles (Kendall et al. 2008), sólo varias publicaciones recientes sobre miembros de la familia *Alphaflexiviridae* han ilustrado por primera vez tanto la estructura a alta resolución de la proteína de la cápside individual, como la estructura a alta resolución del virión de varias especies de esta familia (Yang et al. 2012, Agirrezabala et al. 2015, DiMaio et al. 2015). Estos trabajos han mostrado las diferencias estructurales existentes entre los virus filamentosos flexibles y los rígidos como el TMV, porque aunque ambos tipos de virus presentan viriones filamentosos con una organización helicoidal, las proteínas que los conforman son completamente distintas.

La organización helicoidal del virión de los virus filamentosos flexibles fue identificada a mediados del s. XX, y posteriormente se comprobó que los parámetros de simetría helicoidal que definen la estructura del virión son similares en varias familias de este grupo de virus (*Potyviridae* y *Alphaflexiviridae*), apuntando a un plegamiento común para las proteínas de la cápside (Varma et al. 1968, Kendall et al. 2008). Sin embargo, hasta ahora estas hipótesis no se podían esclarecer por la ausencia de modelos atómicos de la CP para miembros de distintas familias de virus filamentosos flexibles. La determinación estructural de un miembro de otra familia de virus filamentosos flexibles amplía el conocimiento sobre este grupo de virus y permite encontrar puntos en común entre ellos que pueden ser extrapolados al resto de familias, identificando así características estructurales que expliquen las similitudes existentes y definan a este grupo de virus.

En este trabajo se ha realizado la caracterización estructural a alta resolución del virión de WMV por crio-ME. Para ello, se ha utilizado el procesamiento helicoidal usando miles de segmentos de viriones de WMV para finalmente obtener un mapa de densidad 3D a una resolución global de 4,0 Å. Este mapa de densidad nos ha permitido identificar y caracterizar la proteína de la cápside de WMV y su mecanismo de unión al RNA del virus, así como el mecanismo de ensamblaje con otras proteínas de la cápside para constituir el virión. Estos resultados muestran por primera vez la estructura a alta resolución de un virión de un virus de la familia *Potyviridae*, y revelan la enorme similitud existente con los miembros caracterizados de la familia *Alphaflexiviridae*.

5.1 Simetría helicoidal en virus filamentosos flexibles

Gracias a las características estructurales del virión de WMV y el resto de virus filamentosos se puede usar el procesamiento helicoidal para procesar los segmentos (partículas) de los viriones de las micrografías de crio-microscopía electrónica. Es indispensable conocer los valores correctos que definen la simetría de la cápside del virus, ya que si se utilizan valores incorrectos durante el procesamiento daría lugar a mapas de densidad 3D artefactuales que no reflejarían la realidad de los datos.

Los valores de simetría helicoidal para el virión de un miembro de la familia *Potyviridae*, el *Virus del mosaico de la soja* (SMV), fueron determinados con anterioridad con datos a baja resolución (8,8 subunidades/vuelta; 33 Å de paso de hélice) (Kendall et al. 2008), y son muy parecidos a los valores calculados para distintos miembros de la familia *Alphaflexiviridae* usando diferentes aproximaciones (aproximadamente 8,76 subunidades/vuelta; 34,60 Å de paso de hélice; -41.1° de ángulo azimutal) (Agirrezabala et al. 2015, DiMaio et al. 2015, Kendall et al. 2008). Igualmente, también son muy similares a los valores de simetría descritos para *Closteroviridae* (aproximadamente 8 subunidades/vuelta, 33 Å de paso de hélice) (Tollin et al. 1992), poniendo de manifiesto la similitud en la disposición de las proteínas de la cápside en los viriones de virus filamentosos flexibles. Los valores de simetría para WMV calculados previamente en nuestro laboratorio (8,8 subunidades/vuelta; 35,60 Å de paso de hélice), como era de esperar, resultaron ser muy parecidos a los de SMV y los del resto de virus filamentosos flexibles caracterizados, y nos sirvieron como valores iniciales para el refinamiento de los datos de alta resolución de WMV. Durante el propio procesamiento

podimos mejorar estos valores de simetría iniciales para ajustarnos más a los valores reales que definen la simetría de la cápside (8,81 subunidades/vuelta; 35,12 Å de paso de hélice; -40,87° de ángulo azimutal) y así disminuir los posibles errores provocados por imponer una simetría incorrecta. Con estos valores, hemos podido obtener un mapa de densidad 3D con resoluciones locales próximas a 3,5 Å, en el que se definen claramente los elementos de estructura secundaria de la proteína de la cápside de WMV y se identifican densidades correspondientes a cadenas laterales de algunos residuos, de modo que la calidad observada en nuestro mapa de densidad nos ayuda a validar el procesamiento y los valores de simetría empleados.

Dado que el virión de WMV presenta flexibilidad, sospechamos que quizás esta característica podría afectar de algún modo a la simetría del virión y que esta fuera variable a lo largo de la estructura del mismo; sin embargo, hemos visto que la mayoría (93%) de los segmentos de virión aislados de las micrografías presentan los mismos valores de simetría helicoidal, por lo que la simetría de WMV parece ser bastante uniforme a lo largo del virión.

5.2 Estructura de la proteína de la cápside y ensamblaje del virión de WMV

La predicción de desorden para la proteína de la cápside de WMV muestra una zona central ordenada, principalmente compuesta por hélices α , rodeada por extremos con un grado de desorden mayor (Figura 4.9). Esta coincide con las predicciones de desorden realizadas para diferentes especies de las cuatro familias que componen el grupo de virus filamentosos flexibles (Solovyev and Makarov 2016). En todas estas predicciones, la zona central ordenada abarca alrededor de 150 residuos, y las zonas desordenadas de los extremos varían notablemente incluso entre géneros de la misma familia, especialmente el extremo N-terminal. Con la caracterización estructural a alta resolución de varios miembros del género *Potexvirus* se pudo comprobar que la proteína de la cápside de al menos algunos miembros de la familia *Alphaflexiviridae* presenta esta organización. A partir de esta caracterización estructural se pudo determinar que el extremo N-terminal de la proteína de la cápside de estos virus estaba formado por dos dominios: el dominio IA, que es la parte del N-terminal que interacciona con la CP adyacente, y cuya densidad es visible en los mapas de crio-ME; y el dominio NID que corresponde con una región desordenada en el exterior del virión, que

por su naturaleza flexible no es visible en los mapas de crio-ME. En la caracterización estructural del virión de WMV realizada en este trabajo hemos determinado que la proteína de la cápside de al menos un miembro del género *Potyvirus* (fam. *Potyviridae*) también presenta la misma organización que indican las predicciones. Como se aprecia en nuestro modelo atómico, esta presenta una zona central ordenada rica en hélices α flanqueada por dos extremos terminales, al igual que los miembros descritos del género *Potexvirus*. El extremo N-terminal de la proteína de la cápside de WMV también presenta la región de interacción con la CP adyacente (dominio IA), pero sin embargo, el dominio NID presenta una longitud diferente, siendo alrededor de 60 residuos para WMV y alrededor de 10-40 para los miembros descritos de la familia *Alphaflexiviridae* (Figura 5.1, A).

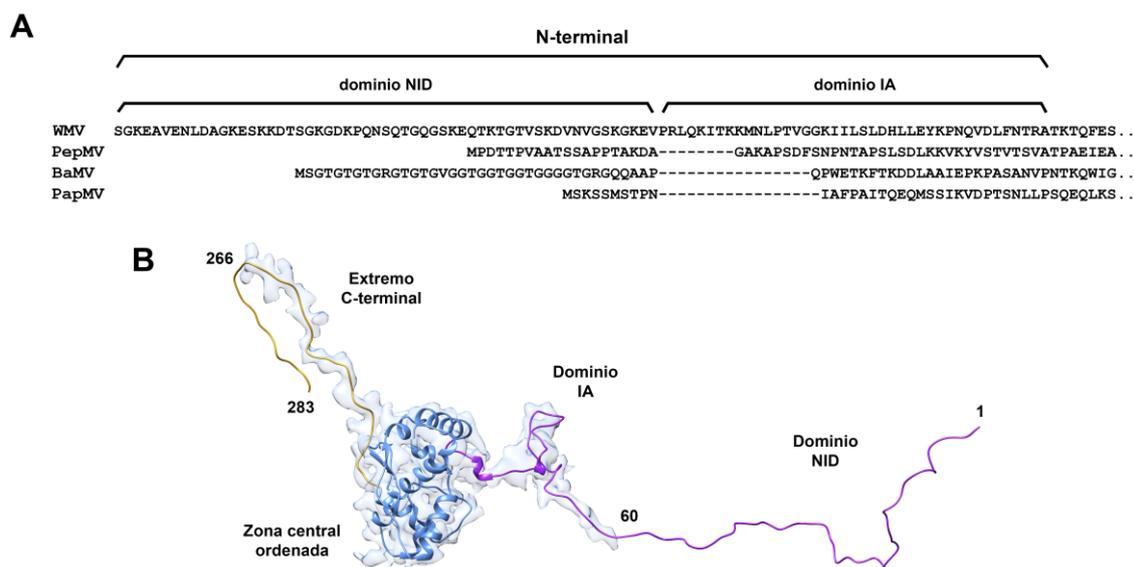


Figura 5.1: Extremos terminales de la proteína de la cápside de WMV. **(A)** Comparación de extremos N-terminal entre los virus filamentosos flexibles con modelo atómico disponible. **(B)** Modelo atómico hipotético de la CP de WMV completa, incluyendo los extremos flexibles no visibles en el mapa de densidad 3D. La zona central ordenada de la CP está coloreada en azul en el modelo atómico, el extremo C-terminal en amarillo y el extremo N-terminal en morado. La densidad en azul transparente corresponde a la parte visible de la CP en el mapa de densidad 3D del virión de WMV.

En nuestro mapa de densidad 3D para el virión de WMV podemos identificar la densidad atribuible al dominio IA del N-terminal (zona del N-terminal que interacciona con la subunidad vecina), sin embargo la densidad para los 60 residuos que constituyen el dominio NID está ausente (Figura 5.1, B), al igual que ocurre con los mapas de densidad para PepMV y BaMV (Agirrezabala et al. 2015, DiMaio et al. 2015) . Esto sugiere que este dominio se dispone hacia el exterior del virión y presenta una elevada flexibilidad. Tal y como vemos en el mapa

de densidad para el virión de WMV, y al igual que ocurre en los miembros de *Potexvirus* descritos, este dominio no está involucrado en el mantenimiento de la estructura de la cápside, de hecho, deleciones del mismo en distintas especies de *Potexvirus* y *Potyvirus* no impiden el ensamblaje del virión (Dolja et al. 1994, Atreya, Atreya, and Pirone 1991, Chapman et al. 1992). Por otro lado, este dominio parece estar involucrado en otros eventos durante la infección del virus como la interacción del virus con el hospedador. Deleciones de este dominio en algunas especies del género *Potexvirus*, como el *Virus X de la patata* (PVX), BaMV y PepMV, causan infecciones más leves con síntomas más suaves (Chapman et al. 1992, Lan et al. 2010, Duff-Farrier et al. 2015). La deleción de este dominio en *Potexvirus* y *Potyvirus* también afecta al movimiento célula-célula, de modo que se cree que este dominio podría interactuar con distintos factores virales y del hospedador involucrados en este proceso (Ozeki et al. 2009). En miembros del género *Potyvirus* también parece ser relevante en el movimiento sistémico de estos virus a través de la planta y parece ser esencial para la transmisión por vectores (Atreya, Atreya, and Pirone 1991, Dolja et al. 1994). En definitiva, la posible naturaleza desordenada del dominio NID podría contribuir a la interacción con diversos factores relevantes durante la infección del virus que condicionan el proceso de infección y la patogenicidad de este (Uversky 2011, Solovyev and Makarov 2016).

Al igual que en los miembros descritos de la familia *Alphaflexiviridae*, el dominio IA de WMV parece ser el principal elemento de la proteína de la cápside involucrado en la interacción con las CPs vecinas. Diferentes ensayos han mostrado cómo deleciones o mutaciones puntuales en el dominio IA impiden la polimerización de estas proteínas entre sí en varias especies del género *Potexvirus* (Lecours et al. 2006, Agirrezabala et al. 2015), así como en especies del género *Potyvirus* (Anindya and Savithri 2003). En WMV el dominio IA del extremo N-terminal adopta una posición muy similar a los miembros de *Potexvirus*, pues igualmente se aloja en una hendidura de la subunidad adyacente y también interactúa con el bolsillo hidrofóbico de la misma. Sin embargo, la zona de contacto entre el dominio IA y la CP adyacente es bastante mayor en WMV, y probablemente también en el resto de miembros del género *Potyvirus*, ya que una vez que alcanza el bolsillo hidrofóbico de la subunidad adyacente se extiende longitudinalmente hasta alcanzar una CP de la vuelta de hélice anterior (Figura 4.14), lo que supone interacciones adicionales que no están presentes en los miembros descritos de *Potexvirus*.

A partir de la caracterización estructural a alta resolución de viriones de *Potexvirus*, la flexibilidad de los virus filamentosos flexibles se atribuye a las características de las CPs y sus extremos flexibles. En teoría, el protagonismo de las interacciones transversales al eje de simetría del virión a través del extremo N-terminal, sumado a las débiles interacciones axiales (longitudinales al eje de simetría del virión) establecidas por el extremo C-terminal,

posibilitarían cierta libertad de movimiento relativo entre subunidades constituyendo una cápside flexible (Solovyev and Makarov 2016). En el virión de WMV vemos como el extremo N-terminal de la CP además de establecer interacciones transversales con la CP adyacente, también establece interacciones axiales con otra CP de la vuelta anterior (Figura 4.15 y Figura 5.2, A), en cambio, al tratarse de un extremo terminal flexible, estas interacciones también podrían permitir la flexibilidad de los viriones de WMV. Por otro lado, al analizar el modelo atómico de la CP de WMV identificamos cadenas laterales de residuos en la zona central ordenada de la CP que por cercanía podrían establecer interacciones axiales con residuos de CPs adyacentes; es el caso de los residuos D142 (N_i) y R227 (N_{i+9}), N144 (N_i) y E209 (N_{i+9}), y entre E165 (N_i) y R212 (N_{i+9}) (Figura 4.12, C). Sin embargo, estos residuos suponen un porcentaje pequeño en la contribución a las interacciones totales al compararlo con la contribución de los extremos terminales.

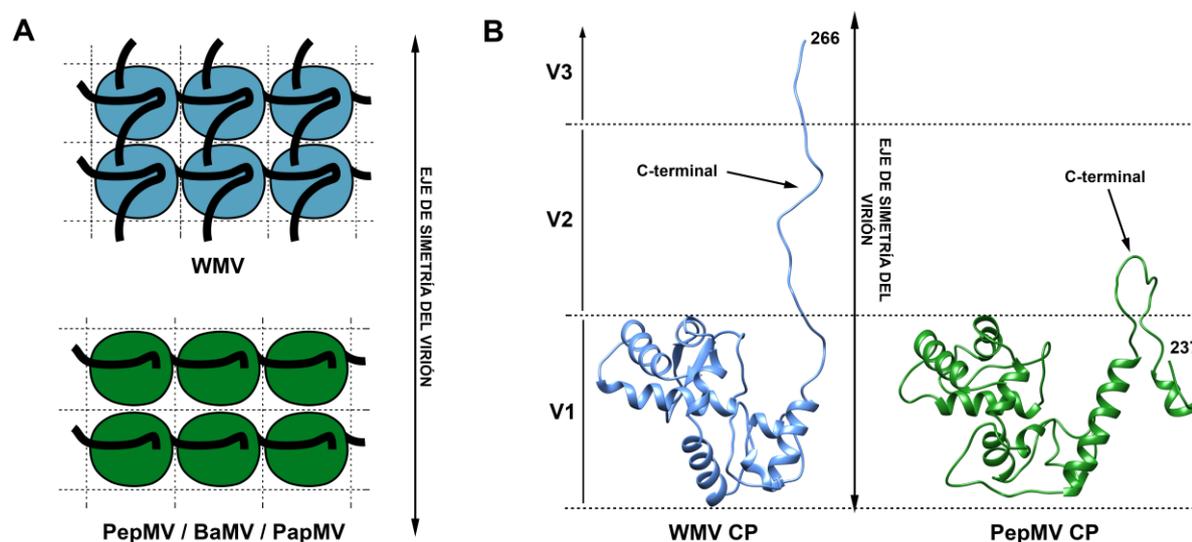


Figura 5.2: Comparación de extremos terminales entre *Potyvirus* y *Potexvirus*. **(A)** Esquema de interacciones del extremo N-terminal: arriba, interacciones axiales y transversales del N-terminal en el virión de WMV (género *Potyvirus*); abajo, interacciones transversales del N-terminal en los viriones de los miembros con modelo atómico disponible del género *Potexvirus* (PepMV, BaMV y PapMV). **(B)** Comparación de extremos C-terminal de la CP de WMV (modelo atómico azul) y la CP de PepMV (modelo atómico verde). Las franjas delimitadas con las líneas de puntos representan diferentes vueltas de hélice en el virión (V1, vuelta 1; V2, vuelta 2; y V3, vuelta 3).

En la proteína de la cápside de WMV, el extremo C-terminal está compuesto por 45 residuos orientados hacia el interior del virión, al igual que ocurre con la CP de los miembros de *Potexvirus*; en cambio, la disposición del C-terminal de WMV en el interior es completamente distinta. En los mapas de densidad 3D de los viriones de PepMV y BaMV se puede trazar en su totalidad la continuidad del extremo C-terminal, que se pliega y se mantiene en la misma

vuelta de hélice en la que se encuentra la CP a la que pertenece. En el caso del mapa de densidad 3D para el virión de WMV, el C-terminal se extiende longitudinalmente al eje de simetría del virión alcanzando dos vueltas de hélice en adelante (Figura 5.2, B). De los 45 residuos que constituyen el extremo C-terminal de la CP de WMV únicamente se pudieron trazar con fiabilidad 28, ya que no pudimos asignar fácilmente densidades para incluir los 17 restantes dentro del modelo atómico. Sin embargo, las densidades poco definidas en la zona central del mapa de densidad correspondiendo con el interior del virión (Figura 4.7, B) se pueden atribuir con bastante certeza a estos residuos. Esta zona del extremo C-terminal probablemente presente cierta flexibilidad y quizás presenta interacciones poco específicas con las regiones equivalentes del C-terminal de las CPs contiguas. Esa flexibilidad y heterogeneidad en los contactos entre estas regiones contribuirían a que, por el efecto del método de partículas individuales y el procesamiento helicoidal, las densidades correspondientes a esta zona no se definan igual que el resto de las partes de la CP que presentan una conformación más homogénea a lo largo virión.

Como vemos en la Figura 4.7, la parte visible del extremo C-terminal de la proteína de la cápside se dispone entre los extremos C-terminal de dos CPs de la vuelta de hélice siguiente y alcanza una CP dos vueltas de hélice por delante. A partir de ahí se plegaría hacia la zona central del virión constituyendo las densidades interiores poco definidas. En esta disposición los extremos C-terminal de las distintas CPs quedan muy próximos entre ellos, probablemente estableciendo contactos que contribuyen a esa disposición, tal y como apreciamos en ciertos residuos de esta zona (Figura 4.11, B). A pesar de que el extremo N-terminal parece tener mayor protagonismo en el ensamblaje del virión de WMV, el extremo C-terminal en nuestro modelo atómico presenta una superficie de contacto bastante relevante (al menos 1375 \AA^2), no mucho menor que la superficie de contacto del extremo N-terminal (1670 \AA^2), por lo que también parece tener cierta implicación en el ensamblaje del virión de WMV. Varios trabajos de investigación muestran cómo mutaciones puntuales en el extremo C-terminal de la CP de algunas especies del género *Potyvirus* impiden la formación de viriones *in vivo*, así como la formación de partículas similares a viriones (VLPs) *in vitro*. Es el caso de los *Potyvirus Virus del mosaico de la soja* (Seo et al. 2013, Anindya and Savithri 2003) y el *Virus del bandeo de las venas del pimiento* (PVBV) (Anindya and Savithri 2003). De manera similar, deleciones de más de 18 residuos del extremo C-terminal en PVX, un miembro del género *Potexvirus*, impiden la formación de VLPs *in vitro* (Zayakina et al. 2008). En cambio, la deleción total del extremo C-terminal (31 residuos) de la CP del *Virus del mosaico del Trébol blanco*, otro miembro del género *Potexvirus*, no impide la formación de filamentos similares a partículas virales (Forster et al. 1992). De modo que, a pesar de presentar una gran superficie de contacto y posibles interacciones con las CPs vecinas, la

relevancia del extremo C-terminal en el ensamblaje de WMV es variable y depende de la especie de virus.

El virión de WMV está constituido por miles de CPs homogéneas que empaquetan el RNA genómico y cada una abarca 5 ribonucleótidos del genoma del virus, de modo que las posibles combinaciones de los 4 ribonucleótidos que constituyen el genoma hacen que los fragmentos de RNA abarcados por cada CP sean heterogéneos. Al usar el método de partículas individuales y emplear la simetría helicoidal en el procesamiento, la combinación de miles de segmentos del virión de WMV resulta en que la densidad correspondiente a la CP proviene de una señal bastante homogénea formada por las proteínas de la cápside del virus pero la densidad correspondiente al RNA genómico proviene de una señal heterogénea formada por las diferentes combinaciones de ribonucleótidos que constituyen la secuencia del genoma de WMV. A partir de la densidad 3D correspondiente al RNA genómico se calculó el modelo atómico para los cinco ribonucleótidos unidos a cada CP. Como vemos en la figura 4.14, en nuestro modelo atómico el fragmento de RNA fue calculado como ribonucleótidos de uracilo orientando las bases nitrogenadas hacia el interior del virión, en cambio en el modelo atómico calculado para PepMV, el fragmento de RNA (también como ribonucleótidos de uracilo) dispone sus bases nitrogenadas hacia el exterior y el esqueleto de fosfatos hacia el interior, presentando una polaridad inversa a nuestro modelo propuesto para WMV (Agirrezabala et al. 2015). Si bien en el mapa de densidad 3D para PepMV la zona correspondiente para el RNA es más ambigua, en nuestro caso, las densidades correspondientes al RNA sugieren la orientación de las bases nitrogenadas del RNA hacia el interior del virión. Sin embargo la definición de las densidades en esta zona no nos permite asegurar con total certeza la orientación del RNA, probablemente debido al efecto de la heterogeneidad del RNA genómico en el procesamiento.

Aunque por ahora no podemos determinar con total seguridad la polaridad del RNA en el interior del virión y la orientación de las bases nitrogenadas y el esqueleto de fosfatos, es inequívoco el modo de interacción del RNA con la CP de WMV. Independientemente del sentido del RNA en el virión, es perfectamente visible en nuestro mapa de densidad 3D, así como en el de PepMV, que una base nitrogenada se acomoda orientándose hacia el interior del bolsillo que forma la CP (Figura 4.14). Además, en nuestro modelo atómico hemos podido identificar residuos que están interaccionando con el RNA independientemente de su orientación (Figura 4.15). Varios estudios muestran cómo mutaciones puntuales de varios residuos en el *Virus del tabaco etch* (TEV) impiden la formación de viriones (Dolja et al. 1994). La zona central ordenada de la CP de WMV presenta una identidad de secuencia del 72% con la zona equivalente de la secuencia de la CP de TEV, de modo que gracias a la generación del modelo atómico de la proteína de la cápside de WMV podemos identificar que

estos residuos mutados en TEV son equivalentes a los residuos R172 y D216 de la CP de WMV, principales protagonistas de la interacción de la CP con el RNA (Figura 4.14 y 4.15). Estos datos apoyan las conclusiones que aporta nuestro modelo atómico y refuerza la relevancia de estos residuos en su implicación en el mantenimiento del virión y la viabilidad del virus, muy probablemente a través de su interacción con el RNA genómico de WMV. Resultados similares se observaron en PepMV; mutantes para los residuos R124, D163 y K196 (equivalentes a los residuos R172, D216 y K236 en WMV) de la CP son incapaces de formar viriones y de llevar a cabo la transmisión célula-célula, probablemente debido a la incapacidad de interacción de la CP con el RNA del virus (Agirrezabala et al. 2015).

5.3 Homología estructural en virus filamentosos flexibles

La obtención del modelo atómico de la proteína de la cápside de WMV junto con la disponibilidad de modelos atómicos de las CPs de varios miembros del género *Potexvirus* nos permite por primera vez comparar la similitud entre CPs de dos familias de virus filamentosos flexibles, la familia *Potyvirdae* y la familia *Alphaflexivirdae*. A simple vista, la zona central de la proteína de la cápside de WMV (residuos 100-240) presenta un plegamiento muy similar al observado en la zona central de las CPs de los miembros del género *Potexvirus*, pese a la inexistencia de identidad y similitud de secuencia entre ellas. Podemos identificar elementos de estructura secundaria equivalentes entre ellas (Figura 4.18) y al comparar la zona central de estos modelos atómicos usando diferentes servidores, los valores de similitud estructural son muy altos en todos los casos ("TM-score") (Figura 4.17). Además, las proteínas de la cápside de los virus descritos de estas dos familias presentan dos extremos N- y C- terminal bastante largos, y en ambas familias el papel del extremo N-terminal es clave para el ensamblaje del virión (Anindya and Savithri 2003, Agirrezabala et al. 2015). En cambio, dentro de la zona central ordenada de la CP de WMV, la región que une la hélice G1 con la hélice H7 (residuos 219-227) (Figura 4.10, B) presenta algunas diferencias con la región equivalente en los miembros descritos de *Potexvirus* (residuos 166-185). Como vemos en la Figura 5.3, A, en *Potexvirus* esta región es 12 residuos más larga y alcanza casi la zona exterior del virión; en cambio, en WMV, al presentar menor longitud, este lazo es más pequeño y libera una zona en la cara exterior del virión que es ocupada por el N-terminal de la subunidad adyacente como parte del mecanismo de interacción entre CPs. Esta es una de las principales diferencias observadas dentro de la

zona central ordenada entre WMV y *Potexvirus* y, junto con las diferencias de longitud observadas entre los extremos N-terminal de las CPs analizadas, esta diferencia de 12 residuos permite al N-terminal de WMV acomodarse formando el ángulo de 90° y alcanzar la subunidad de la vuelta anterior (Figuras 4.12 y 4.16).

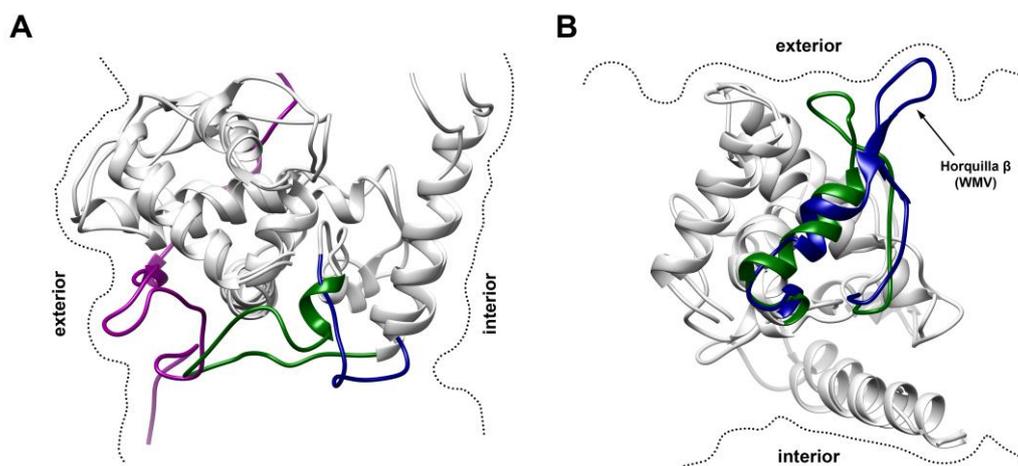


Figura 5.3: Comparación por alineamiento estructural de la zona central ordenada de las proteínas de la cápside de WMV y PepMV (gen. *Potexvirus*). Ambos modelos atómicos están coloreados en gris resaltando en azul (WMV) y verde (PepMV) las zonas variables. **(A)** Vista lateral de los modelos atómicos de las CPs. En azul, los residuos 219-227 de la CP de WMV; en verde los residuos 166-185 de la CP de PepMV ocupando la zona equivalente donde se aloja el extremo N-terminal (morado) de la CP adyacente de WMV. **(B)** Vista superior de los modelos atómicos de las CPs. En azul, el fragmento entre los residuos S140 y R172 de la CP de WMV, incluyendo la horquilla β; en verde, el fragmento entre los residuos S92 y R124 de la CP de PepMV.

Las CPs analizadas de estas dos familias de virus filamentosos flexibles son nucleoproteínas y su función principal es unirse al RNA del virus y empaquetarlo de forma ordenada para constituir el virión. Tal y como indican los resultados, las CPs de ambas familias presentan sitios de unión al RNA idénticos, constituidos por un bolsillo de unión donde se acomoda el RNA genómico orientando una base nitrogenada en el interior del bolsillo. Además, varios residuos ubicados en este bolsillo están conservados en la misma posición en todas las estructuras disponibles (a excepción de BaMV en el que la arginina está sustituida por una histidina), lo que pone de manifiesto su relevancia funcional (Figura 4.19). La conservación de estos residuos fue identificada con anterioridad entre algunos miembros de estas familias (Dolja et al. 1991), sin embargo, debido a la ausencia de estructuras para estas proteínas, no se ha podido determinar la relevancia de estos residuos a lo largo de las cuatro familias de virus filamentosos flexibles. Tal y como vemos en las secuencias consenso para las familias de virus filamentosos flexibles (Figura 4.20) los residuos equivalentes a S140, R172 y D216 en la CP de WMV están conservados en casi todas las CPs analizadas de la familia

Potyviriidae. De manera similar, los residuos equivalentes a S92, R124 y D163 en PepMV también están conservados en casi todas las CPs analizadas de la familia *Alphaflexiviridae*. Al obtener las secuencias consenso para las CPs de las familias *Betaflexiviridae* (género *Carlavirus* por un lado, y resto de géneros por otro) y *Closteroviridae*, tres residuos (serina, arginina y ácido aspártico) aparecen conservados en casi todos sus miembros. Por la distancia entre ellos podemos inferir que estos tres residuos conservados en las especies de *Closteroviridae* y *Betaflexiviridae* son equivalentes a los residuos conservados en *Potyviriidae* y *Alphaflexiviridae*, probablemente se encuentran cumpliendo la misma función formando parte del sitio de unión a RNA y serán relevantes en la interacción de la CP con este. La distancia media entre los residuos conservados R y D en las cuatro familias es muy similar, en torno a 40 residuos, y varía muy poco entre especies. En cambio, la distancia entre los residuos S y R presenta una mayor heterogeneidad: tanto en las familias *Potyviriidae* como *Alphaflexiviridae*, así como en los miembros del género *Carlavirus* de la familia *Betaflexiviridae* la distancia entre estos dos residuos es bastante similar (alrededor de 30 residuos), y presenta poca variabilidad entre especies; en la familia *Closteroviridae* esta distancia es algo mayor (alrededor de 40 residuos); y en la familia *Betaflexiviridae* (excluyendo el género *Carlavirus*) la distancia es casi de 60 residuos y varía bastante entre los distintos miembros (Figura 4.20). La variabilidad en la distancia entre estos dos residuos (S y R) no necesariamente tendría que afectar al plegamiento común de las CPs de estas familias, ya que la zona que conecta estos dos residuos (serina 140 y arginina 172 en WMV) corresponde con la región de la horquilla β en WMV que está expuesta hacia el exterior del virión y variaciones en la longitud de esta región podrían acomodarse sin afectar al plegamiento general de la CP y sin afectar al modo de formación del virión (Figura 5.3, B). Junto con el fragmento que une la hélice G1 con la hélice H7 (Figura 5.3, A) esta región dentro de la zona central ordenada es la que mayores diferencias muestra entre los modelos atómicos disponibles, ya que la horquilla β presente en la CP de WMV no se encuentra en los miembros descritos de *Potexvirus* y adquiere una disposición ligeramente distinta en esta zona, ya que debe permitir la acomodación del extremo N-terminal procedente de la CP de la vuelta siguiente (Figura 5.3, B).

Las cuatro familias de virus filamentosos flexibles presentan viriones muy similares entre ellas; como resultado de este trabajo se ha podido confirmar que al menos dos familias aparentemente no relacionadas (*Potyviriidae* y *Alphaflexiviridae*) presentan CPs con un plegamiento muy similar pese a la ausencia de homología de secuencia entre ellas. Aunque no existen datos estructurales para las CPs de miembros de la familia *Betaflexiviridae* y *Closteroviridae*, la similitud ya observada entre las familias *Potyviriidae* y *Alphaflexiviridae*

junto con la existencia de estos tres residuos conservados a lo largo de estas cuatro familias señala un plegamiento común para las CPs de los virus filamentosos flexibles.

Los virus filamentosos flexibles constituyen uno de los mayores grupos de patógenos vegetales y su efecto sobre los cultivos tiene importantes repercusiones económicas. Aparte de medidas de contención, control sobre los vectores que transmiten estos virus o el uso de cultivos resistentes, a día de hoy no existen tratamientos que afecten directamente a estos virus para controlar las infecciones. Dado que el sitio de unión a RNA de la CP de estos virus involucrando los residuos S, R y D parece estar conservado en todas las familias de virus filamentosos flexibles, supone una potente diana para el desarrollo de compuestos antivirales que puedan interferir con el empaquetamiento del genoma, el ensamble de los viriones de este grupo de virus y en definitiva con cualquier proceso del ciclo de vida del virus en el que la interacción CP-ssRNA sea necesaria, como la diseminación del virus en el interior de la planta a través de los movimientos célula-célula.

5.4 Homología estructural entre nucleoproteínas de virus ssRNA

Al buscar homólogos estructurales para la zona central ordenada de la CP de WMV en los servidores Dali y MATRAS (Holm and Rosenström 2010, Kawabata 2003), como cabría esperar, las proteínas más similares fueron las CPs de los miembros descritos del género *Potexvirus*. Sin embargo, aunque con menor similitud, también aparecieron como resultado de la búsqueda las nucleoproteínas de varios virus (-)ssRNA de genoma segmentado (sNSV) que infectan animales; en concreto, miembros de la familia *Orthomyxoviridae* y varios miembros del orden *Bunyavirales*, en concreto aquellos con estructura conocida del género *Phlebovirus* (familia *Phenuiviridae*). La similitud estructural entre las nucleoproteínas del género *Phlebovirus* y las CPs del género *Potexvirus* ha sido descrita con anterioridad (Agirrezabala et al. 2015), sin embargo la similitud estructural observada entre la CP de WMV y una región de la NP del virus Influenza A es totalmente novedosa. Aunque se ha señalado con anterioridad la existencia de cierto parecido entre las NPs de Influenza A y RVFV (“Rift Valley Fever Virus”), del gen. *Phlebovirus*, (Abrescia et al. 2012), la existencia de homología estructural entre ellas nunca había sido puesta de manifiesto (Ruigrok, Crépin, and Kolakofsky 2011). De hecho, el valor “TM-score” resultante de comparar las regiones similares de estas NPs entre sí con el servidor TM-align es inferior a 0,5 y por lo tanto

insuficiente para considerar la existencia de homología estructural por simple comparación entre ellas. En cambio, al comparar las regiones que presentan similitud de las NPs de Influenza A y RVFV con la zona central ordenada de la CP de WMV, los valores “TM-score” en ambos casos son superiores a 0,5 lo que indicaría la existencia de homología estructural de ambas NPs con la CP de WMV. Aparece, por lo tanto, una relación de homología estructural entre las NPs de Influenza A y RVFV a través de la comparación con la CP de WMV. Valores similares también indican la existencia de homología estructural entre la zona central ordenada de la CP de WMV y determinadas regiones de NPs del resto de miembros con estructura descrita de la familia *Orthomyxoviridae* (géneros *Influenzavirus B* e *Isavirus*), y con regiones concretas de las NPs del resto de miembros con estructura descrita del género *Phlebovirus*.

A diferencia de las CP de los virus filamentosos flexibles, que forman cápsides virales desnudas constituidas únicamente por el RNA genómico y la proteína de la cápside, las NPs de los virus sNSV se encargan de empaquetar el genoma viral formando una RNP que a su vez se encuentra contenida dentro de una envuelta lipídica (Figura 5.4). Los miembros del orden *Bunyvirales* organizan el genoma en RNPs flexibles en el interior de una membrana que integra glicoproteínas ordenadas en disposición icosaédrica (Huiskonen et al. 2009, Freiberg et al. 2008) (Figura 5.4, B). Además, en el interior del virión, asociadas a las RNP, se encuentran varias copias de la RNA polimerasa del virus (proteína L) (Elliott 2008). En cambio, las RNPs que empaquetan el genoma en los miembros de la familia *Orthomyxoviridae* forman una doble hélice con dos hebras de NPs protegidas por una envuelta lipídica de forma variable recubierta por diferentes proteínas virales (Arranz et al. 2012) (Figura 5.4, C). Al igual que los anteriores, los viriones de la familia *Orthomyxoviridae* también presentan varias copias de la RNA polimerasa viral asociadas a las RNPs (Stubbs and Te Velthuis 2014).

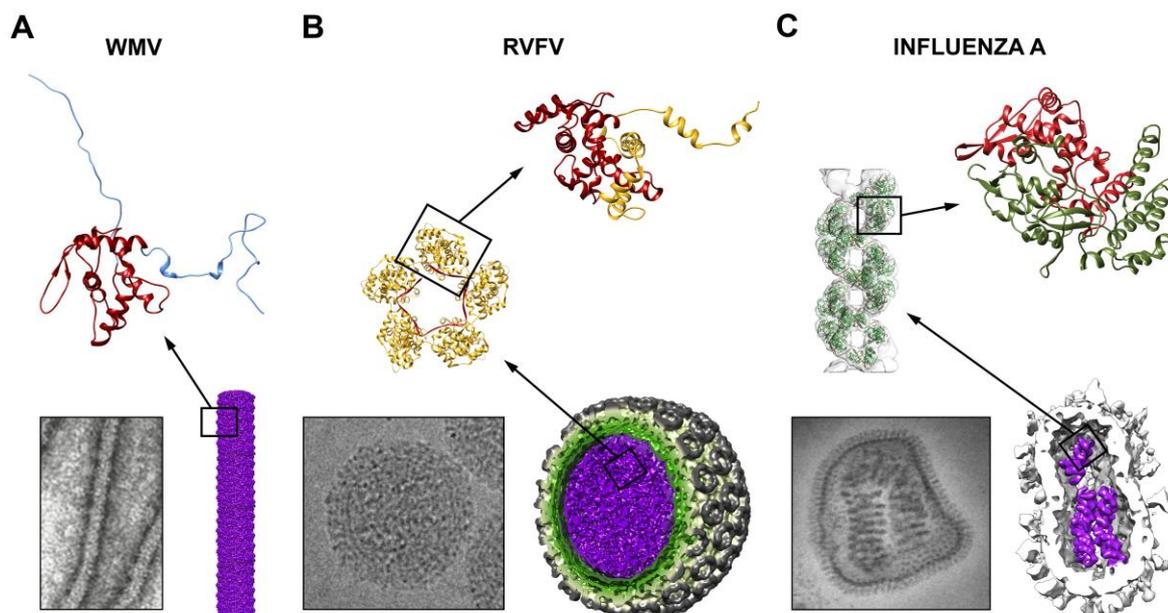


Figura 5.4: Comparación de morfología y RNP de WMV, RVFV e Influenza A. En todos los casos se muestra una imagen de microscopía electrónica de los viriones de cada virus, una reconstrucción 3D del virión y el modelo atómico de las NPs. (A) Virión del virus filamentosos flexible WMV (fam. *Potyviridae*); mapa 3D resultante de este trabajo coloreado en morado y modelo 3D resultante de este trabajo: la parte central ordenada en común con las NPs analizadas está coloreada en rojo oscuro, los extremos N-t y C-t en azul. (B) Virión de RVFV (“*Rift Valley Fever Virus*” gen. *Phlebovirus*, fam. *Phenuiviridae*); dentro del mapa 3D del virión (EMDB-5124 (Sherman et al. 2009)) las RNPs están coloreadas en morado; el modelo atómico de una parte de la RNP formada por 5 NPs se presenta coloreado en amarillo (NP) y rojo (RNA) (código pdb 4H5O (Raymond et al. 2012)); en una NP individual, la zona de topología común al resto de NPs analizadas se muestra coloreada en rojo, el resto de la NP en amarillo. (C) Virión de Influenza A (fam. *Orthomyxoviridae*). Se muestra un mapa de densidad 3D del virión con las RNPs en su interior coloreadas en morado (Arranz et al. 2012); un detalle de la densidad correspondiente a la RNP en gris semitransparente y el modelo atómico coloreado en verde en el interior (EMDB-2205 (Arranz et al. 2012)); en una NP individual (código pdb 3zdp (Chenavas et al. 2013)) la zona con estructura similar al resto de NPs analizadas se muestra coloreada en rojo, el resto de la NP en verde.

Las CPs de los virus filamentosos flexibles también son nucleoproteínas (NPs), ya que su función es la interacción con el RNA genómico del virus y su empaquetamiento, y aunque formen viriones tan distintos de los de los virus sNSV, comparten función con las NPs de estos (Figura 5.4). En todos los casos, la región en común entre todas las NPs de los virus analizados es la zona de interacción con el RNA del virus (Figura 4.21), de modo que esta similitud estructural está íntimamente ligada a su función como NPs. El hecho de que estas proteínas tengan la misma función, altos valores “TM-score” de similitud estructural, un modo muy similar de interacción con el RNA alojado en una hendidura, así como que los extremos N-t y C-t de las regiones similares acaben en posiciones parecidas, refuerza la idea de existencia de homología estructural entre estas proteínas. La presencia de homología estructural entre proteínas normalmente es interpretada como el resultado de un origen

evolutivo común (homología evolutiva) en lugar de como un resultado de convergencia evolutiva (analogía evolutiva). Esto está respaldado por el hecho de que la estructura de las proteínas está más conservada que su secuencia de amino ácidos (Illergård, Ardell, and Elofsson 2009) y que los eventos de convergencia evolutiva entre dominios de proteínas es algo infrecuente (Gough 2005). La similitud estructural entre las NPs analizadas está respaldada por su función como proteínas virales encargadas del empaquetamiento del RNA genómico del virus, de modo que podemos suponer que los genes de las NPs de los virus filamentosos flexibles y de al menos dos familias de sNSV comparten un ancestro común. Además, la ausencia de homología estructural por comparación directa entre las NPs de las familias de sNSV y su existencia al compararlas por separado con la CP de WMV sugiere que esta última presenta un plegamiento más cercano al ancestro común a todas ellas (Figura 5.5).

Los virus La Crosse y TSWT, recientemente asignados a las familias *Peribunyaviridae* y *Tospoviridae* respectivamente, incluidos en el orden *Bunyvirales*, se encontraban hasta hace poco incluidos en la familia *Bunyaviridae* junto con los miembros del género *Phlebovirus* (actualmente incluidos en familia *Phenuiviridae*) (Maes et al. 2018). Las NPs de estos dos virus, que presentan homología estructural entre ellas (TM-score mayor de 0,5), aparecieron en las búsquedas de estructuras similares a la zona central ordenada de la CP de WMV con Dali y MATRAS y muestran una topología parecida a la observada en el resto de NPs comparadas, incluyendo un sitio de unión a RNA parecido y misma posición de los extremos N-t y C-t en las zonas comparadas (Figura 4.21). Sin embargo, pese a su relación filogenética con los miembros del género *Phlebovirus*, los valores “TM-score” resultantes de la comparación de las regiones similares de las NPs de los virus “La Crosse” y TSWT con estos y con el resto de NPs analizadas son inferiores a 0,5, insuficiente para afirmar la existencia de homología estructural. Otras NPs descritas del orden *Bunyvirales*, pertenecientes a las familias *Hantaviridae* y *Nairoviridae*, también tienen una región con una topología similar a las NPs analizadas, y en el caso de los miembros de la familia *Hantaviridae* esta región presenta homología estructural con las NPs de las familias *Peribunyaviridae* y *Tospoviridae* según los valores de TM-align, lo que sugiere una relación evolutiva más cercana entre ellas que al resto de familias del orden (Figura 5.5). En definitiva, la topología general entre las NPs del orden *Bunyvirales*, conservando en algunos casos homología estructural, sugiere un ancestro común entre ellas, sin embargo, los genes de estas NPs parecen haber divergido de tal manera dentro de este orden que la similitud se vuelve más difusa entre algunas familias. Si esto es así, la familia *Phenuiviridae*, y en concreto los miembros del género *Phlebovirus* presentarían unas NPs más cercanas al ancestro común dentro de este orden,

ya que estas mantienen homología estructural con NPs de familias ajenas a este orden como son *Alphaflexiviridae* y *Potyviridae* del grupo de virus filamentosos flexibles.

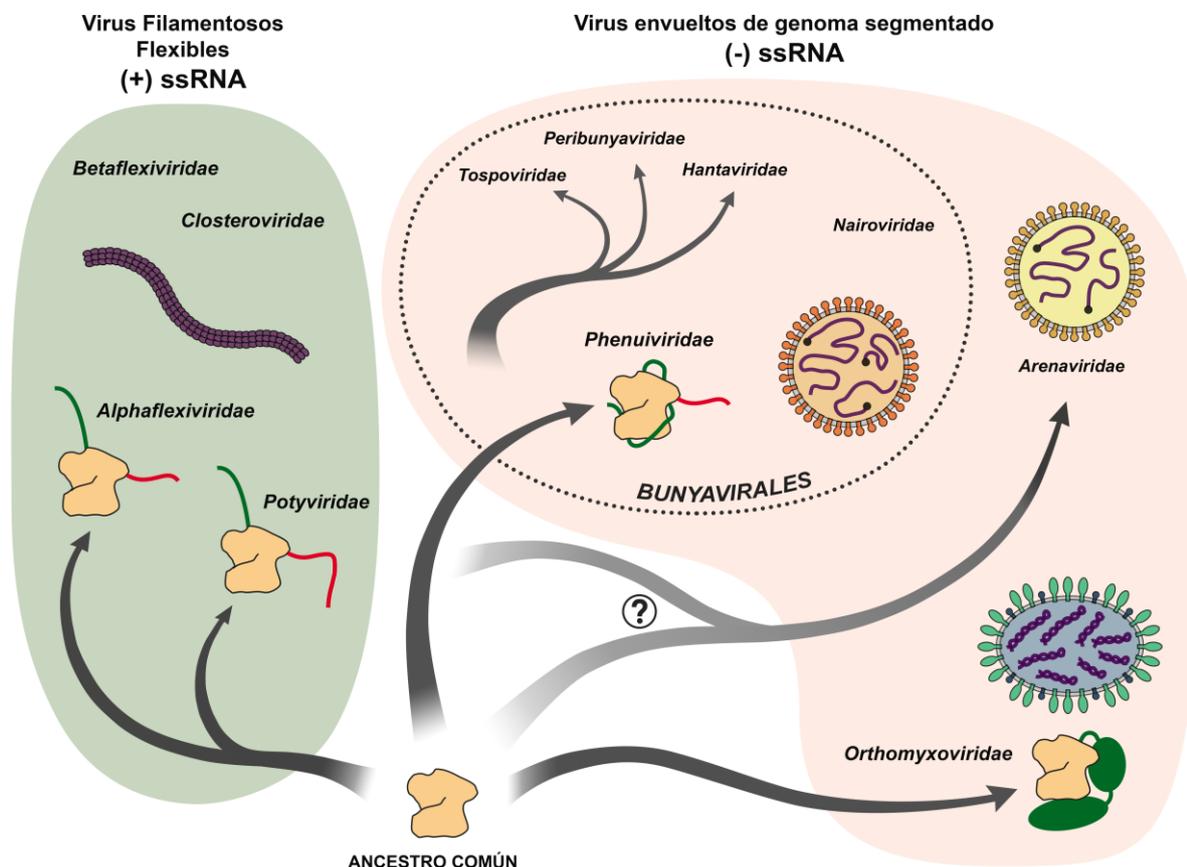


Figura 5.5: Modelo evolutivo de nucleoproteínas de virus ssRNA. Los dos conjuntos de virus analizados están separados en dos grupos principales: virus que infectan plantas constituidos por nucleoproteínas desnudas formando un filamento (izquierda), y virus envueltos que infectan mayoritariamente animales (derecha). La zona en común entre nucleoproteínas se muestra como una figura irregular coloreada en marrón y las extensiones N- y C- terminal se presentan en rojo y verde respectivamente.

Algo similar ocurre con la familia *Arenaviridae*; las NPs de esta familia de sNSV presentan un dominio de unión al RNA genómico del virus que muestra ligeras similitudes con las NPs analizadas y sólo presenta homología estructural significativa ("TM-score" mayor de 0,5) con la región equivalente de la NP de RVFV (género *Phlebovirus*). De modo que, quizás los genes de las NPs de la familia *Arenaviridae* podrían haber divergido de un ancestro común con los genes de las NPs analizadas de sSNV y virus filamentosos flexibles, o con los genes de las NPs del género *Phlebovirus* en particular (Figura 5.5).

En general, las regiones de las NPs analizadas muestran una topología bastante similar asociada a la función que desempeñan, lo que sugiere un ancestro común a todas ellas y pone de manifiesto la existencia de una relación evolutiva entre los virus sNSV y los virus filamentosos flexibles. Dentro de esta relación evolutiva, las NPs de sNSV son bastante diversas, presentando dominios adicionales únicos dependiendo de cada familia. Esto sugiere una gran divergencia evolutiva entre las NPs de estas familias de virus, probablemente favorecida por estar protegidas por envueltas lipídicas, lo que explica que los valores de similitud estructural obtenidos en algunas de las comparaciones realizadas no reflejen homología estructural entre ellas, pese a la existencia de una topología general muy similar en regiones concretas de estas NPs.

La evolución de los virus con genoma de RNA es difícil de determinar, sin embargo, se cree que los eventos de transmisión horizontal de virus (HVT) entre hospedadores juegan un papel muy relevante en la misma (Dolja and Koonin 2018, Shi, Zhang, and Holmes 2018). Estos, junto con la transferencia de genes entre virus y hospedador parecen ser mecanismos comunes en la evolución de los virus (Geoghegan, Duchêne, and Holmes 2017, Aiewsakun and Katzourakis 2015). Por ejemplo, secuencias de la CP del potyvirus PVY se han encontrado insertadas en genomas de vides (Tanne and Sela 2005) y varios elementos virales endógenos derivados de bunyavirus y orthomyxovirus han sido identificados en genomas de insectos y crustáceos (Katzourakis and Gifford 2010, Thézé et al. 2014). Varios trabajos han puesto de manifiesto la enorme diversidad genética de los virus de RNA en invertebrados (Li et al. 2015, Shi, Zhang, and Holmes 2018); los análisis filogenéticos de los viromas de estos virus muestran eventos frecuentes de transferencia horizontal de genes y transferencia horizontal de virus entre hospedadores, especialmente desde invertebrados a plantas y animales, suponiendo los primeros un nicho para la expansión evolutiva de los virus RNA de eucariotas (Koonin, Dolja, and Krupovic 2015, Dolja and Koonin 2018). Los insectos juegan un papel fundamental como vectores en la transmisión de los virus de familias analizadas en este trabajo, y estos eventos comunes podrían ser los responsables de la distribución de un gen de una NP a lo largo de familias de virus con hospedadores muy distintos; dicho gen posteriormente habría divergido dentro de cada nuevo grupo de virus hasta dar lugar a las distintas NPs analizadas.

5.4.1 Nueva familia de plegamiento para NPs de virus ssRNA

Además de las clasificaciones habituales de los virus en base a la organización de su genoma (Clasificación Baltimore) (Baltimore 1971) o la clasificación llevada a cabo por el “International Committee on Taxonomy of Viruses” (ICTV) se ha planteado una clasificación basada en la estructura de las proteínas virales (Abrescia et al. 2012). Esta clasificación se apoya en la conservación de determinados plegamientos en la estructura de las proteínas a lo largo de largas distancias evolutivas (Caetano-Anollés and Nasir 2012, Caetano-Anollés and Caetano-Anollés 2003, Lundin et al. 2012, Illergård, Ardell, and Elofsson 2009). Esto puede aplicarse especialmente a las proteínas de la cápside de los virus y su mecanismo de ensamblaje debido a la presión evolutiva sobre estas para mantener eficientemente la morfología de la partícula viral (Abrescia et al. 2012). En base a esta clasificación se han definido cuatro linajes principales que abarcan decenas de familias de virus agrupados según el plegamiento que presentan principalmente sus proteínas de la cápside, quedando definidos estos linajes por varias superfamilias de plegamiento (Abrescia et al. 2012, Nasir and Caetano-Anollés 2017).

Estos linajes abarcan principalmente los virus de cápsides icosaédricas, dejando sin agrupar, entre otros, gran parte de los virus envueltos y los virus filamentosos helicoidales. En los virus envueltos que no presentan una cápside como los virus icosaédricos supone un problema la elección de una proteína viral que tenga el potencial para ser usada como proteína de referencia para la clasificación de estas familias de virus. Una de las proteínas candidatas en este tipo de virus son las nucleoproteínas que empaquetan el genoma (Abrescia et al. 2012). Un claro ejemplo de este tipo de virus son los virus (-)ssRNA, compuestos por aquellos que presentan el genoma sin segmentar (nsNSV) ubicados en el orden *Mononegavirales*, y aquellos que presentan el genoma segmentado (sNSV), actualmente agrupados en el orden *Bunyavirales* y las familias *Orthomyxoviridae* y *Arenaviridae*. Estas familias no están ubicadas en ningún linaje en base a la estructura de sus NPs, sin embargo, los virus nsNSV tienen unas NPs muy parecidas en todas las familias que componen el orden *Mononegavirales*, presentando un plegamiento que podría caracterizar a este orden, muy distinto al observado para las NPs de sNSV (Abrescia et al. 2012).

En el caso de los virus filamentosos flexibles, las proteínas candidatas para esta clasificación sin duda serían las CPs que componen la cápside. Al igual que los virus nsNSV y sNSV, los virus filamentosos flexibles no se encuentran agrupados en ningún linaje en base a la estructura de sus CPs. En cambio, a la vista de los resultados obtenidos, el plegamiento común observado entre las NPs de las familias que componen los virus sNSV, especialmente entre las familias *Orthomyxoviridae* y *Phenuiviridae*, y las familias que componen el grupo de

virus filamentosos flexibles podría caracterizar a este conjunto de virus. De hecho, al comparar las distancias evolutivas obtenidas a partir de los valores de similitud estructural calculados con TM-align entre las NPs de nsNSV, sNSV y virus filamentosos flexibles se obtienen dos grupos claramente definidos que abarcan las NPs de las familias de nsNSV por un lado, y las familias *Orthomyxoviridae*, *Phenuiviridae*, *Potyviridae* y *Alphaflexiviridae* por otro (Figura 5.6). Las NPs de las familias *Peribunyaviridae* y *Tospoviridae* también analizadas en este trabajo, pese a presentar una topología similar a la observada para la familia *Phenuiviridae*, aparecen como un grupo aparte en la comparación debido a la gran divergencia observada en las NPs de sNSV.

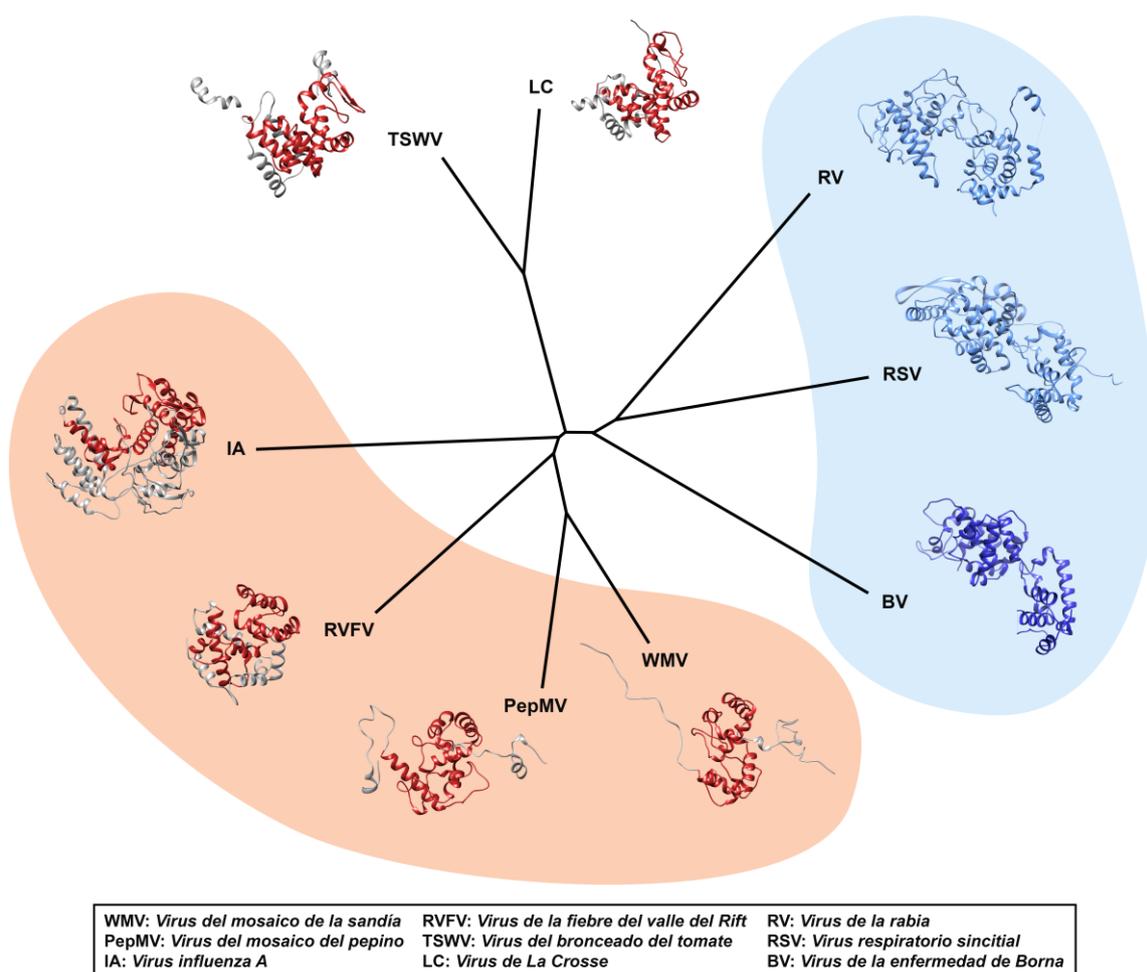


Figura 5.6: Árbol filogenético sin raíz basado en similitud estructural de varias nucleoproteínas de virus ssRNA. A partir de valores de similitud estructural se ha elaborado una matriz de distancias evolutivas (Anexo II) (Lakshmi et al. 2015, Agarwal et al. 2009) que se ha usado para elaborar el árbol con PHYLIP (Felsenstein 1989). Los modelos atómicos usados para la comparación son: RV, código pdb 2gtt; RSV, código pdb 2wj8; BV, código pdb 1n93; TSWV, código pdb 5ip2; LC, 4bgp; RVFV, 3ov9; Influenza A, 3zdp; y PepMV, 5fn1. En los modelos atómicos, la zona comparada de las NPs de sNSV y virus filamentosos flexibles está coloreada en rojo y en gris las extensiones N- y C- terminal.

Dentro del contexto de la clasificación de virus en base la estructura que presenten determinadas proteínas virales (Abrescia et al. 2012, Nasir and Caetano-Anollés 2017), el plegamiento común observado en algunas regiones de las NPs analizadas en este trabajo podría definir a un nuevo conjunto de virus, claramente diferenciado del grupo nsNSV, que abarque al menos algunas de las familias de sNSV ((-)ssRNA) y los virus filamentosos flexibles ((+)ssRNA).

CAPÍTULO 6
CONCLUSIONES

CONCLUSIONES

1. En este trabajo se ha determinado por primera vez la estructura del virión de un miembro de la familia Potyviridae a alta resolución, con una resolución global estimada de 4 Å. Los detalles del mapa de densidad 3D han permitido generar un modelo atómico 3D tanto para la CP como para el ssRNA genómico de WMV.
2. El virión filamentoso de la cepa M116 de WMV presenta unos 850 nm de longitud y 15 nm de diámetro, siguiendo una organización helicoidal de 8,81 unidades asimétricas por vuelta y 35,15 Å de paso de vuelta, valores muy similares a los del resto de virus filamentosos flexibles caracterizados.
3. El extremo N-terminal de la CP de WMV se orienta hacia el exterior del virión está formado por 100 residuos, de los cuales aparentemente los 60 primeros conforman el dominio NID. Los 40 residuos restantes constituyen el dominio IA involucrado en el ensamblaje del virión estableciendo contactos axiales y longitudinales con la CP adyacente.
4. El extremo C-terminal de la CP de WMV se localiza en el interior del filamento y está formado por 44 residuos que se disponen en su mayoría longitudinalmente al eje de simetría del virión, y aparentemente podrían establecer contactos con las CPs adyacentes que aparentemente contribuyen al ensamblaje del virión.
5. La CP de WMV muestra un plegamiento común con las CPs descritas para los miembros de la familia *Alphaflexiviridae*, lo que pone de manifiesto un plegamiento común para las CPs de los virus filamentosos flexibles.
6. Los residuos S140, R172, D216 que forman parte del bolsillo de interacción con el RNA de la CP de WMV están conservados en las CPs de los virus filamentosos flexibles, y dada su relevancia en la estabilidad de estos virus constituyen una diana a tener en cuenta en el desarrollo de posibles compuestos antivirales.
7. Las CPs de los virus filamentosos flexibles presentan homología estructural con NPs de miembros de las familias *Phenuiviridae* y *Orthomyxoviridae*, señalando un posible relación evolutiva entre las NPs que forman las cápsides helicoidales de los virus filamentosos flexibles de (+)ssRNA y las NPs que empaquetan el genoma de los virus de (-)ssRNA segmentado

ANEXO I

DISCUSIÓN

```
      :  --  -----  -      H7      ---
SecA :  --  S-----S-S TTHHHHHHHHH---H
218  :YE--VTS-----K-TPNRAREAIAQMKA---AA: 239
      * * * *
172  :FDGLYDPTCLASELPYDAPSEIDRMAYATFKTIQIKI: 208
SecB :GGGGT TTS  S  SS TT HHHHHHHHHHHHHHHHH
      :
      H7
```

Zona central de CP de WMV frente a zona central de CP de PapMV:

SecA: zona central de CP de WMV
SecB: zona central de CP de PapMV

[SIMILARITY] Seq 16.9 % Sec 79.8 % Exp 74.2 % CRMS 3.37 Å DRMS 2.80 Å
[SCORE] ScSSE 393.5 ScEnv 2371.2 ScDis 71022.6 Rdis 37.3 (%) Rsse 24.4 (%)
[RELIABILITY] Superfamily 84.6 % Fold 98.0 %

```
      : H1      H2      E1      E2
SecA : HHHHHHHHHHHHHHT STTHHHHHHHHHHHHS  S  SSEE STT  EEE
101  :TKTQFESWYSAVKVEYDLNDEQMGVIMNGFMVWCIDNGTSPDVNGVWVMMDGEEQVEYPL: 160
      * * * * *
35  :SQEQLKSVSTLMVA-AKVPAASVTTVALELVNFCYDNGSSAYTTVTGPSSI----PEISL: 89
SecB : HHHHHHHHHHHHHH-TT GGGHHHHHHHHHHHHHHH TT  SB SS----TT BH
      : H1      -      H2      - - - - H

      :H3      ---  H4      H5      H6
SecA :HHHHHT ---SS SHHHHTS SHHHHHHHHHHHHS  HHHHTT  GGGGGG SS
161  :KPIVENA---KPTLRQIMHHFSDAAEAYIEMRNSESPYMPRYGLLRNLRARELARYAFDF: 217
      ** * * * *
90  :AQLASIVKASGTSLRKFCRYFAPIIWNLRDCK-----MAPANWEASGY-KPSAKFAAFDF: 143
SecB :HHHHHHHHHTTS HHHHHHTHHHHHHHTTT-----S  TTTT - GGGGGG STT
      :3      H4      H5      - - - - -

      : -  -----  H7
SecA : -  -----SSS TTHHHHHHH
218  :Y-EVT-----SKTPNRAREAIAQM: 235
      * * * *
144  :FDGVENPAAMQPPSGLTRSPTQEERIANATN: 174
SecB :GGGTT TTS  TT  SS  HHHHHHHH
      :
      H6
```

Las figuras de mérito para el alineamiento 3D generadas por MATRAS son: Seq, identidad de secuencia (%); Sec, identidad de estructura secundaria (%); CRMS, desviación estándar de la media cuadrática (Å) de las posiciones de los átomos de los carbonos α de los residuos alineados; DRMS, desviación estándar de la media cuadrática (Å) de las posiciones de los átomos de los carbonos β de los residuos alineados; ScSSE, valor SSE (elementos de estructura secundaria); ScDis, valor de distancias entre modelos; Rdis, valor de distancias normalizado (%); Rsse, valor SSE normalizado; "Reliability" ("superfamily" y "fold"), indica la probabilidad de que el par de estructuras tengan el mismo plegamiento y formen parte de la misma familia de plegamiento en la base de datos SCOP 1.59.

Anexo I.b: Representación del alineamiento de los modelos 3D de la zona central de la CPs de WMV y las NP de varios virus (-)ssRNA.

Zona central de CP de WMV frente a zona central de NP de Influenza A:

SecA: Zona central de CP de WMV

SecB: Zona central de NP de Influenza A

[SIMILARITY] Seq 7.1 % Sec 74.0 % Exp 73.2 % CRMS 5.58 A DRMS 4.09 A
 [SCORE] ScSSE 367.0 ScDis 43448.2 Rdis 16.9 (%) Rsse 12.2 (%)
 [RELIABILITY] Superfamily 33.7 % Fold 66.3 %

```

      : H1                H--2                -----                E
SecA : HHHHHHHHHHHHHHT  STH--HHHHHHHHH-----HHHS  S  SE
101 :TKTQFESWYSAVKVEYDLNDEQM--GVIMNGFMVW-----CIDNGTSPDVNGV: 146
      *      *      *      *
32 :MIDGIGRFYIQMCTELKLSDYEGRLIQNSLTIERMVLSAFDERRNKYLEEHpSAGKDPKK: 91
SecB :HHHHHHHHHHHHHHHT  HHHHT  HHHHHHHHHHHHHHHHTTSSSS  TT SSTTE
      :                H2      H3                E

      :1      -----      E2  H3      -----      H4      H5
SecA :EE  S-----TT  EEE  HHHHHT  S-----S  SHHHHTS  SHHHHHHHHHHHHHS
147 :WVMMD-----GEEQVEYPLKPIVENAK-----PTLRQIMHHFSDAAEAYIEMRNSES: 193
      *      *      *      *
92 :TGGPIYRRVDGKWRRELILYDKEEIRRIWRQANNGDDATAGLTHMMIWHSNLNDATYQRT: 151
SecB :E  EEEEEETT EEEEE  EEHHHHHHHHHHHHTTTS  HHHHHHHHHHHHHHHHHH  H
      :1  E2      E3      E4H4                H5                H

      :      H6                -----      -  -  -                H7
SecA :      HHHHTT  -----GGGGGG  S-S  -  -  -  SSS  TTHHHHHHHH
194 :PYMPRYGLLRNLRA-----RELARYAFD-FY--EVTSKTPNRAREAIAQM: 235
      *
152 :R-----ALVRTGMDPRMCSLMQGSTLPRRSGAAGA AVKGVGTMVMEIIR: 195
SecB :H-----HHHTTT  GGGGGG  TTTTS  SS  TTTGGG  HHHHHHHHHH
      :6-----                H7
  
```

Zona central de CP de WMV frente a zona central de NP de RVFV:

SecA: Zona central de CP de WMV

SecB: Zona central de NP de RVFV

[SIMILARITY] Seq 8.0 % Sec 67.2 % Exp 76.8 % CRMS 5.02 A DRMS 3.65 A
 [SCORE] ScSSE 370.9 ScDis 46332.8 Rdis 23.2 (%) Rsse 13.9 (%)
 [RELIABILITY] Superfamily 50.7 % Fold 85.7 %

```

      :                H2                E1                E2--  H
SecA :HHHHHHHHHHHHHT  STHHHHHHHHHHHHHHS  S  SEEE  STT  EE--E  H
104 :QFESWYSAVKVEYDLNDEQMGVIMNGFMVWCIDNGTSPDVNGVWVMMDGEEQVEY--PLK: 161
      *      *      *      *      *      *
34 :DARRVIELLKQYG---GADWEKDAKKMIVLALTRGN---KPRRMMKMSKEGKATVEALI: 87
SecB : THHHHHHHHHHS---TTTHHHHHHHHHHHHHT  S---  HHHHTTTS  HHHHHHHHHH
      :  H1      ---  H2                ---  H3                H4

      :---3  ----      H4      H5                H6                ---
SecA :---HHH---HT  SS  SHHHHTS  SHHHHHHHHHHHHHS  HHHHTT  ---GGGG
      :---PIV---ENAKPTLRQIMHHFSDAAEAYIEMRNSESPYMPRYGLLRNLRA---RELA: 211
      **
88 :NKYKLKEGNPSRDELTLRVAALAGRTCQALVVL-----SEWLPVTGTTMDGLSPAYPR: 142
SecB :HHH  EESSS  TT  EEHHHHHHHTHHHHHHHHHHS-----SSSSSS  HHHHHHSTT  G
      :  E1      E2H5      H6                -----                H7
  
```

DISCUSIÓN

```
      :      --      ----      H7
SecA :GG SS -- SSS---- TTHHHHHHHHHHH
212 :RYAFDFYE--VTSKT----PNRAREAIAQMCAA: 239
      *                *
143 :HMMHPSFAGMVDPSLPGDYLRAILDAAHSLYLLQF: 176
SecB :GGSSGGGGGT TTS HHHHHHHHHHHHHHHHH
      :                H8
```

Zona central de CP de WMV frente a zona central de NP de La Crosse:

SecA: Zona central de CP de WMV
SecB: Zona central de NP de La Crosse

[SIMILARITY] Seq 5.4 % Sec 63.4 % Exp 68.8 % CRMS 5.72 Å DRMS 4.42 Å
[SCORE] ScSSE 316.2 ScDis 25721.8 Rdis 12.2 (%) Rsse 10.6 (%)
[RELIABILITY] Superfamily 21.5 % Fold 46.2 %

```
      :H2      --      ----- E1      E2 H3      --      -
SecA :HHHHHHHHHHHHHS -- S -----SEEE STT EEE HHHHH--T SS S-
123 :MGVIMNGFMVWCIDNGT--SPDVN-----GVWVMDGEEQVEYPLKPIVE--NAKPTL-: 171
      *                *
31 :AESLNLAAVRIFFLNAAKAKAALSRKPERKANPKFGE---WQVEVINNHFPGNRNNPIG: 86
SecB :GGG HHHHHHHHHHHHHHHHHHTT SEE EETT----EE EE SGGGT
      :      H2                E1 E2 ----E3 E4

      :--H4      H5      ----- --      H6      --      -
SecA :--HHHHTS SHHHHHHHHHHH-----HS -- HHHHTT -- G-GGGG
      :--RQIMHHFSDAAEAYIEMRNS-----ESP--YMPRYGLLRNL--RAR-ELARYAF: 215
      * *                * *
87 :NNDLTIHRLSGYLARWVLDQYNENDDSQHELIRTTIINPIAESNGVGWDSGPEIYLSFF: 146
SecB :TT HHHHHHHHHHHHHHHHH HHHHHHHHH HHHHTTT GGG HHHHHTS
      :      H3                H4                H5                H6

      :      -      H7
SecA :SS - SSS TTHHHHHHHHHHH
216 :DFY-EVTSKTPNRAREAIAQMCAA: 238

147 :PGTEMFLETFFKYPLTIGIHRVKQ: 170
SecB :TTGGGGTTTTTTTHHHHHHHHHHT
      :                H7
```

Las figuras de mérito para el alineamiento 3D generadas por MATRAS son: Seq, identidad de secuencia (%); Sec, identidad de estructura secundaria (%); CRMS, desviación estándar de la media cuadrática (Å) de las posiciones de los átomos de los carbonos α de los residuos alineados; DRMS, desviación estándar de la media cuadrática (Å) de las posiciones de los átomos de los carbonos β de los residuos alineados; ScSSE, valor SSE (elementos de estructura secundaria); ScDis, valor de distancias entre modelos; Rdis, valor de distancias normalizado (%); Rsse, valor SSE normalizado; "Reliability" ("superfamily" y "fold"), indica la probabilidad de que el par de estructuras tengan el mismo plegamiento y formen parte de la misma familia de plegamiento en la base de datos SCOP 1.59.

ANEXO II

Anexo II: Matriz de distancias evolutivas basada en la similitud estructural entre NPs de varias familias de virus ssRNA.

	WMV	PepMV	RVFV	TSWV	LC	IA	RV	RSV	BV
WMV	-								
PepMV	34,44	-							
RVFV	47,54	48,34	-						
TSWV	58,90	59,02	56,78	-					
LC	59,04	60,36	57,94	32,09	-				
IA	51,53	57,00	60,83	63,58	65,10	-			
RV	59,99	61,66	61,63	68,71	70,73	63,72	-		
RSV	58,38	59,49	62,81	60,99	62,52	64,35	56,71	-	
BV	57,48	62,18	59,82	64,98	75,92	69,41	70,41	58,14	-

WMV, Virus del mosaico de la sandía; **PepMV**, Virus del mosaico del pepino; **RVFV**, Virus de la fiebre del valle del Rift; **TSWV**, Virus del bronceado del tomate; **LC**, Virus de La Crosse; **IA**, Virus Influenza A; **RV**, Virus de la rabia; **RSV**, Virus respiratorio sincitial; **BV**, Virus de la enfermedad de Borna.

Las distancias evolutivas entre los pares de NPs se calcularon en base a varios valores resultantes del alineamiento 3D de los modelos atómicos entre usando TM-align, en concreto el valor RMSD y el número de residuos topológicamente equivalentes, como se detalla en materiales y métodos.

ANEXO III

Anexo III: Secuencia proteína de la CP de MWV M116

1 SGKEAVENLDAGKESKKDTSGKGDKPQNSQTGQGSKEQTKTGTVSKDVNV 50
51 GSKGKEVPRLQKITKKMNLPTVGGKIILSLDHLLEYKPNQVDLFNTRATK 100
101 TQFESWYSAVKVEYDLNDEQMGVIMNGFMVWCIDNGTSPDVNGVWVMMDG 150
151 EEQVEYPLKPIVENAKPTLRQIMHHFSDAAEAYIEMRNSESPYMPRYGLL 200
201 RNLRDRELARYAFDFYEVTSKTPNRAREAIAQMKAALAGINSRLFGLDG 250
151 NISTNSENTERHTARDVNQNMHTLLGMGPPQ 281

BIBLIOGRAFÍA

A

- Abrescia, N. G., D. H. Bamford, J. M. Grimes, and D. I. Stuart. 2012. "Structure unifies the viral universe." *Annu Rev Biochem* 81:795-822. doi: 10.1146/annurev-biochem-060910-095130.
- Adams, M. J., J. F. Antoniw, and F. Beaudoin. 2005. "Overview and analysis of the polyprotein cleavage sites in the family Potyviridae." *Mol Plant Pathol* 6 (4):471-87. doi: 10.1111/j.1364-3703.2005.00296.x.
- Adams, M.J. 2008. *Flexiviruses. Encyclopedia of viruses (Third Edition)*. Elsevier.
- Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. 2010. "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Crystallogr D Biol Crystallogr* 66 (Pt 2):213-21. doi: 10.1107/S0907444909052925.
- Agarwal, G., M. Rajavel, B. Gopal, and N. Srinivasan. 2009. "Structure-based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold." *PLoS One* 4 (5):e5736. doi: 10.1371/journal.pone.0005736.
- Agirrezabala, X., E. Méndez-López, G. Lasso, M. A. Sánchez-Pina, M. Aranda, and M. Valle. 2015. "The near-atomic cryoEM structure of a flexible filamentous plant virus shows homology of its coat protein with nucleoproteins of animal viruses." *Elife* 4:e11795. doi: 10.7554/eLife.11795.
- Agrios, George N. 2005. *Plant Diseases Caused by Viruses*. Plant Pathology (Fifth edition). Elsevier Academic Press.
- Aiewsakun, P., and A. Katzourakis. 2015. "Endogenous viruses: Connecting recent and ancient viral evolution." *Virology* 479-480:26-37. doi: 10.1016/j.virol.2015.02.011.
- Anderson, P. K., A. A. Cunningham, N. G. Patel, F. J. Morales, P. R. Epstein, and P. Daszak. 2004. "Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers." *Trends Ecol Evol* 19 (10):535-44. doi: 10.1016/j.tree.2004.07.021.
- Anindya, R., and H. S. Savithri. 2003. "Surface-exposed amino- and carboxy-terminal residues are crucial for the initiation of assembly in Pepper vein banding virus: a flexuous rod-shaped virus." *Virology* 316 (2):325-36.
- Aranda, MA, and J Freitas-Astúa. 2017. Ecology and diversity of plant viruses, and epidemiology of plant virus-induced diseases. *Annals of applied biology*.
- Arranz, R., R. Coloma, F. J. Chichón, J. J. Conesa, J. L. Carrascosa, J. M. Valpuesta, J. Ortín, and J. Martín-Benito. 2012. "The structure of native influenza virion ribonucleoproteins." *Science* 338 (6114):1634-7. doi: 10.1126/science.1228172.
- Atreya, P. L., C. D. Atreya, and T. P. Pirone. 1991. "Amino acid substitutions in the coat protein result in loss of insect transmissibility of a plant virus." *Proc Natl Acad Sci U S A* 88 (17):7887-91.

B

- Baltimore, D. 1971. "Expression of animal virus genomes." *Bacteriol Rev* 35 (3):235-41.
- Bamford, D. H. 2003. "Do viruses form lineages across different domains of life?" *Res Microbiol* 154 (4):231-6. doi: 10.1016/S0923-2508(03)00065-2.
- Bartesaghi, A., C. Aguerrebere, V. Falconieri, S. Banerjee, L. A. Earl, X. Zhu, N. Grigorieff, J. L. S. Milne, G. Sapiro, X. Wu, and S. Subramaniam. 2018. "Atomic Resolution Cryo-EM Structure of β -Galactosidase." *Structure*. doi: 10.1016/j.str.2018.04.004.
- Bharat, T. A., J. D. Riches, L. Kolesnikova, S. Welsch, V. Krähling, N. Davey, M. L. Parsy, S. Becker, and J. A. Briggs. 2011. "Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells." *PLoS Biol* 9 (11):e1001196. doi: 10.1371/journal.pbio.1001196.
- Bloomer, A. C., J. N. Champness, G. Bricogne, R. Staden, and A. Klug. 1978. "Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits." *Nature* 276 (5686):362-8.
- Booth, T. F., M. J. Rabb, and D. R. Beniac. 2013. "How do filovirus filaments bend without breaking?" *Trends Microbiol* 21 (11):583-93. doi: 10.1016/j.tim.2013.08.001.
- Brister, J. R., D. Ako-Adjei, Y. Bao, and O. Blinkova. 2015. "NCBI viral genomes resource." *Nucleic Acids Res* 43 (Database issue):D571-7. doi: 10.1093/nar/gku1207.

C

- Caetano-Anollés, G., and D. Caetano-Anollés. 2003. "An evolutionarily structured universe of protein architecture." *Genome Res* 13 (7):1563-71. doi: 10.1101/gr.1161903.
- Caetano-Anollés, G., and A. Nasir. 2012. "Benefits of using molecular structure and abundance in phylogenomic analysis." *Front Genet* 3:172. doi: 10.3389/fgene.2012.00172.
- Cambra, M, N Capote, A Myrta, and G Llácer. 2006. *Plum Pox Virus* and the estimated cost associated with sharka disease. OEPP/EPPO. *Bulletin OEPP/EPPO Bulletin* 36.
- Cambra, M., M. T. Gorris, C. Marroquín, M. P. Román, A. Olmos, M. C. Martínez, A. H. de Mendoza, A. López, and L. Navarro. 2000. "Incidence and epidemiology of Citrus tristeza virus in the Valencian community of Spain." *Virus Res* 71 (1-2):85-95.
- Chapman, S., G. Hills, J. Watts, and D. Baulcombe. 1992. "Mutational analysis of the coat protein gene of potato virus X: effects on virion morphology and viral pathogenicity." *Virology* 191 (1):223-30.
- Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. 2010. "MolProbity: all-atom structure validation for macromolecular crystallography." *Acta Crystallogr D Biol Crystallogr* 66 (Pt 1):12-21. doi: 10.1107/S0907444909042073.
- Chenavas, S., L. F. Estrozi, A. Slama-Schwok, B. Delmas, C. Di Primo, F. Baudin, X. Li, T. Crépin, and R. W. Ruigrok. 2013. "Monomeric nucleoprotein of influenza A virus." *PLoS Pathog* 9 (3):e1003275. doi: 10.1371/journal.ppat.1003275.

- Chung, B. Y., W. A. Miller, J. F. Atkins, and A. E. Firth. 2008. "An overlapping essential gene in the Potyviridae." *Proc Natl Acad Sci U S A* 105 (15):5897-902. doi: 10.1073/pnas.0800468105.
- Clare, D. K., and E. V. Orlova. 2010. "4.6A Cryo-EM reconstruction of tobacco mosaic virus from images recorded at 300 keV on a 4k x 4k CCD camera." *J Struct Biol* 171 (3):303-8. doi: 10.1016/j.jsb.2010.06.011.
- Clare, D. K., E. V. Pechnikova, E. V. Skurat, V. V. Makarov, O. S. Sokolova, A. G. Solovyev, and E. V. Orlova. 2015. "Novel Inter-Subunit Contacts in Barley Stripe Mosaic Virus Revealed by Cryo-Electron Microscopy." *Structure* 23 (10):1815-1826. doi: 10.1016/j.str.2015.06.028.
- Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner. 2004. "WebLogo: a sequence logo generator." *Genome Res* 14 (6):1188-90. doi: 10.1101/gr.849004.

D

- Desfosses, A., R. Ciuffa, I. Gutsche, and C. Sachse. 2014. "SPRING - an image processing package for single-particle based helical reconstruction from electron cryomicrographs." *J Struct Biol* 185 (1):15-26. doi: 10.1016/j.jsb.2013.11.003.
- DiMaio, F., C. C. Chen, X. Yu, B. Frenz, Y. H. Hsu, N. S. Lin, and E. H. Egelman. 2015. "The molecular basis for flexibility in the flexible filamentous plant viruses." *Nat Struct Mol Biol* 22 (8):642-4. doi: 10.1038/nsmb.3054.
- Dolja, V. V., R. Haldeman, N. L. Robertson, W. G. Dougherty, and J. C. Carrington. 1994. "Distinct functions of capsid protein in assembly and movement of tobacco etch potyvirus in plants." *EMBO J* 13 (6):1482-91.
- Dolja, V. V., and E. V. Koonin. 2018. "Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer." *Virus Res* 244:36-52. doi: 10.1016/j.virusres.2017.10.020.
- Dolja, Valerian V., Vitaly P. Boyko, Alexey A. Agranovsky, and Eugene V. Koonin. 1991. Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: Two families with distinct patterns of sequence and probably structure conservation. *Virology*.
- Duff-Farrier, C. R., A. M. Bailey, N. Boonham, and G. D. Foster. 2015. "A pathogenicity determinant maps to the N-terminal coat protein region of the Pepino mosaic virus genome." *Mol Plant Pathol* 16 (3):308-15. doi: 10.1111/mpp.12184.

E

- Elliott, RM. 2008. Bunyaviruses: General Features. *Encyclopedia of Virology (Third Edition)*. Elsevier.
- Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. "Features and development of Coot." *Acta Crystallogr D Biol Crystallogr* 66 (Pt 4):486-501. doi: 10.1107/S0907444910007493.
- Eskelin, K., A. Hafrén, K. I. Rantalainen, and K. Mäkinen. 2011. "Potyviral VPg enhances viral RNA Translation and inhibits reporter mRNA translation in planta." *J Virol* 85 (17):9210-21. doi: 10.1128/JVI.00052-11.

F

- Fauquet, CM. 2008. *Taxonomy, Classification and Nomenclature of Viruses. Eyclopedia of Viruses (Third Edition)*. Elsevier.
- Felsenstein, J. 1989. PHYLIP-Phylogeny inference package (Version 3.2). Cladistics.
- Fogolari, F., A. Corazza, V. Yarra, A. Jalaru, P. Viglino, and G. Esposito. 2012. "Bluues: a program for the analysis of the electrostatic properties of proteins based on generalized Born radii." *BMC Bioinformatics* 13 Suppl 4:S18. doi: 10.1186/1471-2105-13-S4-S18.
- Forster, R. L., D. L. Beck, P. J. Guilford, D. M. Voot, C. J. Van Dolleweerd, and M. T. Andersen. 1992. "The coat protein of white clover mosaic potexvirus has a role in facilitating cell-to-cell transport in plants." *Virology* 191 (1):480-4.
- Forterre, P. 2010. "Defining life: the virus viewpoint." *Orig Life Evol Biosph* 40 (2):151-60. doi: 10.1007/s11084-010-9194-1.
- Forterre, P., and D. Prangishvili. 2009. "The origin of viruses." *Res Microbiol* 160 (7):466-72. doi: 10.1016/j.resmic.2009.07.008.
- Freiberg, A. N., M. B. Sherman, M. C. Morais, M. R. Holbrook, and S. J. Watowich. 2008. "Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography." *J Virol* 82 (21):10341-8. doi: 10.1128/JVI.01191-08.
- Fromm, S. A., T. A. Bharat, A. J. Jakobi, W. J. Hagen, and C. Sachse. 2015. "Seeing tobacco mosaic virus through direct electron detectors." *J Struct Biol* 189 (2):87-97. doi: 10.1016/j.jsb.2014.12.002.

G

- Ge, P., and Z. H. Zhou. 2011. "Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches." *Proc Natl Acad Sci U S A* 108 (23):9637-42. doi: 10.1073/pnas.1018104108.
- Geoghegan, J. L., S. Duchêne, and E. C. Holmes. 2017. "Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families." *PLoS Pathog* 13 (2):e1006215. doi: 10.1371/journal.ppat.1006215.
- Gergerich, Rose C., and Valerian V. Dolja. 2006. Introduction to Plant Viruses. the Invisible Foe. The Plant Health Instructor.
- Ghigo, E., J. Kartenbeck, P. Lien, L. Pelkmans, C. Capo, J. L. Mege, and D. Raoult. 2008. "Ameobal pathogen mimivirus infects macrophages through phagocytosis." *PLoS Pathog* 4 (6):e1000087. doi: 10.1371/journal.ppat.1000087.
- Gibbs, A. J., K. Ohshima, M. J. Phillips, and M. J. Gibbs. 2008. "The prehistory of potyviruses: their initial radiation was during the dawn of agriculture." *PLoS One* 3 (6):e2523. doi: 10.1371/journal.pone.0002523.
- Gough, J. 2005. "Convergent evolution of domain architectures (is rare)." *Bioinformatics* 21 (8):1464-71. doi: 10.1093/bioinformatics/bti204.

Gómez, P., AM Rodríguez-Hernández, B Moury, and MA Aranda. 2009. Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *European Journal of Plant Pathology*.

Gómez, P., R. Sempere, and M. A. Aranda. 2012. "Pepino mosaic virus and Tomato torrado virus: two emerging viruses affecting tomato crops in the Mediterranean basin." *Adv Virus Res* 84:505-32. doi: 10.1016/B978-0-12-394314-9.00014-2.

H

Harris, A., G. Cardone, D. C. Winkler, J. B. Heymann, M. Brecher, J. M. White, and A. C. Steven. 2006. "Influenza virus pleiomorphy characterized by cryoelectron tomography." *Proc Natl Acad Sci U S A* 103 (50):19123-7. doi: 10.1073/pnas.0607614103.

He, S., and S. H. W. Scheres. 2017. "Helical reconstruction in RELION." *J Struct Biol* 198 (3):163-176. doi: 10.1016/j.jsb.2017.02.003.

Heinlein, M. 2015. "Plant virus replication and movement." *Virology* 479-480:657-71. doi: 10.1016/j.virol.2015.01.025.

Hipper, C., V. Brault, V. Ziegler-Graff, and F. Revers. 2013. "Viral and cellular factors involved in Phloem transport of plant viruses." *Front Plant Sci* 4:154. doi: 10.3389/fpls.2013.00154.

Holm, L., and P. Rosenström. 2010. "Dali server: conservation mapping in 3D." *Nucleic Acids Res* 38 (Web Server issue):W545-9. doi: 10.1093/nar/gkq366.

Huiskonen, J. T., A. K. Overby, F. Weber, and K. Grunewald. 2009. "Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: evidence for GN-GC glycoprotein heterodimers." *J Virol* 83 (8):3762-9. doi: 10.1128/JVI.02483-08.

Hulo, C., E. de Castro, P. Masson, L. Bougueleret, A. Bairoch, I. Xenarios, and P. Le Mercier. 2011. "ViralZone: a knowledge resource to understand virus diversity." *Nucleic Acids Res* 39 (Database issue):D576-82. doi: 10.1093/nar/gkq901.

I

Illergård, K., D. H. Ardell, and A. Elofsson. 2009. "Structure is three to ten times more conserved than sequence--a study of structural response in protein cores." *Proteins* 77 (3):499-508. doi: 10.1002/prot.22458.

J

Jones, D. T., and D. Cozzetto. 2015. "DISOPRED3: precise disordered region predictions with annotated protein-binding activity." *Bioinformatics* 31 (6):857-63. doi: 10.1093/bioinformatics/btu744.

K

Karasev, A. V., V. P. Boyko, S. Gowda, O. V. Nikolaeva, M. E. Hilf, E. V. Koonin, C. L. Niblett, K. Cline, D. J. Gumpf, and R. F. Lee. 1995. "Complete sequence of the citrus tristeza virus RNA genome." *Virology* 208 (2):511-20.

- Katzourakis, A., and R. J. Gifford. 2010. "Endogenous viral elements in animal genomes." *PLoS Genet* 6 (11):e1001191. doi: 10.1371/journal.pgen.1001191.
- Kawabata, T. 2003. "MATRAS: A program for protein 3D structure comparison." *Nucleic Acids Res* 31 (13):3367-9.
- Kekarainen, T., H. Savilahti, and J. P. Valkonen. 2002. "Functional genomics on potato virus A: virus genome-wide map of sites essential for virus propagation." *Genome Res* 12 (4):584-94. doi: 10.1101/gr.220702.
- Kendall, A., M. McDonald, W. Bian, T. Bowles, S. C. Baumgarten, J. Shi, P. L. Stewart, E. Bullitt, D. Gore, T. C. Irving, W. M. Havens, S. A. Ghabrial, J. S. Wall, and G. Stubbs. 2008. "Structure of flexible filamentous plant viruses." *J Virol* 82 (19):9546-54. doi: 10.1128/JVI.00895-08.
- Kerlan, C, and Moury B. 2008. Potato Virus Y. Encyclopedia of viruses (Third Edition). Elsevier.
- Kim, D. E., D. Chivian, and D. Baker. 2004. "Protein structure prediction and analysis using the Robetta server." *Nucleic Acids Res* 32 (Web Server issue):W526-31. doi: 10.1093/nar/gkh468.
- King, A.M., M.J. Adams, E.B. Carstens, and E.J. Lefkowitz. 2012. *Virus Taxonomy. Ninth Report of the International Committee of Viruses.*: ICTV.
- Klug, A. 1999. "The tobacco mosaic virus particle: structure and assembly." *Philos Trans R Soc Lond B Biol Sci* 354 (1383):531-5. doi: 10.1098/rstb.1999.0404.
- Knott, G., and C. Genoud. 2013. "Is EM dead?" *J Cell Sci* 126 (Pt 20):4545-52. doi: 10.1242/jcs.124123.
- Koonin, E. V., and V. V. Dolja. 1993. "Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences." *Crit Rev Biochem Mol Biol* 28 (5):375-430. doi: 10.3109/10409239309078440.
- Koonin, E. V., V. V. Dolja, and M. Krupovic. 2015. "Origins and evolution of viruses of eukaryotes: The ultimate modularity." *Virology* 479-480:2-25. doi: 10.1016/j.virol.2015.02.039.
- Krissinel, E., and K. Henrick. 2007. "Inference of macromolecular assemblies from crystalline state." *J Mol Biol* 372 (3):774-97. doi: 10.1016/j.jmb.2007.05.022.
- Kucukelbir, A., F. J. Sigworth, and H. D. Tagare. 2014. "Quantifying the local resolution of cryo-EM density maps." *Nat Methods* 11 (1):63-5. doi: 10.1038/nmeth.2727.
- L**
- Lacroix, C., A. Jolles, E. W. Seabloom, A. G. Power, C. E. Mitchell, and E. T. Borer. 2014. "Non-random biodiversity loss underlies predictable increases in viral disease prevalence." *J R Soc Interface* 11 (92):20130947. doi: 10.1098/rsif.2013.0947.
- Lakshmi, B., M. Mishra, N. Srinivasan, and G. Archunan. 2015. "Structure-Based Phylogenetic Analysis of the Lipocalin Superfamily." *PLoS One* 10 (8):e0135507. doi: 10.1371/journal.pone.0135507.

- Lan, P., W. B. Yeh, C. W. Tsai, and N. S. Lin. 2010. "A unique glycine-rich motif at the N-terminal region of Bamboo mosaic virus coat protein is required for symptom expression." *Mol Plant Microbe Interact* 23 (7):903-14. doi: 10.1094/MPMI-23-7-0903.
- Lecoq, H, and C Desbiez. 2008. *Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus. Encyclopedia of Viruses (Third Edition)*. Elsevier.
- Lecours, K., M. H. Tremblay, M. E. Gagné, S. M. Gagné, and D. Leclerc. 2006. "Purification and biochemical characterization of a monomeric form of papaya mosaic potexvirus coat protein." *Protein Expr Purif* 47 (1):273-80. doi: 10.1016/j.pep.2005.10.013.
- Lefkowitz, E. J., D. M. Dempsey, R. C. Hendrickson, R. J. Orton, S. G. Siddell, and D. B. Smith. 2018. "Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV)." *Nucleic Acids Res* 46 (D1):D708-D717. doi: 10.1093/nar/gkx932.
- Li, C. X., M. Shi, J. H. Tian, X. D. Lin, Y. J. Kang, L. J. Chen, X. C. Qin, J. Xu, E. C. Holmes, and Y. Z. Zhang. 2015. "Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses." *Elife* 4. doi: 10.7554/eLife.05378.
- Li, X., P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, and Y. Cheng. 2013. "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM." *Nat Methods* 10 (6):584-90. doi: 10.1038/nmeth.2472.
- Lucas, W. J. 2006. "Plant viral movement proteins: agents for cell-to-cell trafficking of viral genomes." *Virology* 344 (1):169-84. doi: 10.1016/j.virol.2005.09.026.
- Lundin, D., A. M. Poole, B. M. Sjöberg, and M. Högbom. 2012. "Use of structural phylogenetic networks for classification of the ferritin-like superfamily." *J Biol Chem* 287 (24):20565-75. doi: 10.1074/jbc.M112.367458.
- López-Moya, J.J., and J.A. García. 2008. *Potyviruses. Encyclopedia of viruses (Third Edition)*. Elsevier.

M

- Maes, P., S. V. Alkhovsky, Y. Bào, M. Beer, M. Birkhead, T. Briese, M. J. Buchmeier, C. H. Calisher, R. N. Charrel, I. R. Choi, C. S. Clegg, J. C. de la Torre, E. Delwart, J. L. DeRisi, P. L. Di Bello, F. Di Serio, M. Digiaro, V. V. Dolja, C. Drosten, T. Z. Druciarek, J. Du, H. Ebihara, T. Elbeaino, R. C. Gergerich, A. N. Gillis, J. J. Gonzalez, A. L. Haenni, J. Hepojoki, U. Hetzel, T. Hồ, N. Hóng, R. K. Jain, P. Jansen van Vuren, Q. Jin, M. G. Jonson, S. Junglen, K. E. Keller, A. Kemp, A. Kipar, N. O. Kondov, E. V. Koonin, R. Kormelink, Y. Korzyukov, M. Krupovic, A. J. Lambert, A. G. Laney, M. LeBreton, I. S. Lukashevich, M. Marklewitz, W. Markotter, G. P. Martelli, R. R. Martin, N. Mielke-Ehret, H. P. Mühlbach, B. Navarro, T. F. F. Ng, M. R. T. Nunes, G. Palacios, J. T. Pawęska, C. J. Peters, A. Plyusnin, S. R. Radoshitzky, V. Romanowski, P. Salmenperä, M. S. Salvato, H. Sanfaçon, T. Sasaya, C. Schmaljohn, B. S. Schneider, Y. Shirako, S. Siddell, T. A. Sironen, M. D. Stenglein, N. Storm, H. Sudini, R. B. Tesh, I. E. Tzanetakis, M. Uppala, O. Vapalahti, N. Vasilakis, P. J. Walker, G. Wáng, L. Wáng, Y. Wáng, T. Wèi, M. R. Wiley, Y. I. Wolf, N. D. Wolfe, Z. Wú, W. Xú, L. Yang, Z. Yáng, S. D. Yeh, Y. Z. Zhāng, Y. Zhèng, X. Zhou, C. Zhū, F. Zirkel, and J. H. Kuhn. 2018. "Taxonomy of the family Arenaviridae

- and the order Bunyavirales: update 2018." *Arch Virol*. doi: 10.1007/s00705-018-3843-5.
- Martelli, G. P., M. J. Adams, J. F. Kreuze, and V. V. Dolja. 2007. "Family Flexiviridae: a case study in virion and genome plasticity." *Annu Rev Phytopathol* 45:73-100. doi: 10.1146/annurev.phyto.45.062806.094401.
- Marvin, D. A., L. C. Welsh, M. F. Symmons, W. R. Scott, and S. K. Straus. 2006. "Molecular structure of fd (f1, M13) filamentous bacteriophage refined with respect to X-ray fibre diffraction and solid-state NMR data supports specific models of phage assembly at the bacterial membrane." *J Mol Biol* 355 (2):294-309. doi: 10.1016/j.jmb.2005.10.048.
- McGuffin, L. J., K. Bryson, and D. T. Jones. 2000. "The PSIPRED protein structure prediction server." *Bioinformatics* 16 (4):404-5.
- Melgarejo, P, J García-Jiménez, MC Jordá, MM López, MF Andrés, and M Duran-Vila. 2010. *Patógenos de plantas descritos en España.*: Ministerio de Medio Ambiente y Medio Rural y Marino.
- Menéndez-Conejero, R., T. H. Nguyen, A. K. Singh, G. N. Condezo, R. E. Marschang, M. J. van Raaij, and C. San Martín. 2017. "Structure of a Reptilian Adenovirus Reveals a Phage Tailspike Fold Stabilizing a Vertebrate Virus Capsid." *Structure* 25 (10):1562-1573.e5. doi: 10.1016/j.str.2017.08.007.
- Merk, A., A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M. I. Davis, R. Pragani, M. B. Boxer, L. A. Earl, J. L. S. Milne, and S. Subramaniam. 2016. "Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery." *Cell* 165 (7):1698-1707. doi: 10.1016/j.cell.2016.05.040.
- Mindell, J. A., and N. Grigorieff. 2003. "Accurate determination of local defocus and specimen tilt in electron microscopy." *J Struct Biol* 142 (3):334-47.
- Modrow, Susanne, Dietrich Falke, Uwe Tryuen, and Hermann Schätzl. 2010. "Viral Proliferation and Replication." In *Molecular Virology*, 31-38. Springer Reference.
- Mäkinen, K., and A. Hafrén. 2014. "Intracellular coordination of potyviral RNA functions in infection." *Front Plant Sci* 5:110. doi: 10.3389/fpls.2014.00110.

N

- Namba, K., and G. Stubbs. 1986. "Structure of tobacco mosaic virus at 3.6 Å resolution: implications for assembly." *Science* 231 (4744):1401-6.
- Narayanan, K. B., and S. S. Han. 2017. "Helical plant viral nanoparticles-bioinspired synthesis of nanomaterials and nanostructures." *Bioinspir Biomim* 12 (3):031001. doi: 10.1088/1748-3190/aa6bfd.
- Nasir, A., and G. Caetano-Anollés. 2017. "Identification of Capsid/Coat Related Protein Folds and Their Utility for Virus Classification." *Front Microbiol* 8:380. doi: 10.3389/fmicb.2017.00380.

O

- Oerke, EC. 2006. Crop losses to pests. *The Journal of Agricultural Science*.

Orlova, E. V., and H. R. Saibil. 2011. "Structural analysis of macromolecular assemblies by electron microscopy." *Chem Rev* 111 (12):7710-48. doi: 10.1021/cr100353t.

Ozeki, J., M. Hashimoto, K. Komatsu, K. Maejima, M. Himeno, H. Senshu, T. Kawanishi, S. Kagiwada, Y. Yamaji, and S. Namba. 2009. "The N-terminal region of the *Plantago asiatica* mosaic virus coat protein is required for cell-to-cell movement but is dispensable for virion assembly." *Mol Plant Microbe Interact* 22 (6):677-85. doi: 10.1094/MPMI-22-6-0677.

P

Pagán, I., P. González-Jara, A. Moreno-Letelier, M. Rodelo-Urrego, A. Fraile, D. Piñero, and F. García-Arenal. 2012. "Effect of biodiversity changes in disease risk: exploring disease emergence in a plant-virus system." *PLoS Pathog* 8 (7):e1002796. doi: 10.1371/journal.ppat.1002796.

Penczek, P. A., R. A. Grassucci, and J. Frank. 1994. "The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles." *Ultramicroscopy* 53 (3):251-70.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. "UCSF Chimera--a visualization system for exploratory research and analysis." *J Comput Chem* 25 (13):1605-12. doi: 10.1002/jcc.20084.

Pintilie, G. D., J. Zhang, T. D. Goddard, W. Chiu, and D. C. Gossard. 2010. "Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions." *J Struct Biol* 170 (3):427-38. doi: 10.1016/j.jsb.2010.03.007.

Poranen, M. M., R. Daugelavicius, and D. H. Bamford. 2002. "Common principles in viral entry." *Annu Rev Microbiol* 56:521-38. doi: 10.1146/annurev.micro.56.012302.160643.

Prasad, B. V., and M. F. Schmid. 2012. "Principles of virus structural organization." *Adv Exp Med Biol* 726:17-47. doi: 10.1007/978-1-4614-0980-9_3.

R

Radermacher, M. 1988. "Three-dimensional reconstruction of single particles from random and nonrandom tilt series." *J Electron Microscop Tech* 9 (4):359-94. doi: 10.1002/jemt.1060090405.

Radermacher, M., T. Wagenknecht, A. Verschoor, and J. Frank. 1987. "Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*." *J Microsc* 146 (Pt 2):113-36.

Raymond, D. D., M. E. Piper, S. R. Gerrard, G. Skiniotis, and J. L. Smith. 2012. "Phleboviruses encapsidate their genomes by sequestering RNA bases." *Proc Natl Acad Sci U S A* 109 (47):19208-13. doi: 10.1073/pnas.1213553109.

Reddy, V. S., and J. E. Johnson. 2005. "Structure-derived insights into virus assembly." *Adv Virus Res* 64:45-68. doi: 10.1016/S0065-3527(05)64003-1.

Revers, F., and J. A. García. 2015a. "Molecular biology of potyviruses." *Adv Virus Res* 92:101-99. doi: 10.1016/bs.aivir.2014.11.006.

- Revers, Frédéric, and Juan Antonio García. 2015b. *Molecular Biology of Potyviruses. Advances in Virus Research*. Vol. 92: Elsevier.
- Rossmann, M. G. 2013. "Structure of viruses: a short history." *Q Rev Biophys* 46 (2):133-80. doi: 10.1017/S0033583513000012.
- Ruigrok, R. W., T. Crépin, and D. Kolakofsky. 2011. "Nucleoproteins and nucleocapsids of negative-strand RNA viruses." *Curr Opin Microbiol* 14 (4):504-10. doi: 10.1016/j.mib.2011.07.011.
- Rybicki, E. P. 2015. "A Top Ten list for economically important plant viruses." *Arch Virol* 160 (1):17-20. doi: 10.1007/s00705-014-2295-9.
- Rybicki, E. P., and G. Pietersen. 1999. "Plant virus disease problems in the developing world." *Adv Virus Res* 53:127-75.

S

- Sastry, KS. 2013. Impact of Virus and Viroid Diseases on Crop Yields. In: *Plant Virus and Viroids Diseases in the Tropics*.: Springer, Dordrecht.
- Scheres, S. H. 2010. "Classification of structural heterogeneity by maximum-likelihood methods." *Methods Enzymol* 482:295-320. doi: 10.1016/S0076-6879(10)82012-9.
- Scheres, S. H., and S. Chen. 2012. "Prevention of overfitting in cryo-EM structure determination." *Nat Methods* 9 (9):853-4. doi: 10.1038/nmeth.2115.
- Schneider, T. D., and R. M. Stephens. 1990. "Sequence logos: a new way to display consensus sequences." *Nucleic Acids Res* 18 (20):6097-100.
- Scholthof, K. B., S. Adkins, H. Czosnek, P. Palukaitis, E. Jacquot, T. Hohn, B. Hohn, K. Saunders, T. Candresse, P. Ahlquist, C. Hemenway, and G. D. Foster. 2011. "Top 10 plant viruses in molecular plant pathology." *Mol Plant Pathol* 12 (9):938-54. doi: 10.1111/j.1364-3703.2011.00752.x.
- Seo, J. K., M. S. Vo Phan, S. H. Kang, H. S. Choi, and K. H. Kim. 2013. "The charged residues in the surface-exposed C-terminus of the Soybean mosaic virus coat protein are critical for cell-to-cell movement." *Virology* 446 (1-2):95-101. doi: 10.1016/j.virol.2013.07.033.
- Shaikh, T. R., H. Gao, W. T. Baxter, F. J. Asturias, N. Boisset, A. Leith, and J. Frank. 2008. "SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs." *Nat Protoc* 3 (12):1941-74. doi: 10.1038/nprot.2008.156.
- Sherman, M. B., A. N. Freiberg, M. R. Holbrook, and S. J. Watowich. 2009. "Single-particle cryo-electron microscopy of Rift Valley fever virus." *Virology* 387 (1):11-5. doi: 10.1016/j.virol.2009.02.038.
- Shi, M., Y. Z. Zhang, and E. C. Holmes. 2018. "Meta-transcriptomics and the evolutionary biology of RNA viruses." *Virus Res* 243:83-90. doi: 10.1016/j.virusres.2017.10.016.
- Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, and D. G. Higgins. 2011. "Fast, scalable

generation of high-quality protein multiple sequence alignments using Clustal Omega." *Mol Syst Biol* 7:539. doi: 10.1038/msb.2011.75.

Simmonds, P., M. J. Adams, M. Benkő, M. Breitbart, J. R. Brister, E. B. Carstens, A. J. Davison, E. Delwart, A. E. Gorbalenya, B. Harrach, R. Hull, A. M. King, E. V. Koonin, M. Krupovic, J. H. Kuhn, E. J. Lefkowitz, M. L. Nibert, R. Orton, M. J. Roossinck, S. Sabanadzovic, M. B. Sullivan, C. A. Suttle, R. B. Tesh, R. A. van der Vlugt, A. Varsani, and F. M. Zerbini. 2017. "Consensus statement: Virus taxonomy in the age of metagenomics." *Nat Rev Microbiol* 15 (3):161-168. doi: 10.1038/nrmicro.2016.177.

Solovyev, A. G., and V. V. Makarov. 2016. "Helical capsids of plant viruses: architecture with structural lability." *J Gen Virol* 97 (8):1739-54. doi: 10.1099/jgv.0.000524.

Sorzano, C. O., S. Jonić, C. El-Bez, J. M. Carazo, S. De Carlo, P. Thévenaz, and M. Unser. 2004. "A multiresolution approach to orientation assignment in 3D electron microscopy of single particles." *J Struct Biol* 146 (3):381-92. doi: 10.1016/j.jsb.2004.01.006.

Stubbs, G., and A. Kendall. 2012. "Helical viruses." *Adv Exp Med Biol* 726:631-58. doi: 10.1007/978-1-4614-0980-9_28.

Stubbs, G., S. Warren, and K. Holmes. 1977. "Structure of RNA and RNA binding site in tobacco mosaic virus from 4-A map calculated from X-ray fibre diagrams." *Nature* 267 (5608):216-21.

Stubbs, T. M., and A. J. Te Velthuis. 2014. "The RNA-dependent RNA polymerase of the influenza A virus." *Future Virol* 9 (9):863-876. doi: 10.2217/fvl.14.66.

T

Tang, G., L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke. 2007. "EMAN2: an extensible image processing suite for electron microscopy." *J Struct Biol* 157 (1):38-46. doi: 10.1016/j.jsb.2006.05.009.

Tanne, E., and I. Sela. 2005. "Occurrence of a DNA sequence of a non-retro RNA virus in a host plant genome and its expression: evidence for recombination between viral and host RNAs." *Virology* 332 (2):614-22. doi: 10.1016/j.virol.2004.11.007.

Thézé, J., S. Leclercq, B. Moumen, R. Cordaux, and C. Gilbert. 2014. "Remarkable diversity of endogenous viruses in a crustacean genome." *Genome Biol Evol* 6 (8):2129-40. doi: 10.1093/gbe/evu163.

Tollin, P., H. R. Wilson, I. M. Roberts, and A. F. Murrant. 1992. "Diffraction studies of the particles of two closteroviruses: heracleum latent virus and heracleum virus 6." *J Gen Virol* 73 (Pt 11):3045-8. doi: 10.1099/0022-1317-73-11-3045.

U

Urcuqui-Inchima, S., A. L. Haenni, and F. Bernardi. 2001. "Potyvirus proteins: a wealth of functions." *Virus Res* 74 (1-2):157-75.

Uversky, V. N. 2011. "Intrinsically disordered proteins from A to Z." *Int J Biochem Cell Biol* 43 (8):1090-103. doi: 10.1016/j.biocel.2011.04.001.

V

Van Heel, M. 1987. "Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction." *Ultramicroscopy* 21 (2):111-23.

Varma, A., A. J. Gibbs, R. D. Woods, and J. T. Finch. 1968. "Some observations on the structure of the filamentous particles of several plant viruses." *J Gen Virol* 2 (1):107-14. doi: 10.1099/0022-1317-2-1-107.

W

Wei, T., T. S. Huang, J. McNeil, J. F. Laliberté, J. Hong, R. S. Nelson, and A. Wang. 2010. "Sequential recruitment of the endoplasmic reticulum and chloroplasts for plant potyvirus replication." *J Virol* 84 (2):799-809. doi: 10.1128/JVI.01824-09.

Wylie, S. J., M. Adams, C. Chalam, J. Kreuze, J. J. López-Moya, K. Ohshima, S. Praveen, F. Rabenstein, D. Stenger, A. Wang, F. M. Zerbini, and Ictv Report Consortium. 2017. "ICTV Virus Taxonomy Profile: Potyviridae." *J Gen Virol* 98 (3):352-354. doi: 10.1099/jgv.0.000740.

Y

Yang, S., T. Wang, J. Bohon, M. Gagné, M. Bolduc, D. Leclerc, and H. Li. 2012. "Crystal structure of the coat protein of the flexible filamentous papaya mosaic virus." *J Mol Biol* 422 (2):263-73. doi: 10.1016/j.jmb.2012.05.032.

Ye, Q., R. M. Krug, and Y. J. Tao. 2006. "The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA." *Nature* 444 (7122):1078-82. doi: 10.1038/nature05379.

Z

Zayakina, O., M. Arkhipenko, S. Kozlovsky, N. Nikitin, A. Smirnov, P. Susi, N. Rodionova, O. Karpova, and J. Atabekov. 2008. "Mutagenic analysis of potato virus X movement protein (TGBp1) and the coat protein (CP): in vitro TGBp1-CP binding and viral RNA translation activation." *Mol Plant Pathol* 9 (1):37-44. doi: 10.1111/j.1364-3703.2007.00445.x.

Zhang, Y. 2008. "I-TASSER server for protein 3D structure prediction." *BMC Bioinformatics* 9:40. doi: 10.1186/1471-2105-9-40.

Zhang, Y., and J. Skolnick. 2005. "TM-align: a protein structure alignment algorithm based on the TM-score." *Nucleic Acids Res* 33 (7):2302-9. doi: 10.1093/nar/gki524.

PUBLICACIONES

PLANT VIRUSES

Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses

Miguel Zamora,¹ Eduardo Méndez-López,² Xabier Agirrezabala,¹ Rebeca Cuesta,¹ José L. Lavín,¹ M. Amelia Sánchez-Pina,² Miguel A. Aranda,² Mikel Valle^{1*}

Potyvirus constitute the second largest genus of plant viruses and cause important economic losses in a large variety of crops; however, the atomic structure of their particles remains unknown. Infective potyvirus virions are long flexuous filaments where coat protein (CP) subunits assemble in helical mode bound to a monopartite positive-sense single-stranded RNA [(+)ssRNA] genome. We present the cryo-electron microscopy (cryoEM) structure of the potyvirus watermelon mosaic virus at a resolution of 4.0 Å. The atomic model shows a conserved fold for the CPs of flexible filamentous plant viruses, including a universally conserved RNA binding pocket, which is a potential target for antiviral compounds. This conserved fold of the CP is widely distributed in eukaryotic viruses and is also shared by nucleoproteins of enveloped viruses with segmented (–)ssRNA (negative-sense ssRNA) genomes, including influenza viruses.

INTRODUCTION

Flexible filamentous viruses constitute one of the largest groups of plant pathogens and comprise more than 380 species [ICTV (International Committee on Taxonomy of Viruses) 2016 release] distributed in four families (Alphaflexiviridae, Betaflexiviridae, Closteroviridae, and Potyviridae). Their virions are nonenveloped and flexible rod-shaped particles where multiple copies of the coat protein (CP) are arranged in helical symmetry while bound to a monopartite positive-sense single-stranded RNA [(+)ssRNA] (1). The genomic structure of these viruses differs significantly between families, and their CPs display low sequence similarity; however, their overall virion architecture (at least at a low resolution) is very similar (1). Hence, it is thought that CPs from flexuous rod-shaped plant viruses share a common, putatively homologous fold and have a common evolutionary origin (2). Atomic structures for CPs are only available for several potexviruses (genus *Potexvirus*, family Alphaflexiviridae) (3–5), and the putative structural homology between groups is uncertain. The genus *Potyvirus* (family Potyviridae) is the largest group of flexible filamentous viruses and comprises nearly 200 species, about 15% of all known plant viruses (6). They are transmitted horizontally, mainly by aphids, and some species can spread vertically to the progeny through seeds (7). Potyviruses are widely distributed, infect a large variety of angiosperms in tropical and subtropical regions, and cause large economic losses in crops. Watermelon mosaic virus (WMV) is a potyvirus with one of the widest host ranges and is present worldwide. It infects more than 170 different plant species from 26 families, including cucurbits and legumes (8). WMV infections induce mosaics, deformations, and discoloration of leaves and fruits and cause serious yield and quality losses in infected crops.

RESULTS AND DISCUSSION

By using cryo-electron microscopy (cryoEM), we have structurally characterized WMV virions isolated from infected squash plants. Following single particle-based helical image processing in RELION (9), we determined a three-dimensional (3D) density map at a resolu-

tion of 4.0 Å, which was then used to calculate an atomic model for both CP and ssRNA (Fig. 1 and fig. S1). The helical structure of WMV displays a left-handed helix with a diameter of 130 Å, contains about 8.8 CP subunits per turn and a pitch of 35.2 Å (Fig. 1, A and B), and is very similar to the helical structure of other flexuous plant viruses (1, 3, 4, 10). The WMV CP exhibits a core domain rich in α helices and two long N- and C-terminal arms (Fig. 1C). The atomic structure lacks the first 59 amino acids at the N-terminal region because there is no density attributable to this part in the cryoEM map (Fig. 1, A to C). This region of the polypeptide faces the solvent and seems to be flexible. The atomic structure also misses the last 17 residues at the C terminus. This fraction of the CP is at the inner side of the filament, but the signal in the cryoEM map is scattered (Fig. 1B), and the polypeptide cannot be traced with accuracy, suggesting the presence of a disordered short segment at the C-terminal end. As shown before for potexviruses (3, 4), the flexible N- and C-terminal arms mediate the polymerization of CP subunits in the virion (Fig. 1D), allow for relative movements between CPs, and explain the flexuous nature of the viral particles. However, whereas in potexviruses, the N-terminal arm only supports side-by-side contacts between CPs (3, 4), in WMV, each N-terminal region bridges the adjacent subunit and, by a 90° turn, reaches a second subunit in the next turn of the helix (Fig. 1, A and D). This interaction between subunits at different turns is favored by a complementary surface potential between an electropositive N-terminal arm and a polymerization groove flanked by electronegative regions (Fig. 1D). Thus, in WMV CP, the N-terminal arm has a dual role of supporting side-by-side and longitudinal polymerization for helical construction.

High-resolution density for the CP allows for accurate atomic modeling and shows that the WMV CP shares the same fold with the reported CPs from potexviruses (fig. S2) (3–5). Nonetheless, the density for the ssRNA is the average of RNA segments of different composition along the virions. Thus, the atomic model for the ssRNA cannot be traced unambiguously. The current RNA polarity is opposite to the one calculated for the potexvirus Pepino mosaic virus (PepMV) (4). However, the RNA binding mode in WMV is very similar to the one observed previously for PepMV virion, and both share key characteristics (4). Each WMV CP subunit covers five nucleotides of the ssRNA (Fig. 2A). Whereas four of the nucleotides orient their bases toward the inner side of the virion (Fig. 2B), one of the nucleotides goes deep into an RNA binding pocket (nucleotide position labeled as U₄ in Fig. 2).

¹Molecular Recognition and Host-Pathogen Interactions, Center for Cooperative Research in Biosciences, CIC bioGUNE, 48160 Derio, Spain. ²Centro de Edafología y Biología Aplicada del Segura (CEBAS), Consejo Superior de Investigaciones Científicas (CSIC), Espinardo, 30100 Murcia, Spain.

*Corresponding author. Email: mvalle@cicbiogune.es

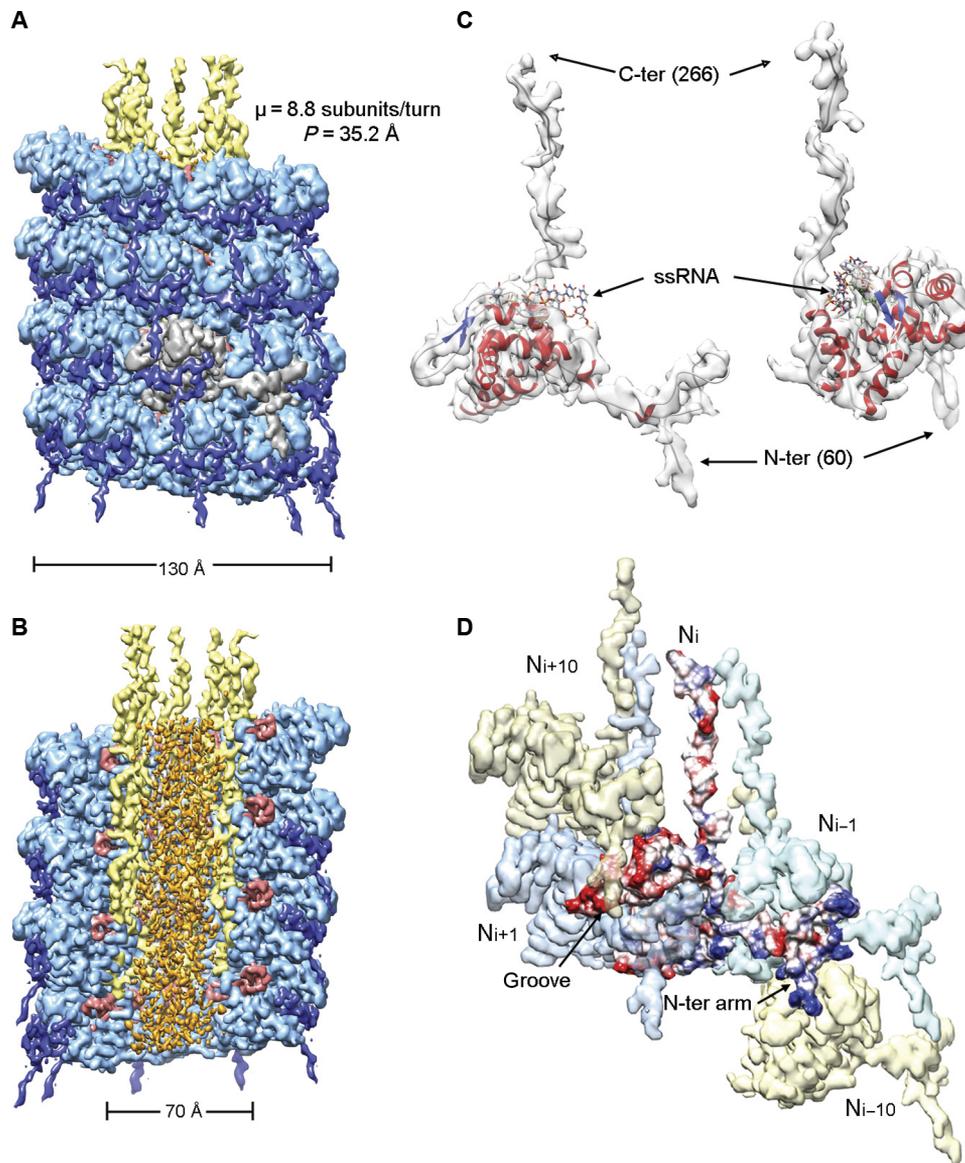


Fig. 1. Near-atomic cryoEM structure of WMV virions. (A and B) Renderings of the 3D cryoEM map [cutaway mode in (B)] calculated for WMV. Segmented densities are depicted as follows: WMV CP core region (light blue), N-terminal arm (dark blue), C-terminal arm (yellow), scattered densities at the inner side of the helix (orange), and density for the ssRNA (red). One of the CP subunits is seen as gray in (A). (C) Two views of the atomic model calculated for WMV CP within the semitransparent density for a subunit segmented from the cryoEM map. (D) Segmented densities for WMV CP subunits are depicted. The subunit N_i is colored on the basis of the calculated electrostatic surface potential.

Essentially, the interactions between Ser¹⁴⁰ and Arg¹⁷², with consecutive phosphate groups in the RNA (Fig. 2B), hold the flanked nucleoside deep inside this pocket, where its base is surrounded by Arg¹⁷², Asp²¹⁶, and Lys²³⁶ of WMV CP (Fig. 2C).

These four residues are located at the same position in the atomic structure of PepMV CP (4), suggesting the presence of a conserved RNA binding site. Despite the low sequence similarity among CPs from different flexuous rod-shaped plant viruses, the Ser, Arg, and Asp residues are strictly conserved throughout the four families (Fig. 3), and the available atomic structures place them in exactly the same positions within this RNA binding pocket (Fig. 3, A to C). In the consensus sequences of CPs from different families (Fig. 3D), it is also noted that the distance between invariant Arg and Asp is kept constant, and in WMV

CP, this segment (R172↔D216 in Fig. 2D) runs next to the RNA binding site. However, there is a large variability in the number of residues separating conserved Ser and Arg residues. In WMV CP, this fragment of the protein (S140↔R172 in Fig. 2D) protrudes away from the RNA site and faces the solvent, and hence, it could accept the insertion of several residues without breaking the protein-RNA interaction. Although the presence of these three conserved amino acids in flexible filamentous plant viruses was described earlier (11), the lack of structural information precluded the identification of the conserved RNA binding site. Thus, even in the absence of structural data for CPs other than the few potexviruses already mentioned and of the current potyvirus, the conservation of key residues points to a universal fold for all these CPs, a structure that includes three invariant amino acids that

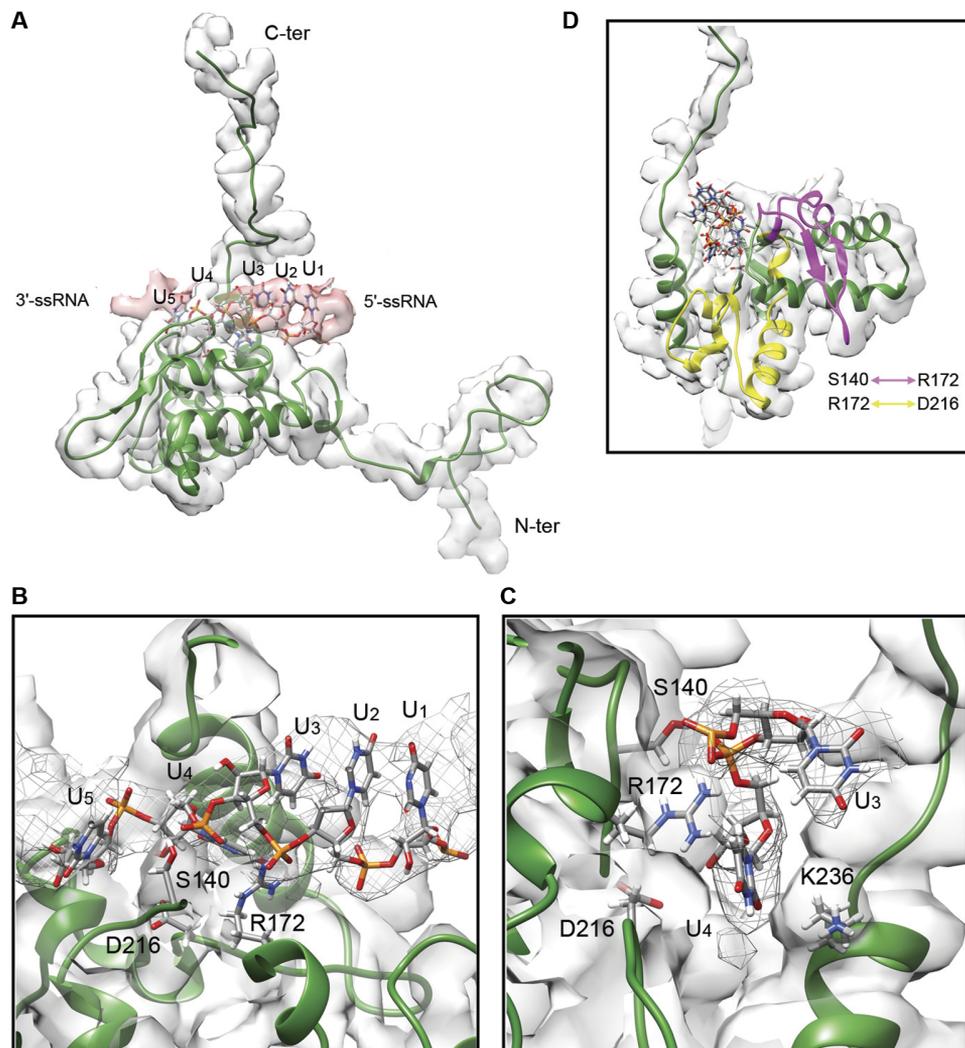


Fig. 2. Protein-ssRNA interactions in WMV. (A) Atomic models for WMV CP and ssRNA (ssRNA modeled as a polyU) are seen inside the densities for protein (gray) and RNA (red). (B and C) Closeup views of the RNA binding pocket within WMV CP. The CP is seen in ribbons, and some of the amino acids that interact with the ssRNA are displayed. Density for the RNA is rendered in gray mesh. (D) The ribbon representation for WMV CP is seen as green, but the segments of the protein linking the amino acids that bind to RNA are highlighted in magenta (from Ser¹⁴⁰ to Arg¹⁷²) and yellow (from Arg¹⁷² to Asp²¹⁶).

bind directly to the genomic ssRNA. From a functional perspective, mutations in the conserved Arg and Asp residues of the CP are known to impair *in vitro* assembly of the potyvirus Johnsongrass mosaic virus (12) and to block the assembly and cell-to-cell movement of the potyvirus PepMV in plants (4). Currently, apart from containment measures, chemical vector control, and use of resistant crop varieties, there is no available treatment to eradicate this group of plant pathogens. The conserved RNA binding site in CPs from flexible filamentous viruses is a clear target for the search of antiviral compounds that might interfere with the assembly and genome packaging of a large number of economically relevant plant viruses.

Apart from a conserved fold for the CPs in flexuous filamentous plant viruses, we searched for structural homologs for the core region of WMV CP (excluding the extended N- and C-terminal regions) using the web servers Dali (13) and Matras (14). The hits with higher scores by both methods are indeed the CPs from potexviruses (Fig. 4 and figs. S3 and S4), but additional homologs are detected in nucleoproteins (NPs) from members of Orthomyxoviridae and Bunyaviridae families, including all

phleboviruses with known NP structure, such as Rift Valley fever virus, influenza virus, La Crosse virus, and Tomato spotted wilt virus (Fig. 4). The bona fide structural homology along all those proteins is based on (i) the TM-alignment scores (fig. S3) (15) that weigh their topology resemblance, (ii) the high probability of belonging to the same fold family, as calculated by Matras (fig. S4) (14), (iii) the identical location of the N- and C-terminal ends, (iv) the nature of viral proteins that bind genomic ssRNA, and (v) the position of the ssRNA binding groove. For influenza virus (family Orthomyxoviridae), there is no NP-RNA atomic structure, but the proposed groove for ssRNA binding (16) is at a comparable location (indicated by an arrow in Fig. 4). The structural similarity between CPs from potexviruses and NPs from phleboviruses was described previously (4), but it was limited to these two groups. Our current results greatly expand the universe of this structural homology and suggests that the CPs from flexible filamentous plant viruses, non-enveloped and with monopartite (+)ssRNA genomes, share a common evolutionary origin with NPs from several enveloped viruses with segmented (–)ssRNA (negative-sense ssRNA) that infect animals and

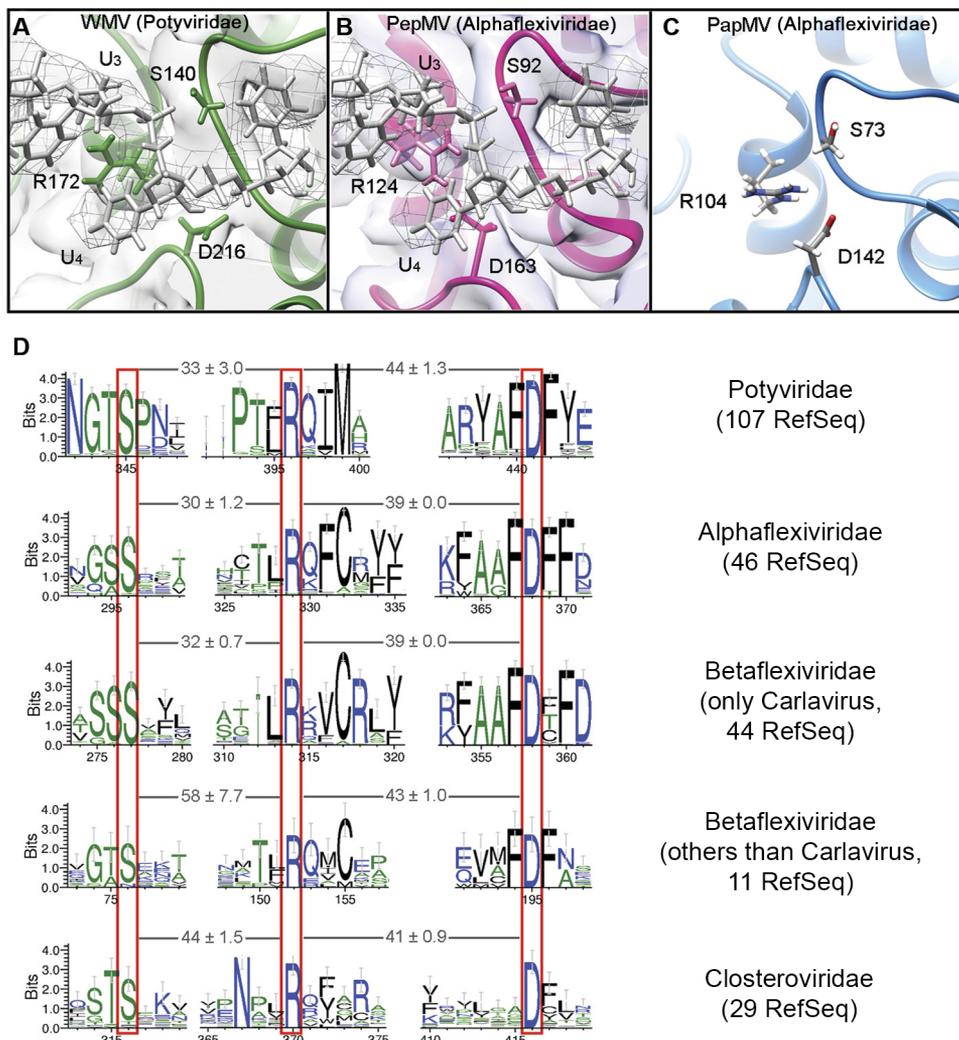


Fig. 3. Conserved RNA binding pocket in flexible filamentous plant viruses. (A to C) Close-up views of the RNA binding pockets of the CPs from WMV (A), PepMV [Protein Data Bank (PDB) code: 5FN1 (4)] (B), and papaya mosaic virus (PapMV) [PDB code: 4DOX (5)] (C). For clarity, (A) and (B) show the ssRNA model calculated for WMV virion. The structure for PapMV CP was calculated for isolated and RNA-free protein by crystallography (5). (D) Consensus sequence logos for the CPs from different families of flexuous filamentous plant viruses. The distance was measured by the number of residues (average and SD) between invariant amino acids (Ser, Arg, and Asp), indicated by gray lines and numbers. The number of reference sequences (36) aligned in each family is also indicated. The color scheme of amino acid symbols is as follows: blue for hydrophilic, green for neutral, and black for hydrophobic.

plants. Overall, these are viral proteins that protect the genome of ssRNA viruses that infect eukaryotes, although the structural design of their infective particles differs significantly (fig. S5). The members of family Bunyaviridae exhibit loose ribonucleoproteins (RNPs) protected inside a membrane that contains inserted viral glycoproteins arranged in an icosahedral fashion (17, 18). On the other hand, in influenza virus, the genomic material is arranged in double-helical RNPs with two NP strands of opposite polarity, and these RNPs are inside a pleomorphic lipid envelope decorated by several viral proteins (19). These two different designs for bunyaviruses and orthomyxoviruses also differ from the helical naked filaments of flexuous plant viruses; nevertheless, the proteins that cover their ssRNA genomes in all those settings display similar folds. Notably, no structural homology has been previously described between NPs from influenza and bunyaviruses, and this has only emerged when atomic structures for CPs guide the structural comparison. This suggests that the structure of CPs from flexible rod-shaped viruses conserves a fold closer to a common ancestor protein where the

homology with the other two groups can still be recognized. On the contrary, NPs from the groups of enveloped viruses with segmented RNA genomes have strongly diverged. Regardless of the evolutionary origin of this protein fold and of the mechanisms that transferred it to different viral groups, it has widely spread along eukaryotic ssRNA viruses.

MATERIALS AND METHODS

WMV inoculation and purification

Carborundum-dusted cotyledons of squash plants were inoculated with an homogenate consisting of a dried material from WMV-M116-infected squash leaves ground in 30 mM sodium phosphate (pH 8.0) (20). The inoculated plants were grown in a greenhouse (16-hour photoperiod; 25°/18°C day/night) for 4 weeks. Young systemically infected leaves were harvested, and WMV particles were purified following a previously described method by Moreno *et al.* (21) with some modifications. Infected leaves (60 g) were ground in liquid nitrogen and

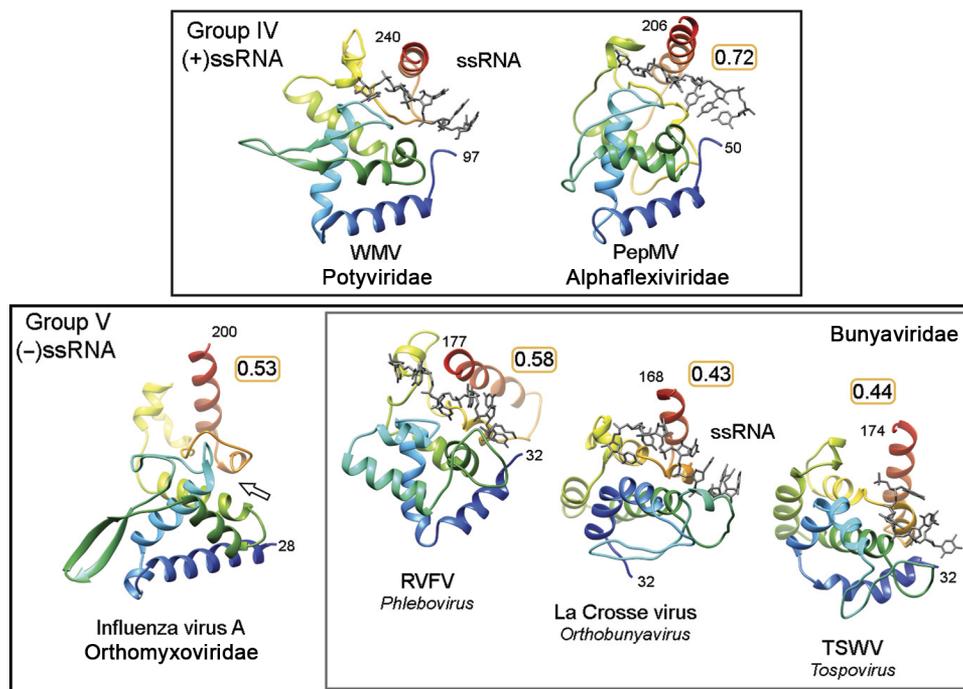


Fig. 4. Conserved fold in eukaryotic ssRNA viruses. Ribbon representations for the CP and NP regions showing structural homology. The regions of the proteins are depicted in rainbow colors (from N- to C-terminal ends) and are seen together with bound ssRNA (whenever available). TM-scores (15) for the structural alignments between WMV CP and each of the atomic structures are seen inside the orange boxes. Values of TM-scores around 0.5 are indicative of proteins with the same fold. Depicted structures are as follows: WMV CP and ssRNA from the current work, PepMV CP and ssRNA (PDB code: 5FN1) (4), influenza virus A NP (PDB code: 3ZDP) (40), Rift Valley fever virus (RVFV) NP in complex with ssRNA (PDB code: 4H5O) (41), La Crosse virus NP and ssRNA (PDB code: 4BHH) (42), and Tomato spotted wilt virus (TSWV) NP in complex with ssRNA (PDB code: 5IP2) (43).

homogenized in a buffer containing 0.5 M K_2HPO_4 (pH 7.5), 5 mM EDTA, 10 mM diethyldithiocarbamic acid, and 20 mM Na_2SO_3 (5 ml/g). The mixture was stirred for 15 min at 4°C and centrifuged for 10 min at 7500g. The supernatant was filtered through five layers of cheesecloth and stirred for 1 hour at 4°C after adding Triton X-100 [1% (v/v)]. Virions were precipitated by ultracentrifugation for 90 min at 300,000g, and the pellet was resuspended for 2 hours with constant stirring at 4°C in buffer A [50 mM sodium citrate and 20 mM Na_2SO_3 (pH adjusted at 7.5 with 0.25 M citric acid)] with 1% Triton X-100. The solution was centrifuged at low speed to clarify the preparation. The supernatant was mixed with chloroform [10% (v/v)] and centrifuged for 10 min at 15,000g. The aqueous phase was overlaid onto 5 ml of a 30% sucrose cushion and centrifuged for 110 min at 245,000g. The pellet was resuspended overnight in buffer A with constant stirring at 4°C. After a low-speed centrifugation, Cs_2SO_4 (0.26 g/ml) was added to the supernatant, and a density gradient was formed by ultracentrifugation for 30 hours at 28,000 rpm in an SW 28 rotor. The WMV particles concentrated in an opaque band were collected using a syringe. The volume recovered was then diluted and ultracentrifuged for 90 min at 108,000g. The pellet containing the purified virions was resuspended in 200 μ l of buffer A and stored at 4°C.

CryoEM and image processing

The WMV particle solution was applied to Quantifoil R2/2 holey carbon grids covered with a thin layer of carbon followed by grid vitrification in FEI Vitrobot. Data collection was carried out in a Titan Krios FEI electron microscope operated at 300 kV by a K2 direct detector (GATAN). Movie frame images were taken at a nominal magnification of $\times 130,000$

resulting in a sampling of 1.1 $\text{\AA}/\text{pixel}$. Exposures of 5 s in electron counting mode resulted in images with 35 frames and a total dose of 35 $e^-/\text{\AA}^2$. Motion between frames was corrected in micrographs (22) using frames 2 to 27, resulting in an accumulative dose of 27 $e^-/\text{\AA}^2$. The contrast transfer function of the micrographs was estimated using CTFFIND3 (23). An initial set of 915 helices was selected in EMAN2 (24) and processed in SPRING software (25), following the single particle-based helical reconstruction scheme. These manually selected helices were used for a preliminary estimation of the helical symmetry parameters inherent to WMV. From the same micrographs, a second data set was automatically selected in Relion2 (9), resulting in 50,045 segments (boxes of 230 pixels \times 230 pixels, with a step of 29 pixels between segments). The cryoEM density map was also calculated in Relion2. 2D classification rounds and particle sorting allowed us to isolate a set of 39,100 particles corresponding to good-quality filament segments. Starting with a cylinder as a reference map, the image processing of this new set of images yielded a refined cryoEM map at a resolution of 4.0 \AA (fig. S1). RELION local optimization of twist and rise was carried out during 3D refinement setting the symmetry parameters at -40.87° for helical twist and 3.99 \AA for helical rise. 3D classification with angular local search (1.8°) and local searches of symmetry resulted in three classes with different symmetry values. A predominant class (93% of the particles) revealed the same symmetry as that of the previous local optimization of twist and rise results, indicating no relevant heterogeneity in the helical structure among the filament population. Local resolution variability was estimated using Relion2 and ResMap (26), revealing resolutions ranging from 3.4 to 4.6 \AA (fig. S1). Helical symmetry in real space was imposed to the final map to obtain homogeneity among the asymmetric units for molecular modeling.

Atomic model building for WMV CP

An initial atomic model for WMV CP was obtained in Robetta server (<http://rosetta.bakerlab.org>) (27) by comparative modeling based on the PepMV CP structure (PDB code: 5FN1) (4). Density for a single WMV CP subunit was isolated from the cryoEM map by segmentation using the Segger method (28) in Chimera (29), which was also used for rigid-body fitting of the preliminary atomic model within the cutoff density. Further modeling of the WMV CP atomic structure was carried out manually using Coot (30), guided by the cryoEM density, and the stereochemistry of the model was improved by real-space refinement in Phenix (31). After several iterative cycles of model building and optimization, a final refinement in Phenix using noncrystallographic symmetry improved the interfaces between adjacent subunits. For the ssRNA, a modeled polyU was included in the refinement. The final atomic structure was validated in MolProbity (32). The surface electrostatic potential of the atomic structure was calculated on the basis of generalized Born radii in Bluues server (<http://protein.bio.unipd.it/bluues>) (33, 34). To search for structural homologs of the calculated WMV CP atomic model, we used the Dali (http://ekhidna.biocenter.helsinki.fi/dali_server) (13) and Matras (<http://strcomp.protein.osaka-u.ac.jp/matras>) (14, 35) servers, and the similarities were further evaluated by the TM-score calculated using the TM-align algorithm (<http://zhanglab.ccmb.med.umich.edu/TM-align>) (15).

Protein sequence alignments

Protein sequences for the CPs of flexuous filamentous plant viruses were retrieved from the Reference Sequence (RefSeq) database at the National Center for Biotechnology Information (36). The data set includes 237 sequences for the CPs from four families: Alphaflexiviridae (46), Betaflexiviridae (55), Closteroviridae (29), and Potyviridae (107). Multiple sequence alignments were performed within each family using the Clustal Omega (37) web server (www.ebi.ac.uk/Tools/msa/clustalo) on default parameters. The resulting multi-FASTA files provided the consensus sequences displayed as sequence logos using the WebLogo web-based application (<http://weblogo.berkeley.edu>) (38, 39). For the family Betaflexiviridae, the alignment of sequences from genus *Carlavirus* was done independently from the rest of the genera because of the large difference observed in the distance between conserved Ser and Arg residues at the putative RNA binding site. The only exceptions found among the 237 sequences analyzed are the CPs from potexviruses Bamboo mosaic virus and Foxtail mosaic virus, where the conserved Arg is substituted by His, a conservative mutation between basic amino acids.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/3/9/eaa02182/DC1>

fig. S1. CryoEM and atomic model for WMV.

fig. S2. Structural comparison between potyvirus and potexvirus.

fig. S3. Representation of the structural alignment between regions of CPs and NPs from several ssRNA viruses.

fig. S4. Output from the 3D library search using the WMV CP core region as target.

fig. S5. Morphological universe of ssRNA viruses with structural homology between their CPs/NPs.

REFERENCES AND NOTES

1. A. Kendall, M. McDonald, W. Bian, T. Bowles, S. C. Baumgarten, J. Shi, P. L. Stewart, E. Bullitt, D. Gore, T. C. Irving, W. M. Havens, S. A. Ghahrial, J. S. Wall, G. Stubbs, Structure of flexible filamentous plant viruses. *J. Virol.* **82**, 9546–9554 (2008).
2. E. V. Koonin, V. V. Dolja, M. Krupovic, Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **479–480**, 2–25 (2015).
3. F. DiMaio, C.-C. Chen, X. Yu, B. Frenz, Y.-H. Hsu, N.-S. Lin, E. H. Egelman, The molecular basis for flexibility in the flexible filamentous plant viruses. *Nat. Struct. Mol. Biol.* **22**, 642–644 (2015).
4. X. Agirrezabala, E. Méndez-López, G. Lasso, M. A. Sánchez-Pina, M. Aranda, M. Valle, The near-atomic cryoEM structure of a flexible filamentous plant virus shows homology of its coat protein with nucleoproteins of animal viruses. *eLife* **4**, e11795 (2015).
5. S. Yang, T. Wang, J. Bohon, M.-È. L. Gagné, M. Bolduc, D. Leclerc, H. Li, Crystal structure of the coat protein of the flexible filamentous papaya mosaic virus. *J. Mol. Biol.* **422**, 263–273 (2012).
6. S. J. Wylie, M. Adams, C. Chalam, J. Kreuze, J. J. López-Moya, K. Ohshima, S. Praveen, F. Rabenstein, D. Stenger, A. Wang, F. M. Zerbini; ICTV Report Consortium, ICTV Virus Taxonomy Profile: Potyviridae. *J. Gen. Virol.* **98**, 352–354 (2017).
7. F. Revers, J. A. García, Molecular biology of potyviruses. *Adv. Virus Res.* **92**, 101–199 (2015).
8. H. Lecoq, C. Desbiez, Viruses of cucurbit crops in the Mediterranean region: An ever-changing picture. *Adv. Virus Res.* **84**, 67–126 (2012).
9. S. He, S. H. W. Scheres, Helical reconstruction in RELION. *J. Struct. Biol.* **198**, 163–176 (2017).
10. A. Kendall, W. Bian, A. Maris, C. Azzo, J. Groom, D. Williams, J. Shi, P. L. Stewart, J. S. Wall, G. Stubbs, A common structure for the potexviruses. *Virology* **436**, 173–178 (2013).
11. V. V. Dolja, V. P. Boyko, A. A. Agranovsky, E. V. Koonin, Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: Two families with distinct patterns of sequence and probably structure conservation. *Virology* **184**, 79–86 (1991).
12. M. N. Jagadish, D. Huang, C. W. Ward, Site-directed mutagenesis of a potyvirus coat protein and its assembly in *Escherichia coli*. *J. Gen. Virol.* **74**, 893–896 (1993).
13. L. Holm, P. Rosenström, Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
14. T. Kawabata, MATRAS: A program for protein 3D structure comparison. *Nucleic Acids Res.* **31**, 3367–3369 (2003).
15. Y. Zhang, J. Skolnick, TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
16. Q. Ye, R. M. Krug, Y. J. Tao, The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* **444**, 1078–1082 (2006).
17. J. T. Huiskonen, A. K. Overby, F. Weber, K. Grünwald, Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: Evidence for GN-GC glycoprotein heterodimers. *J. Virol.* **83**, 3762–3769 (2009).
18. A. N. Freiberg, M. B. Sherman, M. C. Morais, M. R. Holbrook, S. J. Watowich, Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography. *J. Virol.* **82**, 10341–10348 (2008).
19. R. Arranz, R. Coloma, F. J. Chichón, J. J. Conesa, J. L. Carrascosa, J. M. Valpuesta, J. Ortín, J. Martín-Benito, The structure of native influenza virion ribonucleoproteins. *Science* **338**, 1634–1637 (2012).
20. J. A. Díaz-Pendón, R. Fernández-Muñoz, M. L. Gómez-Guillamón, E. Moriones, Inheritance of resistance to watermelon mosaic virus in *Cucumis melo* that impairs virus accumulation, symptom expression, and aphid transmission. *Phytopathology* **95**, 840–846 (2005).
21. I. M. Moreno, J. M. Malpica, J. A. Díaz-Pendón, E. Moriones, A. Fraile, F. García-Arenal, Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. *Virology* **318**, 451–460 (2004).
22. X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, Y. Cheng, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
23. J. A. Mindell, N. Grigorieff, Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
24. G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, S. J. Ludtke, EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
25. A. Desfosses, R. Ciuffa, I. Gutsche, C. Sachse, SPRING—An image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J. Struct. Biol.* **185**, 15–26 (2014).
26. A. Kucukelbir, F. J. Sigworth, H. D. Tagare, Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
27. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **32**, W526–W531 (2004).
28. G. D. Pintiile, J. Zhang, T. D. Goddard, W. Chiu, D. C. Gossard, Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.* **170**, 427–438 (2010).
29. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

30. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
31. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *PHENIX: A comprehensive Python-based system for macromolecular structure solution*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
32. V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *MolProbity: All-atom structure validation for macromolecular crystallography*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
33. I. Walsh, G. Minervini, A. Corazza, G. Esposito, S. C. E. Tosatto, F. Fogolari, *Blues server: Electrostatic properties of wild-type and mutated protein structures*. *Bioinformatics* **28**, 2189–2190 (2012).
34. F. Fogolari, A. Corazza, V. Yarra, A. Jalaru, P. Viglino, G. Esposito, *Blues: A program for the analysis of the electrostatic properties of proteins based on generalized Born radii*. *BMC Bioinformatics* **13** (suppl. 4), S18 (2012).
35. T. Kawabata, K. Nishikawa, *Protein structure comparison using the markov transition model of evolution*. *Proteins* **41**, 108–122 (2000).
36. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufu, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretidin, Y. Bao, O. Blinkova, V. Brover, V. Chetvermin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, *Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation*. *Nucleic Acids Res.* **44**, D733–D745 (2016).
37. F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. *Mol. Syst. Biol.* **7**, 539 (2011).
38. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, *WebLogo: A sequence logo generator*. *Genome Res.* **14**, 1188–1190 (2004).
39. T. D. Schneider, R. M. Stephens, *Sequence logos: A new way to display consensus sequences*. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
40. S. Chenavas, L. F. Estrozi, A. Slama-Schwok, B. Delmas, C. Di Primo, F. Baudin, X. Li, T. Crépin, R. W. H. Ruigrok, *Monomeric nucleoprotein of influenza A virus*. *PLOS Pathog.* **9**, e1003275 (2013).
41. D. D. Raymond, M. E. Piper, S. R. Gerrard, G. Skiniotis, J. L. Smith, *Phleboviruses encapsidate their genomes by sequestering RNA bases*. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19208–19213 (2012).
42. J. Reguera, H. Malet, F. Weber, S. Cusack, *Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein*. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7246–7251 (2013).
43. K. Komoda, M. Narita, K. Yamashita, I. Tanaka, M. Yao, *Tomato spotted wilt tospovirus nucleocapsid protein-ssRNA complex* (PDB code: 5IP2; Protein Data Bank, 2017); www.rcsb.org/pdb/explore/explore.do?structureId=5IP2.

Acknowledgments: We thank the Spanish Ministry of Economy and Competitiveness (MINECO) for the Severo Ochoa Excellence Accreditation (SEV-2016-0644). We also thank the Netherlands Centre for Electron Nanoscopy (NeCEN) (Leiden, Netherlands) for the collection of cryoEM images at the facility and C. Diebold for the technical assistance. **Funding:** This work was supported by grants from MINECO (BFU2015-66326-P to M.V. and AGL2015-65838-R to M.A.A. and M.A.S.-P.). **Author contributions:** M.Z., R.C., X.A., and M.V. prepared the cryoEM samples, processed the images, and generated the cryoEM maps and atomic models. E.M.-L., M.A.S.-P., and M.A.A. grew the plants and isolated the WMV virions from infected plants. J.L.L. performed the alignment of protein sequences. M.A.A. and M.V. designed the research and wrote the article with contributions from the rest of the authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or in Supplementary Materials. Additional data related to this paper may be requested from the authors. The cryoEM map for WMV and the calculated atomic model for its CP and ssRNA are available in the Electron Microscopy Data Bank (www.emdatabank.org) and PDB (www.rcsb.org) under accession codes EMD-3785 and 5ODV, respectively.

Submitted 27 June 2017

Accepted 18 August 2017

Published 20 September 2017

10.1126/sciadv.aao2182

Citation: M. Zamora, E. Méndez-López, X. Agirrezabala, R. Cuesta, J. L. Lavín, M. A. Sánchez-Pina, M. A. Aranda, M. Valle, *Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses*. *Sci. Adv.* **3**, eao2182 (2017).

Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses

Miguel Zamora, Eduardo Méndez-López, Xabier Agirrezabala, Rebeca Cuesta, José L. Lavín, M. Amelia Sánchez-Pina, Miguel A. Aranda and Mikel Valle

Sci Adv 3 (9), eaao2182.
DOI: 10.1126/sciadv.aao2182

ARTICLE TOOLS

<http://advances.sciencemag.org/content/3/9/eaao2182>

SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2017/09/18/3.9.eaao2182.DC1>

REFERENCES

This article cites 42 articles, 9 of which you can access for free
<http://advances.sciencemag.org/content/3/9/eaao2182#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Supplementary Materials for

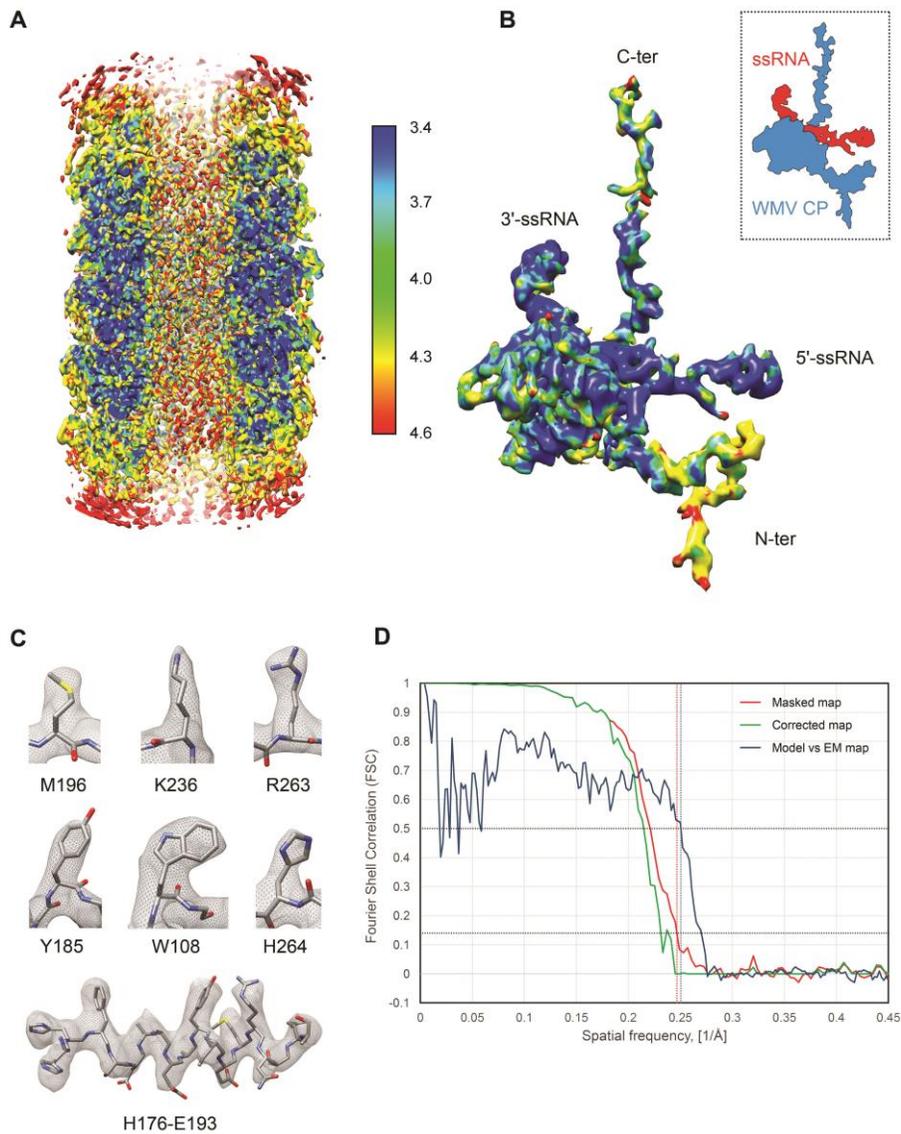
Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses

Miguel Zamora, Eduardo Méndez-López, Xabier Agirrezabala, Rebeca Cuesta, José L. Lavín, M. Amelia Sánchez-Pina, Miguel A. Aranda, Mikel Valle

Published 20 September 2017, *Sci. Adv.* **3**, eaao2182 (2017)
DOI: 10.1126/sciadv.aao2182

This PDF file includes:

- fig. S1. CryoEM and atomic model for WMV.
- fig. S2. Structural comparison between potyvirus and potexvirus.
- fig. S3. Representation of the structural alignment between regions of CPs and NPs from several ssRNA viruses.
- fig. S4. Output from the 3D library search using the WMV CP core region as target.
- fig. S5. Morphological universe of ssRNA viruses with structural homology between their CPs/NPs.



E

Clashscore, all atoms	0.87	100th/99th percentile ^{1*}
MolProbity score	1.39	100th/95th percentile ^{2*}
Ramachandran outliers	0	0.00 %
Ramachandran favored	178	86.83%
Poor rotamers	0	0.00 %
Favored rotamers	179	99.44%
Bad Backbone bones	0/1698	0.00 %
Bad backbone angles	0/2297	0.00 %

^{1*} 100th percentile compared with N=37 atomic structures with resolution >3Å, and 99th with N=1784 structures in all the range of resolutions.

^{2*} 100th percentile compared with N=342 atomic structures with resolution range 3.25-4.25 Å, and 97th with N=27675 structures in all the range of resolutions.

fig. S1. CryoEM and atomic model for WMV. (A) Iso-surface rendering of the cryoEM map calculated for WMV virions. Colors of the map show estimates of local resolution (color-scale included). (B) Rendering of the density map attributed to one WMV CP subunit after segmentation. Signal for a segment of ssRNA is also included. The color-code indicates local resolution as in (A). The inset shows a cartoon with a schematic depiction of WMV CP and ssRNA. (C) Local features of the cryoEM map together with modeled atomic coordinates. The density depicts accurately the side chain of bulky amino acids. (D) Fourier Shell Correlation (FSC) calculated between fully independent half data sets (gold standard), and between the final cryoEM map and a simulated density derived from the atomic model. The FSC thresholds of 0.14 for the cryoEM data, and of 0.5 for the comparison between cryoEM and the atomic model, estimate the global resolution at 4.0 Å. (E) Table summary of the figures of merit calculated in Molprobity for the validation of the atomic model of WMV CP.

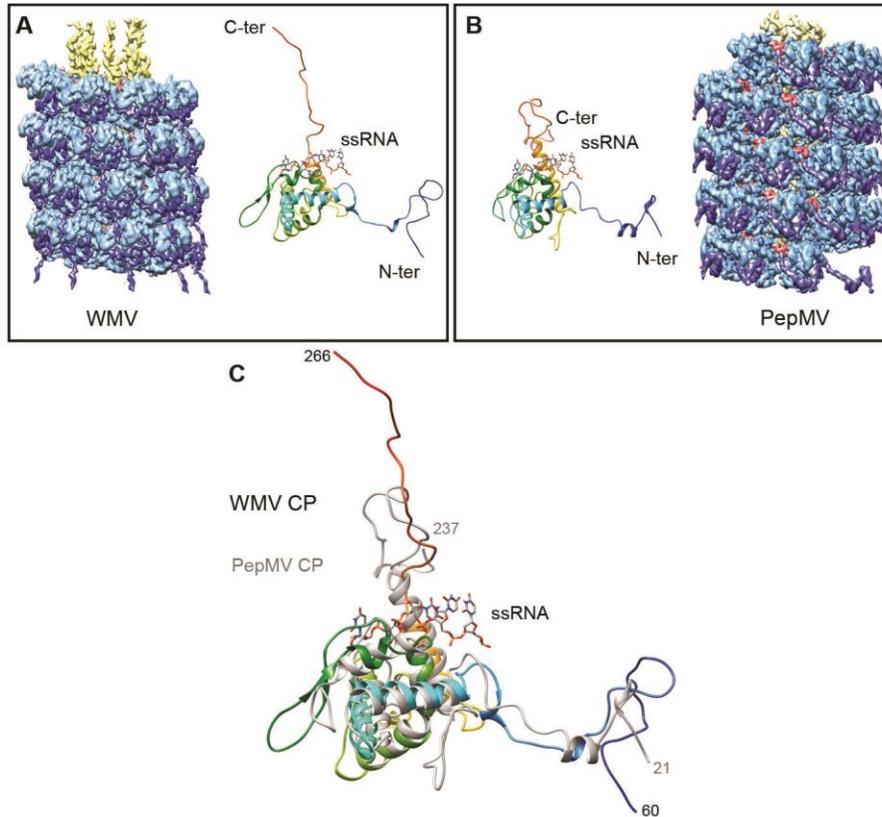


fig. S2. Structural comparison between potyvirus and potexvirus. (A) CryoEM map rendering of potyvirus WMV together with the derived atomic model for WMV CP and the ssRNA. (B) Iso-surface rendering for the cryoEM map for PepMV together with the atomic model for PepMV CP and corresponding ssRNA (4). In both panels A and B, the densities are colored: CP core region (light blue); N-terminal arm (dark blue); C-terminal arm (yellow); density for the ssRNA (red). The ribbons are colored in rainbow mode from N- to C-terminal ends. (C) Superimposition of the two atomic models for WMV CP (Rainbow colors) and PepMV CP (gray). The structural alignment was generated by Matras (13).

WMV CP vs PepMV CP (5FN1)

Aligned length= 135, RMSD= 2.85, Seq_ID=n_identical/n_aligned= 0.178
TM-score= 0.72404

NTRATKTKQFESWYSAVKVEYDLNDEQMGVIMNGFMVWCIDNGTSPDVNGVWVMDGEE-QVEYPLK--PIVENAKPTLRQIMHHFSDAAEAYIEMRNSESPYMPRYGLLRNLDRRELARYAFDFY-EVTS-----KTPNRAREIAQMKAAL-----
.....
TSVATPAEIEALGKIFTA-MGLAANETGPAMWDLARAYADVQSSKSAQLIGATPS---NPALSRRALAAQFDRINITPRQFCMYFAKVVWVWILLDSN---IPANWAKLGYQE-DTKFAAFDFPDGVTNPASLQPADGLIRQPNEKELAAHSVAKYGALARQKIS

WMV CP vs RFEV NP (4H50)

Aligned length= 125, RMSD= 4.04, Seq_ID=n_identical/n_aligned= 0.064
TM-score= 0.57568

NTRATKTKQFESWYSAVKVEYDLNDEQMGVIMNGFMVWCIDNGTSPDVNGVWVMDGEEQVE--YP-LK---PIVEN--AKP-T-L-RQIMHHFSDAAEAYIEMRNSESPYM-PRYGLLRNLR---DRELARYAFDFY-E-VTS-K---TPNRAREIAQMKAAL
.....
-GFDARRVIELLKQ-YG-----GADWEKDAKMMIVLALTRGN-KPRRMMMKMS---KEGKATVEALINKYKLEKE-GNPSRDELTLRVAALAGRTCALV-V----LSEWLPVTGTTMDGLSPAYPRHMHPSFAGMVDPSLPGDYLRALDAHSLYLLQFS

WMV CP vs TSWV NP (5IP2)

Aligned length= 110, RMSD= 4.68, Seq_ID=n_identical/n_aligned= 0.100
TM-score= 0.44432

NTRATKTKQ---FESW---YSAVKVEYDLNDEQMGVIMN-GFMVWCIDNGTSPDVNGVWVMDGEEQVEYF--LKPIVE---NAK-----P-----TLRQIMHHFSDAAEAYIE-M-RNSESPYMP--R-YGLLRNLDRDRE--LARYAF--DFY---EV-TSKTPNRAREIAQMKAAL
.....
FNFKTF--CLEN--LDQ-----IKK-MSIISCLTFLKNRQSIMKVI-----KQSDTFE-GKITIKKTSRIGATDMTFRRLDSLIRVRLVEETGENSENLNTIKSHASHPLIQAYGLPLDDAKSVRLAIMLGGSLPLIASVDSFEMISVVLAIYQDAKY-

WMV CP vs La Crosse virus NP (4BHH)

Aligned length= 103, RMSD= 4.36, Seq_ID=n_identical/n_aligned= 0.068
TM-score= 0.42651

NTRATKTKQFESWYSAVKVEYDLNDEQMGVIMN-GFMVWCIDNGT-SPDVNG-----VWVMDGEEQVEYPLKPIVENA-----K-P----TLRQIM-HHFSDAAEAYIE---M-RNSESPY--MPRYGLLRNLDRDRE--LARYAFDFY-EVTSKTPNRAREIAQMKAAL
.....
-----ES-LNLAAVRIFFLNAAKAKAAL-SRKPERKANPKFG-----EWQV-EVINNHFPGNRNNPIGNNDLTIHRLSGYLARWLDQYNENDESQHELIRTTIINPIAESNGVGWDSGPEIYLSFFPGTEMFLETFFKPYPLTIGIHR-V---

WMV CP vs Influenza A virus NP (3ZDP)

Aligned length= 121, RMSD= 4.51, Seq_ID=n_identical/n_aligned= 0.066
TM-score= 0.52572

NTRATKTKQFESWYSAVKVEYDLNDEQM-G-VIMNGFMVW--C-----IDNGTSPD-----VNGVWVMDGE-----EQVEYPLK--PIVENAK----PTLRQIMHHF-SDAAEAYIEMRNSESPYMPRYGLLRNL--RD-R----E--L-ARYAF-DFY-EVTSKTPNRARE-ATA-QMKAAL----
.....
SVGKMIDGIGRFYIQMCTELKLSDEYGRLIQNSLTIERMVLSAFDERRNKYLEEH---PSAGKDPKKTGGPIY--RRVDGKWRRELILYDKEEIRRIWRQANNDDATAGLTHMIIWHSNLDATYQR-TRA-----LVRTGMDPRMCSLMQGSTLPRRSGAAGAAVKVGTMVWEL-IRM-----IKRG

fig. S3. Representation of the structural alignment between regions of CPs and NPs from several ssRNA viruses. The structure of current atomic model for WMV CP was compared with several viral proteins using TM-align (14). TM-score and RMSD between C α atoms after the 3D alignment are shown. The output includes the alignment of protein sequences where ":" denotes aligned residue pairs of C α atoms where the distance between them is below 5.0 Å, and "." denotes other aligned residues. The compared structures are as in Fig. 4: WMV CP and ssRNA from the current work; PepMV CP, PDB code 5FN1 (4); *Influenza virus A* NP, PDB code 3ZDP (20); RVFV NP, PDB code 4H5O (21); *La Crosse* virus NP, PDB code 4BHH (22); and TSWV NP, PDB code 5IP2 (23).

fig. S4. Output from the 3D library search using the WMV CP core region as target.

The core region of the atomic model for WMV CP was used as the input to search for structural homologues within a library of representative structures from the PDB in Matras (13). The representation of primary sequences shows the alignment of secondary structure elements. The detected structural homologues have Z-score above meaningful threshold ($Z_{sc} > 5$). The picked atomic models within the database are: PepMV CP, PDB code 5FN1 (4); RVFV NP, PDB code 3OV9 (43); *Influenza virus A* NP, PDB code 4DYN (44); and La Crosse virus NP, PDB code 4BGP (22). The figures of merit provided for the structural alignment and similarity measurements are: Naa, number of compared amino acids; Ncmp, Number of compared secondary structure elements; SqID, sequence identity (%) defined as number of identical amino acids divided by Naa; rms, root mean square deviation (Å) of aligned C α atoms; Ssse, score of secondary structure elements; Rsse, normalized Ssse score (%); Sdis, Distance score; Rdis, normalized distance score (%); Zsc, Z score; RelS, Reliability (%) that a pair with this Zsc belongs to the same SCOP superfamily; RelO, Reliability (%) that a pair with this Zsc belongs to the same SCOP fold.

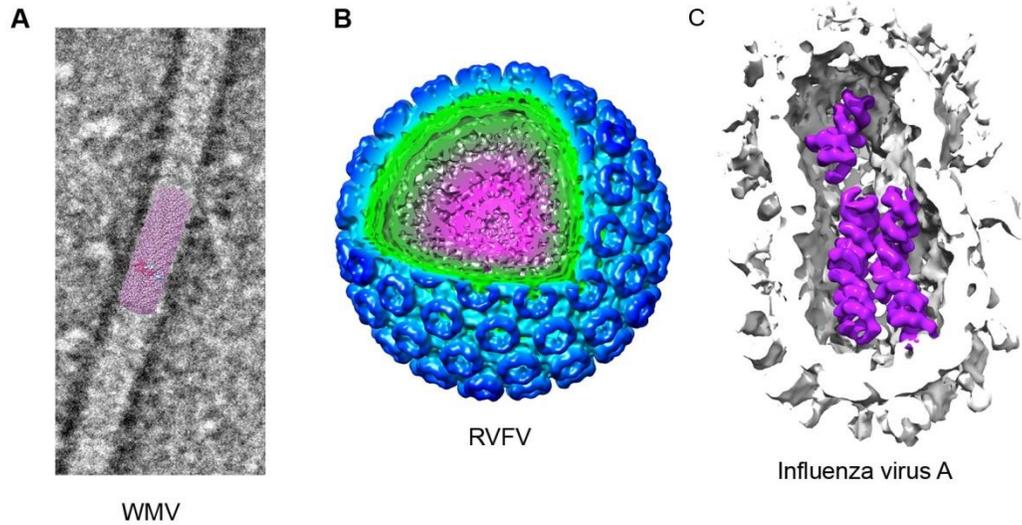


fig. S5. Morphological universe of ssRNA viruses with structural homology between their CPs/NPs. (A) Fraction of a negatively stained WMV particle. Superimposed there is a rendering of the cryoEM density map calculated for the virions. (B) Rendering of the cryoEM map for RVFV (EMD-5124, (45)). (C) Segmented cryoelectron tomogram for influenza A virus (18). Ribonucleoprotein complexes are depicted in purple (image courtesy of J. Martín-Benito). The relative sizes of the depicted structures are proportional to their real dimensions.