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. Lavin,1 M. Amelia Sánchez-Pina,2 Miguel A. Aranda,2 Mikel Valle1*

Potyviruses 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 resolution 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), allowing 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 a RNA binding pocket (nucleotide position labeled as U4 in Fig. 2).

1Molecular Recognition and Host-Pathogen Interactions, Center for Cooperative Research in Biosciences, CIB biogUNE, 48160 Derio, Spain. 2Centro de Edafología y Biología Aplicada del Segura (CEBAS), Consejo Superior de Investigaciones Científicas (CSIC), Espinar, 30100 Murcia, Spain. *Corresponding author. Email: mvalle@cibbiogune.es

Essentially, the interactions between Ser\textsuperscript{140} and Arg\textsuperscript{172}, with consecutive phosphate groups in the RNA (Fig. 2B), hold the flanked nucleoside deep inside this pocket, where its base is surrounded by Arg\textsuperscript{172}, Asp\textsuperscript{216}, and Lys\textsuperscript{236} 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
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 potexvirus 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 hit 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, nonenveloped 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
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; nonetheless, 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°C/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...
homogenized in a buffer containing 0.5 M K$_2$HPO$_4$ (pH 7.5), 5 mM EDTA, 10 mM diethylldithiocarbamic acid, and 20 mM Na$_2$SO$_3$ (5 ml/g). The mixture was stirred for 15 min at 4°C and centrifuged for 10 min at 7500 $g$. 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 300,000 $g$, 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$_2$SO$_3$ (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 15,000 $g$. The aqueous phase was overlaid onto 5 ml of a 30% sucrose cushion and centrifuged for 110 min at 245,000 $g$. The pellet was resuspended overnight in buffer A with constant stirring at 4°C. After a low-speed centrifugation, Cs$_2$SO$_4$ (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 108,000 $g$. The pellet containing the purified virions was resuspended in 200 ml 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 ×130,000 resulting in a sampling of 1.1 Å/pixel. Exposures of 5 s in electron counting mode resulted in images with 35 frames and a total dose of 35 e$^-$/Å$^2$. Motion between frames was corrected in micrographs (22) using frames 2 to 27, resulting in an accumulative dose of 27 e$^-$/Å$^2$. 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 × 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 Å (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 Å 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 Å (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://robetta.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 Bluesys server (http://protein.bio.unipd.it/bluesys) (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://strcprop.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.ccb.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 Cardivirus 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 Tomato yellow leaf curl virus, where the conserved Arg is substituted by His, a conservative mutation between basic amino acids.

REFERENCES AND NOTES


Downloaded from http://advances.sciencemag.org/ on November 18, 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