The molecular basis for flexibility in the flexible filamentous plant viruses

Frank DiMaio1,7, Chun-Chieh Chen2–4,7, Xiong Yu5, Brandon Frenz1, Yau-Heiu Hsu4, Na-Sheng Lin2–3,6 & Edward H Egelman5

Flexible filamentous plant viruses cause more than half the viral crop damage in the world but are also potentially useful for biotechnology. Structural studies began more than 75 years ago but have failed, owing to the virion’s extreme flexibility. We have used cryo-EM to generate an atomic model for bamboo mosaic virus, which reveals flexible N- and C-terminal extensions that allow deformation while still maintaining structural integrity.

The flexible filamentous plant viruses are single-stranded positive (sense)-strand RNA viruses that are widely found and are responsible for more than half the viral crop damage in the world.1 Owing to their low toxicity, they are viewed as potentially useful in biotechnology, for example as biomaterials for vaccine delivery3,4, drug delivery or imaging5. Filamentous plant viruses are broadly classified into the rigid rod-like viruses, such as tobacco mosaic virus (TMV), and the flexible viruses. Although a previous study found ten rigid filamentous plant viruses in substantial quantities in human stool samples,6 the flexible filamentous plant viruses were conspicuously absent, thus suggesting that they can be metabolized, whereas the rod-like viruses cannot. In addition, the flexible filamentous viruses are potentially valuable for recombinant-protein production in plants.7,8 However, all of these applications have been hampered by the absence of atomic structures. Published structural studies of these viruses date from 1941 (ref. 8), but no atomic model has been possible, because the viruses cannot be crystallized and have proven to be too flexible for high-resolution X-ray fiber diffraction or electron cryo-EM. TMV, the first virus to be discovered,9 is a rigid filamentous plant virus that has been a model system in structural biology and virology. Atomic models of TMV have been produced by both X-ray diffraction10 and cryo-EM11. In contrast to the rigid viruses, the flexible filamentous viruses, including potexviruses such as potato virus X, could not generate high-quality fiber diffraction patterns. It was previously suggested, on the basis of low-resolution X-ray fiber diffraction, that all potexviruses may share a common architecture with slightly fewer than nine protein subunits per helical turn12. This conclusion was subsequently strengthened with results from X-ray diffraction and various forms of EM.13,14 Several low-resolution models of the flexible plant viruses have been generated1,13,14, all implicitly assuming that the virions have a right-handed helical pitch as found in TMV.10

Bamboo mosaic virus (BaMV) belongs to the genus Potexvirus, family Alphaflexiviridae. It has a single-stranded RNA genome of about 6.4 kb and has a flexible filamentous morphology with a length of 490 nm and a diameter of 15 nm (ref. 16), and it is built mainly from a single protein, coat protein (CP). Previous research has shown that up to 35 residues of the N terminus of CP can be deleted with no effect on virus replication and assembly.15 Thus, BaMV has been developed as a plant expression vector for vaccine production by replacement of the N-terminal 35 residues with foreign peptides from either foot-and-mouth disease virus9 or infectious bursal disease virus4. We set out to determine an atomic structure for a flexible filamentous plant virus. We imaged both the wild-type BaMV (WT) and a virion containing a deletion of 35 N-terminal residues of CP, denoted BaMV Nd35 (Nd35), using cryo-EM with a direct electron detector (Fig. 1). Starting with power spectra from the filaments (Supplementary Fig. 1), we found the symmetry by trial and error until recognizable secondary-structure elements (rod-like features from α-helices) were observable, by using a previously described approach.18 We found no change in the symmetry between the WT and the Nd35: both have a pitch of ~35 Å with ~8.8 subunits per turn. Upon examination of the crystal structure of a large fragment of the papaya mosaic virus (PapMV) CP (PDB 4DON),14 which has 28% sequence identity with the corresponding region of BaMV, it became obvious that the ~35-Å pitch helix must be left-handed. (Two enantiomorphic reconstructions can be generated that are equally consistent with the images—one right-handed and one left-handed—but the crystal structure could fit into only the left-handed one.) We improved the reconstructions of both the WT and Nd35 BaMV by using a classification approach to remove the large variability in twist and rise, and we used a more homogeneous set of ~54,000 segments for the WT and ~50,000 segments for the Nd35 for the final reconstructions of each. The absence of any substantial structural differences between the two volumes allowed us to use these completely independent reconstructions for an estimate of resolution, which yielded a value of 5.6 Å (Supplementary Fig. 2). We then combined the two sets to generate a reconstruction (Fig. 1b) used for modeling. Owing to the intrinsic variability of the structure, the resolution of the combined reconstruction was not substantially better than

1Department of Biochemistry, University of Washington, Seattle, Washington, USA. 2Department of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan. 3Molecular and Biological Agricultural Sciences Program, Taiwan International Graduate Program, Academia Sinica, Taipei, Taiwan. 4Graduate Institute of Biotechnology, National Chung-Hsing University, Taichung, Taiwan. 5Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia, USA. 6Biotechnology Center, National Chung-Hsing University, Taichung, Taiwan. 7These authors contributed equally to this work. Correspondence should be addressed to E.H.E. (egelman@virginia.edu), N.-S.L. (nslin@sinica.edu.tw) or Y.-H.H. (yhhsu@dragon.nchu.edu.tw).

Received 15 May; accepted 3 June; published online 13 July 2015; doi:10.1038/nsmb.3054
that of either of the individual reconstructions. A comparison between the surface of the combined reconstruction (Fig. 2a) and the atomic model that we have built, filtered to 5.6 Å (Fig. 2b), shows that our estimate of the resolution is very reasonable.

The existing crystal structure of the PapMV fragment (PDB 4DOX) provided us with an initial template for building a full atomic model of BaMV (Fig. 2c–e and Supplementary Fig. 3). We initially docked the template into the density, and the core (residues 33–174 in PapMV, corresponding to residues 62–201 in BaMV) showed good agreement with the experimental data, with real-space correlation of 0.61 over these 143 residues composing the compact core of the crystal structure (Fig. 3a,b). However, this model failed to account for 61 N-terminal BaMV residues (18 of which were present in the template model but were probably stabilized by crystal contacts and poorly fit the experimental density) and 42 C-terminal residues, which were truncated for crystallization. Continuous density was clearly visible for both the missing N-terminal and C-terminal residues on the outside and the inside of the capsid, respectively.

We used Rosetta to build the backbone of the missing C terminus and to rebuild the N terminus (Fig. 3c,d). Because of the length, the relatively low resolution of the local density and the low apparent secondary-structure content of the insertions, building or rebuilding these termini proved especially challenging. Ultimately, we used a new enumerative backbone sampling protocol (Online Methods). Sampling these terminal conformations revealed reasonable convergence of the top-scoring models (Supplementary Fig. 4). An unbroken tube of density, which remained unexplained by the models, seemed likely to correspond to the single-stranded RNA. Comparison to an independent data set (Supplementary Fig. 2) shows similar agreement to the map used for fitting, thus indicating that the model is not overrefined.

The final structure shows a highly intertwined topology (Fig. 2c), in which each subunit makes direct contact with eight other subunits (Fig. 2d,e). As viewed from the outside of the capsid (Fig. 2d), or 5 as the nearest integer. We docked and refined a five-nucleotide sequence from rift valley fever virus (PDB 4H50) because the nucleotide chain in this structure had a very similar radius of curvature to that observed in the density (~23 Å in the crystal structure versus ~30 Å in the density map), and a five-nucleotide stretch (nucleotides 3–7) showed good agreement when docked into the experimental data (Fig. 3e).

Finally, we carried out all-atom refinement of the symmetric full-length model against the experimental density data in Rosetta. After refinement, we selected and compared the lowest-energy structures. These structures showed relatively tight convergence (Supplementary Fig. 4); however, convergence was noticeably worse in the C terminus, thus making the specific side chain interactions stabilizing this long loop ambiguous. Regions of the model that used the crystal structure as a starting point were very well converged, and the final model showed only very modest deviation from the initial crystal structure, with a Cα r.m.s. deviation of 3.1 Å; differences between the two are largely limited to several loops interacting with the single-stranded RNA. Comparison to an independent data set (Supplementary Fig. 2) shows similar agreement to the map used for fitting, thus indicating that the model is not overrefined.

The final structure shows a highly intertwined topology (Fig. 2c), in which each subunit makes direct contact with eight other subunits (Fig. 2d,e). As viewed from the outside of the capsid (Fig. 2d), or 5 as the nearest integer. We docked and refined a five-nucleotide sequence from rift valley fever virus (PDB 4H50) because the nucleotide chain in this structure had a very similar radius of curvature to that observed in the density (~23 Å in the crystal structure versus ~30 Å in the density map), and a five-nucleotide stretch (nucleotides 3–7) showed good agreement when docked into the experimental data (Fig. 3e).
the N terminus of each subunit wraps around the i − 1 subunit, forming a short helix that packs into a hydrophobic cleft on the surface, with Phe45 buried in a small pocket on the surface. An N-terminal loop continues wrapping around the structure, and Trp41 and Trp68 form a stacking interaction. The first 38 N-terminal residues are disordered in our model; this is validated by the striking similarity between the WT and Nd35 reconstructions. The C terminus (Fig. 2e) wraps through the core of the capsid, follows a continuous tube of density, and forms contacts with three subunits in the turn above (i + 7, i + 8 and i + 9) before pointing toward the center of the virion, where the extreme C-terminal residues contact subunit i − 7 in the turn below.

Although the resolution of the data does not permit us to draw conclusions on the nature of the protein-RNA interactions, the model suggests that residues Arg99, Lys132, Lys157, Lys202 and Lys213 all potentially make protein-RNA contacts. Among them, Arg99 was previously found to be part of the potential RNA-binding motifs in the BaMV CP.

This highly intertwined structure of a flexible filamentous virus is in sharp contrast to the highly compact architecture of TMV. The asymmetric unit of TMV is similar in size to that of BaMV, comprising 158 residues in TMV versus 204 residues in BaMV (ignoring the disordered 38 N-terminal residues), with an accessible surface area of 8,787 Å² per subunit for TMV versus 12,922 Å² for BaMV (8,075 Å² not including the extended termini). TMV, however, forms much more extensive contacts with neighboring subunits, with 4,499 Å² (or about 51%) of surface forming the interface in the assembled capsid. However, BaMV’s compact core makes only modest contacts between subunits, with 1,806 Å² (22%) of its surface contacting neighboring subunits; the contacting surface increases to 5,752 Å² (45%), in line with that of TMV, only upon inclusion of the N- and C-terminal extensions, which are connected to the core of the subunit by very flexible linkers. It is this architecture that allows the extensive noncovalent interactions with many surrounding subunits to be maintained in the flexible filamentous viruses, because mechanical forces cause the structures to deform.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. The map and the model have been deposited in the Electron Microscopy Data Bank and the Protein Data Bank, respectively, under accession codes EMD-3020 and PDB 5A2T.
ONLINE METHODS

BaMV preparation. The plasmid pCB is a full-length cDNA infectious clone of BaMV-S (GenBank AF018156) in a pCass2 vector, as described previously. The pCB-Nd35 was derived from pCB by deletion of the N-terminal 35–amino acid sequence of CP. The plasmids (1–2 µg) were used to inoculate Chenopodium quinoa. The inoculated leaves with local lesions were collected 7–10 d after inoculation. BaMV particles were extracted and purified as described previously.

Cryo-EM and image processing. The sample (3 µL, 1–2 µg/µL) was applied to plasma-cleaned lacey carbon grids (Gatan Solarus) and vitrified in a Vitrobot Mark IV (FEI). Grids were imaged in a Titan Krios at 300 keV and recorded using a Falcon II direct electron detector at 1.05 Å per pixel, with seven ‘chunks’ per image. Each chunk, containing multiple frames, represented a dose of ~20 electrons per Å². A total of 914 images (each 4,000 × 4,000) from the WT and 560 images from the Nd35 sample were selected that were free from drift or astigmatism and had a defocus less than 3.0 µm. CTFFind3D (ref. 23) was used to determine the contrast transfer function (CTF), and the range used was from 0.6 to 3.0 µm. The SPIDER package24 was used for most subsequent steps. The CTF was corrected by multiplying each image by the theoretical CTF, both reversing phases where they need to be reversed and improving the signal-to-noise ratio. The program e2helixboxer within EMAN2 (ref. 25) was used for boxing long filaments from the micrographs, and 5,099 and 3,400 such boxes of varying length were generated from the WT and Nd35 samples, respectively. Overlapping boxes, 384 pixels long with an 8-pixel shift between adjacent boxes (98% overlap) were extracted from these long filaments, thus yielding 236,726 segments used for the initial alignments and reconstruction came from the first two chunks.

The determination of the helical symmetry was by trial and error, by searching for a symmetry yielding recognizable secondary structure18. The IHRSR to-noise ratio. The program e2helixboxer within EMAN2 (ref. 25) was used for eliminating segments with large out-of-plane tilt (greater than 9°) and for sorting by axial rise and twist. The sorting was done by generation of ~4.0 Å and a rotation of ~−40.9° per subunit) a preliminary reconstruction was generated by imposing the helical parameters found for each segment, from the first two chunks on segments containing only the first chunk (~20 electrons per Å²), and using these for the back-projection in SPIDER. The Fourier shell correlation (FSC) was generated by comparison of two completely independent reconstructions: the WT and the Nd35 (Supplementary Fig. 2), and the FSC = 0.143 criterion27 was used.

Building and refining atomic models into density. Model building began by docking a crystallized fragment from PapMV (PDB 4DOX) into the experimental density data with Chimera’s ‘dock into density’ tool. The fragment was truncated to the core residues (residues 33–174) that showed good agreement with the density data. However, this model failed to account for 58 N-terminal residues and 42 C-terminal residues. Continuous density was clearly visible for both the missing N-terminal and C-terminal residues at the outside and the inside of the capsid, respectively.

Initially, these termini were rebuilt with RosettaCM, which combines Monte Carlo sampling of backbone fragments with Cartesian space minimization28. However, owing to the length, the relatively low resolution of the local density, and the low apparent secondary-structure content of the insertions, backbone conformational sampling in RosettaCM was poorly converged, and the best-scoring models still poorly fit the density, with large segments outside of the density and substantial amounts of unexplained density remaining. Instead, we used a new enumerative rebuilding strategy in Rosetta to overcome this sampling issue. Rather than sample the entire backbone segment simultaneously, we iteratively sampled short three-residue segments of backbone. By considering only three-residue segments, we could completely explore the space of backbone conformations, given each 3–amino acid segment. For each iteration, we stored a ‘beam’ containing up to 50 of the best solutions; each subsequent iteration attempts to extend each of these solutions and stores up to the best 50 after the next extension. The density data are used to filter obviously wrong solutions (by throwing out solutions with density agreement substantially worse than the best seen over the same stretch of backbone), and additional filters ensure that models stored at each iteration are sufficiently different from one another. Sampling these terminal conformations revealed good convergence of the top-scoring models (Supplementary Fig. 4a) when we looked at the final beam.

After rebuilding the N and C termini, an unbroken tube of density, which presumably corresponded to the single-stranded RNA, remained unexplained by the models. The resolution of the data was unfortunately insufficient to build RNA models de novo with any degree of confidence. From the length of the capsid and genome, we assumed that there were five nucleotides in each asymmetric unit of the capsid. We identified a set of structures that had RNA with a similar radius of curvature (PDB 1C9S, 1RMV, 3PDM, 4BKK, and 4H5O) and considered docking and refining every five-residue segment into the density map. Refinement of the RNA was carried out according to the symmetry of the capsid29, with constraints used to ensure that bond geometry was maintained between adjacent asymmetric units. This refinement (25 different RNA stretches) showed that the best agreement to density was observed for residues 3–7 of 4H5O, a crystal structure of rift valley fever virus. The RNA conformation clashed with residues 85–96 in the docked crystal structure, so these were rebuilt with RosettaCM, with the RNA model present.

Finally, all-atom refinement of the symmetric full-length model against the experimental density data was carried out in Rosetta, with a previously described protocol30. A total of 600 refined models were generated. After refinement, the ten lowest-energy structures were selected and compared. These structures showed relatively tight convergence over most of the structure, with most deviation in the C terminus; this is unsurprising because this is the region with the worst local resolution, possibly owing to conformational heterogeneity in this region. All coordinate and B-factor refinement was carried out against the RNA model present.

An FSC curve comparing the final model to the WT map (Supplementary Fig. 2) shows good agreement between model and map, with FSC of 0.5 at a resolution of about 5 Å, in line with the resolution of the data. The agreement of the model with the Nd35 reconstruction shows a similarly good fit, thus suggesting that the refined model is not overfitted to the density data.