In situ structures of the genome and genome-delivery apparatus in an ssRNA virus

Xinghong Dai[1,2], **Zhihai Li**[2,3,4], **Mason Lai**[3], **Sara Shu**[1], **Yushen Du**[1], **Z. Hong Zhou**[2,3,*], and **Ren Sun**[1,2,*]

[1]Department of Molecular and Medical Pharmacology, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

[2]The California NanoSystems Institute (CNSI), University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

[3]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA

[4]State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, School of Life Sciences, Xiamen University, Xiamen 361102, Fujian, China

Summary

Genome packaging into a protein capsid and its subsequent delivery into a host cell are two fundamental processes in the life cycle of a virus. Unlike dsDNA viruses which pump their genome into a preformed capsid[1-3], ssRNA viruses, such as bacteriophage MS2, co-assemble their capsid with genome[4-7]; however, the structural basis of this co-assembly is poorly understood. MS2 infects *Escherichia coli* via host "sex" pilus (F-pilus)[8] and is the first fully-sequenced organism[9] and a model system for studies of gene translational regulations[10,11], RNA-protein interactions[12-14], and RNA virus assembly[15-17]. Its positive-sense ssRNA genome of 3569 bases is enclosed in a capsid with one maturation protein (MP) monomer and 89 coat protein (CP) dimers arranged in a T=3 icosahedral lattice[18,19]. MP is responsible for attaching the virus to an F-pilus and delivering the viral genome into the host during infection[8], but how the genome is organized and delivered are not known. Here we show the MS2 structure at 3.6Å resolution determined by electron-counting cryo electron microscopy (cryoEM) and asymmetric reconstruction. We traced ~80% backbone of the viral genome, built atomic models for 16 RNA stem-loops, and identified three conserved motifs of RNA-CP interactions among 15 of these stem-loops with diverse sequences. The stem-loop at 3' end of the genome interacts extensively with the MP, which, with just a six-helix bundle and a six-stranded β-sheet, forms a genome-delivery apparatus, and joins 89

CP-dimers to form a capsid. This first atomic description of genome-capsid interactions in a spherical ssRNA virus provides insights into genome delivery via host "sex" pilus and mechanisms underlying ssRNA-capsid co-assembly, and inspires imaginations about links between nucleoprotein complexes and the origin of viruses.

We imaged MS2 particles embedded in vitreous ice with a K2 direct electron detector attached to the end of an energy filter in a Titan Krios electron microscope and averaged more than 330,000 particles to calculate a 3.6Å-resolution reconstruction without imposing any symmetry (Extended Data Fig. 1). The external structure of the asymmetric reconstruction (Fig. 1a) appears similar to the icosahedrally-averaged crystallographic structure of MS2, which contains 90 CP-dimers[18,20]. However, a small density bulge emerges at one of the 2-fold axes and its structure differs from all of the other 89 CP-dimers (Fig. 1b, Supplementary Video 1). Initial analysis of secondary structures of this asymmetric feature and more detailed amino-acid tracing indicated it to be an MP monomer. High-resolution structural features in MP, such as side chains, are consistent with the estimated 3.6Å resolution of the map (Fig. 1d).

The reconstruction also reveals the ssRNA genome density inside the capsid. This density is a bit noisy and somewhat broken in the 3.6Å resolution map but structural features typical of RNA are apparent when the map is low-pass filtered to 6Å resolution (Fig. 1c, Supplementary Video 1), indicating some flexibility of the RNA chain relative to the capsid shell. To sort out possible structural heterogeneity due to the flexible RNA, we carried out further three-dimensional (3D) classification of the dataset and obtained ten structures. These structures are almost identical to each other except for several short, possibly flexible, segments (Extended Data Fig. 2, Supplementary Video 2). The RNA density is not uniformly distributed within the capsid, with the MP-proximal side of the space more densely packed than the distal side (Fig. 1c). The majority of the density shows prominent major and minor grooves (Fig. 1c), hallmarks of double helices, indicating that most of the ssRNA has folded into stem-loops. We identified over 50 stem-loops, most of which contact the capsid at the tip (i.e., the loop region) (Fig. 1c).

Among the stem-loops identified, 16 show clearly resolved individual nucleotides and even features that distinguish purines from pyrimidines (Fig. 1d). Such features enabled us to derive possible sequences for each stem-loop, which were used to search against the viral genome to identify its genuine sequence. Using the identified sequences of these 16 stem-loops as landmarks across the genome, we were able to trace the backbone for more than 80% of the genome, build atomic models for the 16 stem-loops, and identify long-range base-pairing and kissing-loop interactions (Fig. 2, Extended Data Figures 3-9, and Supplementary Data 1-3).

Among the 16 stem-loops, the one at 3' end of the genome interacts with the MP and each of the remaining 15 binds to a CP-dimer (Extended Data Fig. 9). The 15 stem-loops vary greatly in their lengths and sequences (Fig. 3a). Comparison of the 15 stem-loop/CP-dimer structures reveals how these different stem-loops can be specifically recognized by the same CP during genome packaging. The non-sequence-specific, negatively-charged phosphates of the stem-loop RNA backbone interact with the CP-dimer through a patch of positively-

charged/polar residues on the two CP monomers (Lys43, Arg49, Ser51, Lys57 and Lys61 of one CP, and Arg49, Ser51, Ser52, Asn55, Lys57 and Lys61 of the other) (Fig. 3b, c). At individual base level, three conserved motifs of RNA-CP interactions can be identified (Fig. 3d, e, Supplementary Video 3). First, Thr45 and Ser47 of a CP hydrogen-bond with the base of a partially conserved nucleotide (red in Fig. 3a) in the RNA loop (Fig. 3e), preferably an A with two hydrogen bonds (e.g., A1757, Fig. 3e), and alternatively a C as in 3 of the 15 cases with only one hydrogen bond (e.g., C109, Fig. 3h). Second, in 4 of the 15 cases (magenta in Fig. 3a), the Thr45 and Ser47 of the second CP also form two hydrogen bonds with an unpaired and sticking-out base (a purine in all these 4 cases) in the stem region (Fig. 3d, f). But this purine has a flipped conformation as compared to that in the loop region (cf. A1751 in Fig. 3d or G2784 in Fig. 3f vs A1757 in Fig. 3e). Third, Tyr85 of one CP establishes an aromatic-ring stacking interaction with another base in the loop region (e.g., U1756, Fig. 3e). This base (green in Fig. 3a) is preferentially a pyrimidine, but is a purine in 4 of the 15 cases. It can be either next to the Thr45/Ser47-interacting base (Fig. 3e), or separated by a spacer (Fig. 3g). The aromatic-ring stacking is usually extended to include another unpaired base (e.g., A1754, Fig. 3e) also in the loop region, before merging with the base-stacking in the stem region (Fig. 3e). In addition to accommodation of diversities in sequence, these stem-loops also utilize the above described non-sequence-specific interactions and conserved motifs to accommodate difference in local environments (Fig. 3i, j).

Although each of the over 50 stem-loops in the MS2 genome might act as a "packaging signal" for capsid assembly as previously proposed[15-17,21], the higher resolution of the above 16 stem-loops indicates stronger interactions with capsid proteins, and thus perhaps a more important role in the capsid assembly. Three of these 16 stem-loops are consecutive in sequence (stem-loops 1714-1737, 1746-1764, and 1766-1806, Extended Data Fig. 8), cluster together (stem loops 1-3 in Extended Data Fig. 5) and bind three neighboring CP-dimers – a configuration desirable for nucleating capsid assembly. Indeed, the one in the middle of the three, which encompasses the start codon of the replicase gene, was proposed to nucleate capsid assembly[12,22]. Interestingly, the 16 stem-loops are non-uniformly distributed, with most of them clustered around the putative nucleation site (Fig. 2a). Because CP-dimers alone can assemble into both octahedral and wild-type-like icosahedral particles[23], guidance by the RNA stem-loops in the early stage of capsid assembly might be the size-determining factor leading to the assembly of wild-type, icosahedral capsid.

One MP monomer replaces a CP-dimer at one of the icosahedral 2-fold axes, imparting structural changes to the neighboring CPs (Fig. 4a). The MP structure consists of an α-helix domain (aa. 140-225, 269-313, and 375-393) with a bundle of six α-helices, and a β-sheet domain (aa. 1-139, 226-268, and 314-374) with six anti-parallel β-strands sandwiched between an N-terminal loop and a helix-loop-helix motif (Fig. 4b). The MP is slightly tilted off the capsid surface with its α-helix domain inserting into the capsid lumen and its β-sheet domain projecting out (Fig. 4a, Supplementary Video 1).

MP has extensive (>20) interactions with one of the 16 high-resolution stem-loops of the genome structure, i.e., stem-loop 3540-3563 at the 3' end of the genome (Fig. 4c, d). These interactions can be categorized into four types (Fig. 4e-h). First, phosphates in the RNA

backbone can form charge-charge interactions with basic residues in MP (Lys50, Arg43, 62, 358, and 229, Fig. 4e, f), or form hydrogen bonds with polar residues (Tyr60, Trp264, and Gln152, Fig. 4e, f, h) or even with amide of the protein backbone (Val141, Fig. 4f). Second, there are two cases of RNA base stacking with aromatic ring of the protein side chain: U3553 with Phe4, and U3557 with Trp54 (Fig. 4f, g). Third, there are multiple hydrogen bonds between RNA bases and protein side chains, primarily serines and threonines. These include interactions between U3549 and Thr47, U3552 and Asn45 (Fig. 4e), G3554 and Ser258/Thr324, G3555 and Ser58 (Fig. 4f). Lastly, seven hydrogen bonds are formed between RNA bases and carboxyl or amide groups of the protein backbone in the long loop connecting the first two β-strands of the MP (Fig. 4e, g). In this regard, Pro57 and Pro59 (Fig. 4g) might be important in maintaining the twisted conformation of the protein backbone, preventing the formation of β-strand interactions in this segment, and freeing those carboxyl and amide groups for hydrogen-bonding with the RNA bases. It is also noteworthy that the base-pairing in the MP-bound state of this stem-loop is not optimal. If the MP binds to the RNA after the stem-loop has been folded, four base-pairs have to be melted and two new ones formed (Fig. 4d) to accommodate the RNA-MP interactions described above. How this dynamic process is achieved awaits further exploration.

This MP-interacting stem-loop also interacts with CP, albeit much less extensively and in a manner that bears no similarity to that of the other 15 stem-loop/CP-dimer structures. Two phosphates in the stem region of this stem-loop interact with the same CP (yellow in Fig. 4h), one through a hydrogen bond with Asn27 and the other through charge-charge interaction with Arg49. Base C3548, which is sticking out from the loop region, forms a hydrogen bond with the terminal carboxylic acid group of Tyr129 of a second CP (beige in Fig. 4h). It may also stack with the Tyr129 aromatic ring in a third CP (purple in Fig. 4h).

Besides the extensive interactions with the 3' end stem-loop, MP also binds the backbone of three other stem-loops with positively charged Arg and Lys residues (Fig. 5a, b). Three arginines in the β-sheet domain bind a very short stem-loop 3524-3532 in the 3' untranslated region (3'-UTR) of the RNA. The other six arginines and two lysines are clustered at the tip of the α-helix bundle, binding stem-loops 1766-1806 and 1960-1995, with Arg188 inserted into the minor groove of stem-loop 1960-1995 (Fig. 5a).

During infection, MP is responsible for attaching the MS2 virion to the bacterial F-pilus and delivering the genome into the host[8]. Fitting our MS2 model into the structure of MS2/F-pilus complex[19] reveals interactions between F-pilus and the helix-loop-helix motif on one side of the β-sheet domain of MP with all β-strands nearly parallel to the pilus axis (Fig. 5c), resulting in slight tilting of the virion as observed previously[19,24] (Fig. 5d). Our results shed light on why genome release occurs only when MS2 binds to F-pilus on a living bacterium[25], but not when it binds to detached F-pilus[8]. Considering that F-pili in living bacteria are highly dynamic, with extension and retraction accompanied by rotation[26], the driving force for MS2 genome delivery might come from the dynamics of, rather than the binding to, the F-pili. As the pilus retracts, the virion gets stuck outside the cell due to its relatively large size, but MP and its tightly-bound ssRNA genome are pulled together out of the capsid shell, leading to the delivery of this ribonucleoprotein complex into the host. Indeed, MP is delivered into host cell with the genome during infection[27,28] while empty

MS2 capsid is left outside[29], and infectious ribonucleoprotein complexes can be reconstituted by mixing MP and ssRNA genome of MS2[30]. One might even imagine that after the primitive "RNA world", such a minimalist ribonucleoprotein complex could constitute a simple *replicator* to set the stage of evolution towards more sophisticated complexes. Selective pressure towards better fitness of the replicator naturally could have led to the acquisition of a protein coat to protect the RNA from a hostile environment.

Methods

Sample preparation

Coliphage MS2 (ATCC 15597-B1) and its host *Escherichia coli* strain C-3000 (ATCC 15597) were obtained from American Type Culture Collection (ATCC). Bacteria were cultured in ATCC-recommended broth media at 37°C, and infected with MS2 phage at middle log-phase. After complete lysis of the bacteria in a few hours, the lysate was centrifuged at 10,000g for 15min to remove cell debris, and then centrifuged again at 100,000g for 4hr to pellet the phage particles. The pellet was resuspended in buffer containing 50mM Tris (pH7.5), 150mM NaCl, 5mM $CaCl_2$ and 5mM $MgCl_2$. The suspension was applied to an OptiPrep (Sigma-Aldrich) density gradient (10-50% w/v with 10% steps) and centrifuged at 100,000g overnight. The phage band was clearly visible and was collected, diluted with buffer, and centrifuged at 100,000g for 4 hr. The purified phage particles were resuspended in 100μl buffer. To prepare cryoEM sample, an aliquot of 2.5 μl phage sample was applied to a Quantifoil grid, blotted with filter paper and plunge-frozen in liquid ethane with an FEI Vitrobot.

Cryo electron microscopy and movie preprocessing

CryoEM images were collected with Leginon[31] on an FEI Titan Krios electron microscope equipped with Gatan imaging filter (GIF) and K2 Summit direct detection camera. A nominal magnification of 130,000x was used, giving a pixel size of 1.06Å/pixel at the sample level. A slit width of 20eV was set for the energy filter. Movies were recorded with the K2 camera operated in counting mode with an electron dose rate below 8e$^-$/pixel/s. An accumulated dose of 50e$^-$/Å^2 on the sample was fractionated into a movie stack of 29 image frames.

For each of the total 6,080 movies recorded, the frames were aligned for drift correction with the method described previously[32]. The images averaged from all 29 frames were used for initial model searching and structure refinement, while images averaged from the first 14 frames were used for calculation of the final density map. Defocus value was set to -2 μm during imaging session, and was determined with CTFFIND3[33] to be in the range of -0.5 to -3 μm in the images. Particles were picked with Ethan[34] and preprocessed with EMAN[35].

Asymmetric model generation

2D classification of particle images was performed with refine2d of EMAN. The resulted 2D class averages were used to calculate an initial model with the *starticos* command of EMAN[35]. The resulting reconstruction consists of two layers of smooth, featureless spherical shells, with the outer and inner layers corresponding to the averaged densities of

the viral capsid and packaged RNA, respectively. This initial reconstruction was used to run 3D classification of the images using Relion[36], with C1 symmetry applied. Within less than 10 iterations, one of the emerged 3D classes showed densities of separated RNA chains inside the capsid. 3D classification was then started over again with the converged RNA-containing structure as initial model. All of the resulting 10 classes showed similar structures with prominent RNA densities, suggesting that the RNA genome was indeed well organized inside the capsid of MS2 phage and the initial model was solid.

It is noteworthy that two recent efforts in cryoEM reconstruction of a segmented dsRNA virus relied on subtraction of the icosahedrally symmetric capsid contribution from the raw cryoEM images to minimize interference of the symmetric capsid in the orientation search for the asymmetric components[37,38]. Our success in generating the asymmetric model of MS2 without applying such computation-demanding subtraction methods suggests that the asymmetric RNA genome has sufficient power to drive asymmetric orientation search even in the presence of icosahedral capsid signal. This simpler computational strategy opens the door to modeling viral genomes and genome-capsid interactions in spherical viruses.

Structure refinement

The dataset was divided into two random halves, and refined separately against the Relion-generated model using Frealign[39], considering that the latter software demands much less computational resources in processing such a large dataset. The refinement procedure included five rounds of grid search (mode 3) with data points lower than 20Å resolution, followed by several iterations of local refinement (mode 1) with gradually increased resolution range. Particle images binned for 4x or 2x were used throughout the refinement procedure to speed up calculations. The un-binned particle images from averaging 14 of the 29 frames in the drift-corrected image stacks were used in the last few iterations of the refinement. A final density map was calculated by merging the two half datasets, containing a total number of 339,718 particles. The average resolution was determined based on the "gold-standard" FSC (Fourier shell correlation)=0.143 criterion[40]. Local resolutions were assessed with ResMap[41] (Extended Data Fig. 1b).

Atomic model building for proteins

Atomic model of the maturation protein was built *ab initio* with Coot[42]. To model the coat proteins, the crystallographic structure with PDB ID 2MS2[20] was fitted into the density map with Chimera[43], and then manually adjusted with Coot. Only 6 of the 178 copies of the coat protein needed partial modification. All the models were refined with Phenix real space refinement program[44].

Backbone tracing and atomic model building for RNA

In modeling the 16 high-resolution RNA stem-loops, we took advantage of the available RNA moiety model in the crystallographic structure 2B2G[45]. This model was first fitted into our cryoEM densities, mutated into the genuine sequence, and then manually adjusted in Coot. For tracing the backbone of the genome, we low-pass filtered the density map to 6Å resolution so that the RNA backbone became visible and was manually traced with the baton mode in Coot. For some RNA densities that are weak in the final density map (i.e., flexible)

but show better quality in some of the 10 classes from 3D classification, the density map of the best class was used to guide the backbone tracing.

Three constraints were used in conjunction to determine sequences of high-resolution stem-loops and to trace the genome backbone simultaneously. First of all, at 3.6Å resolution, purines and pyrimidines are readily distinguishable based on the fact that densities of purines are relatively fattier than those of pyrimidines as exemplified in Fig. 1d. Therefore, a degenerate sequence can be derived for each stem-loop: R=A/G or Y=C/U was assigned to each nucleotide (nt) in high quality regions, while N=A/G/C/U was assigned to nt that has poor density. This degenerate sequence was then searched against the MS2 genome with the JDSA program[46]. With a typical length of 9-12nt for a high quality stem-loop, and considering that bases in the stem region should be paired (G-U was counted as paired in addition to the Watson-Crick pairs), the degenerate search usually produced fewer than 3 candidates, and a single hit was not uncommon. Constraint from the backbone tracing was then used to further narrow down the candidate sequence and/or to confirm the assignment. For two tandem stem-loops in the path of the tracing, if their assigned sequences are also in tandem and have a distance in the genome agreeing with that in the tracing, then both assignments are likely correct. A third constraint we used to validate the sequence assignments and backbone tracing was the secondary structure prediction of the RNA sequence. Although the full-length MS2 genome of 3569 bases was too long for most RNA secondary structure prediction algorithms, we found that prediction results for sequences in the length of a few hundred bases are consistent among different prediction software and thus more reliable. Therefore, we divided the entire genome into several segments, predicted the secondary structure of each segment with the RNAfold webserver[47], and compared it with the corresponding backbone as identified by the assigned sequences of the high-resolution stem-loops contained in that segment. Matching of the predicted secondary structure and the traced backbone is evident for majority of the genome (Fig. 2c, d, Extended Data Fig. 3-7), thus confirming the validity of the backbone tracing and sequence assignments of the high-resolution stem-loops. From the correspondence between the traced backbone and the predicted secondary structure, we can also roughly assign sequences for most of the low-resolution stem-loops.

Modeling the high-resolution stem-loops individually was challenging due to the relatively low resolution of the genome structure. Ambiguity may arise from multiple results of the degenerate sequence search in some cases and also from generation of the degenerate sequence itself (i.e., ambiguity in assigning purine or pyrimidine to some of the nucleotides based on feature of the density at current resolution). Tracing backbone of the genome solely based on the 6Å resolution map is impractical because of numerous junctions and crossovers in the genome (see Fig. 2b for example). Our strategy combines the two levels of structural information with secondary structure prediction of the genome sequence to eliminate ambiguities by trial and error, dramatically improving reliability of the model. Such strategy is generally applicable for modeling genome organization in many other viruses.

Data availability

The cryoEM density map and the atomic models are deposited to the EMDB and PDB under the accession numbers EMD-8397 and 5TC1, respectively. The traced backbone model and the annotated secondary structure of the MS2 genome are available as supplementary information to the online version of the paper.

Extended Data
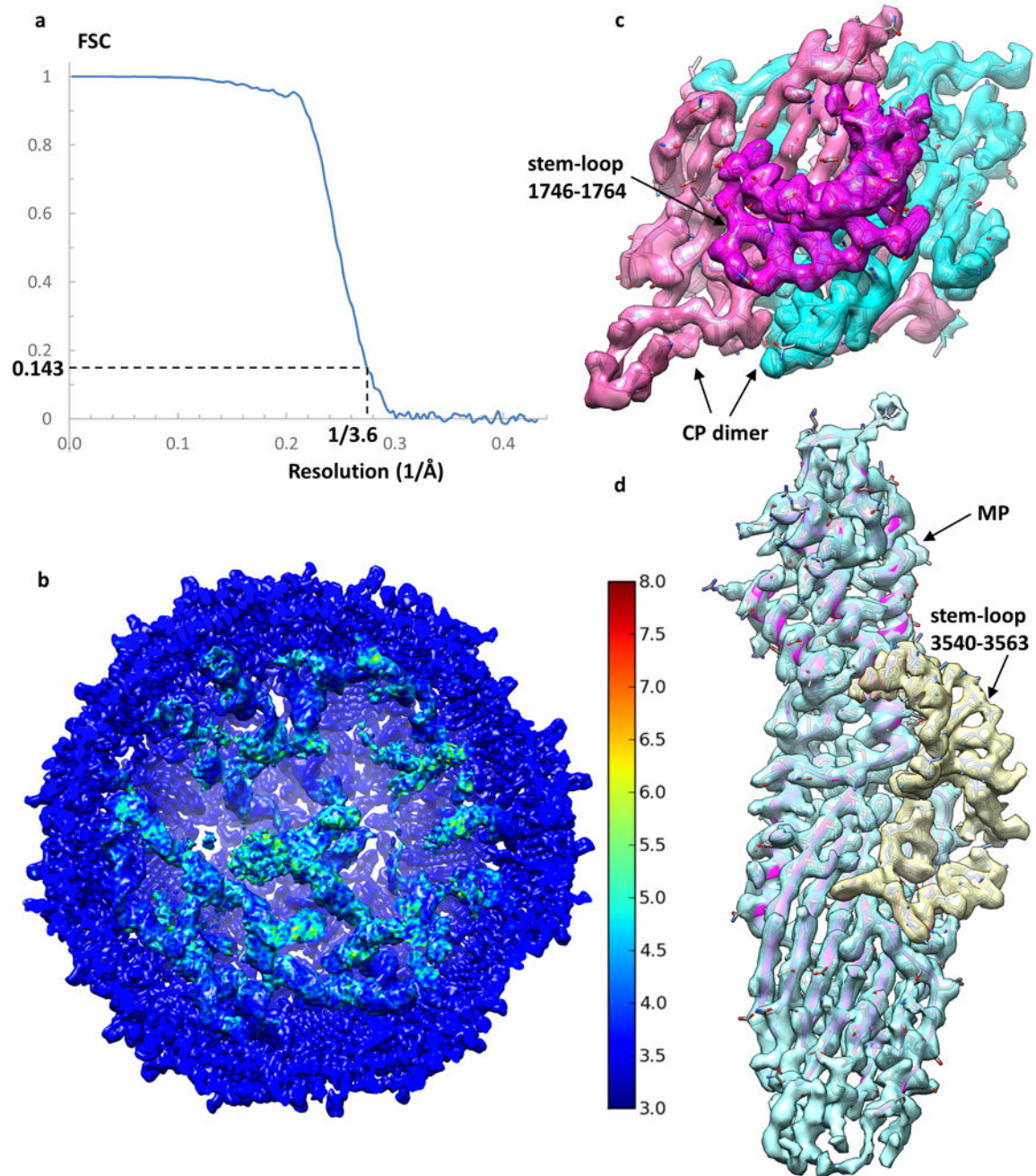
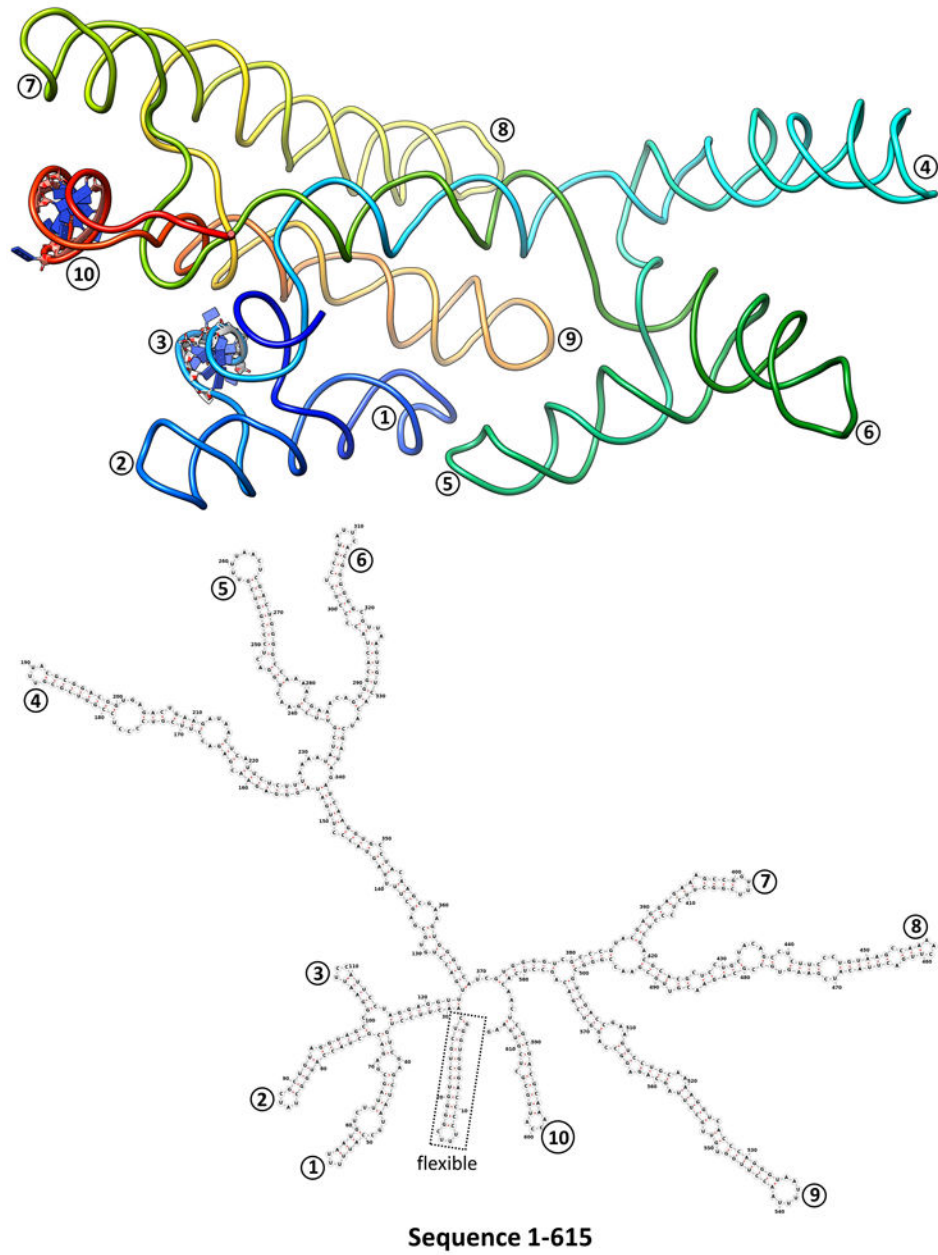**Extended Data Figure 1. Resolution assessment of the cryoEM reconstruction**

a, "Gold-standard" Fourier shell correlation (FSC) curve of the cryoEM reconstruction. The average resolution of the final density map is 3.6Å as determined by the FSC=0.143 criterion[40]. **b**, Local resolution assessed by ResMap[41]. Density voxels are colored according to their local resolution as defined in the color scale on the right side. Only half of the capsid is shown to expose the RNA densities inside. **c, d**, CryoEM densities of a coat protein dimer (**c**) or the maturation protein (**d**) with their bound RNA stem-loops to show quality of the density map. In both cases, the cryoEM densities are semitransparent to show the fitted atomic models of the protein and RNA.
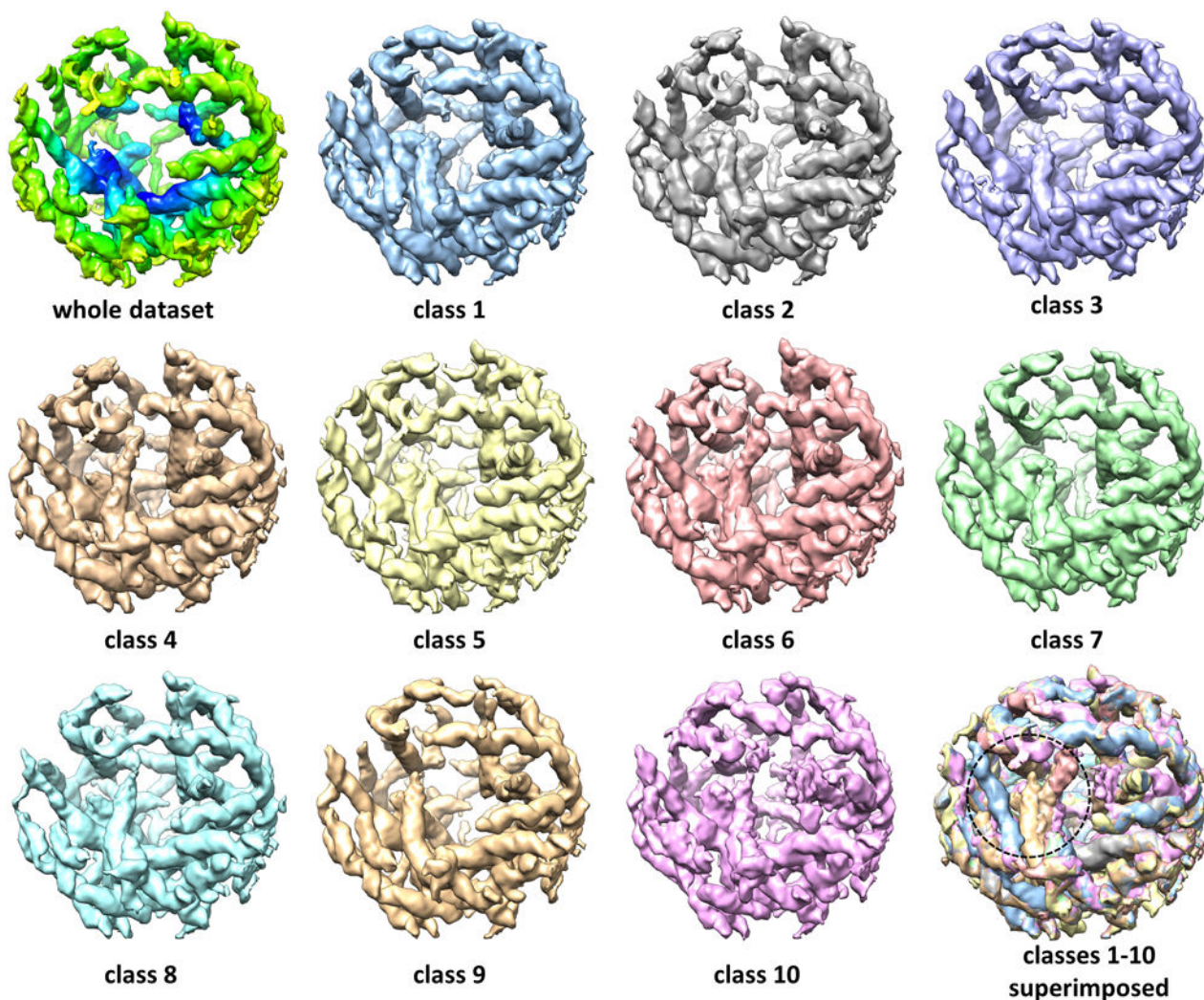
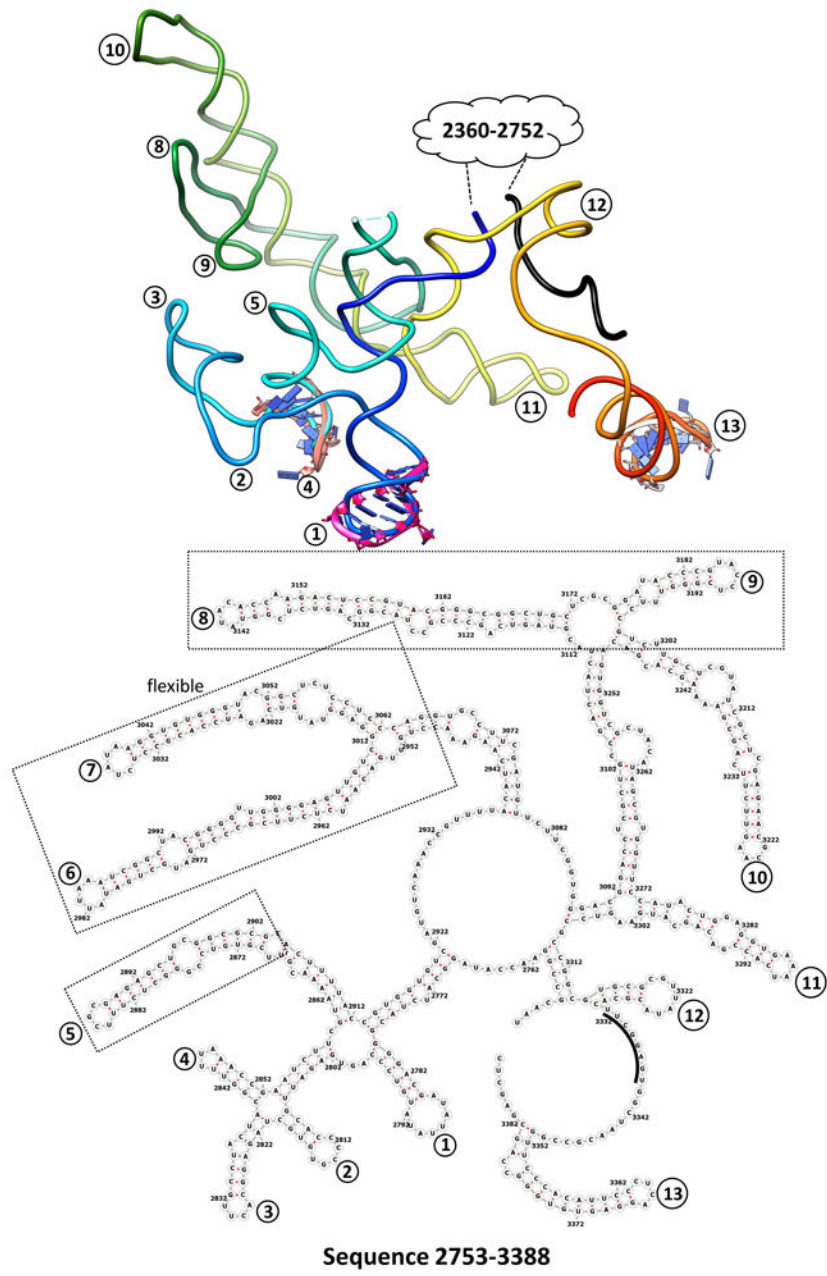**Extended Data Figure 2. 3D classification**

The entire dataset of the cryoEM images were subjected to 3D classification and refinement starting from a single initial model of the asymmetric reconstruction. Ten classes were arbitrarily set. The resulted density maps were compared with the reconstruction of the whole dataset and with each other. Overall structures of the ten classes are almost identical, except for small regions as exemplified by region enclosed in the dashed circle in the superimposed map. The RNA fragments of these regions have multiple conformations, and
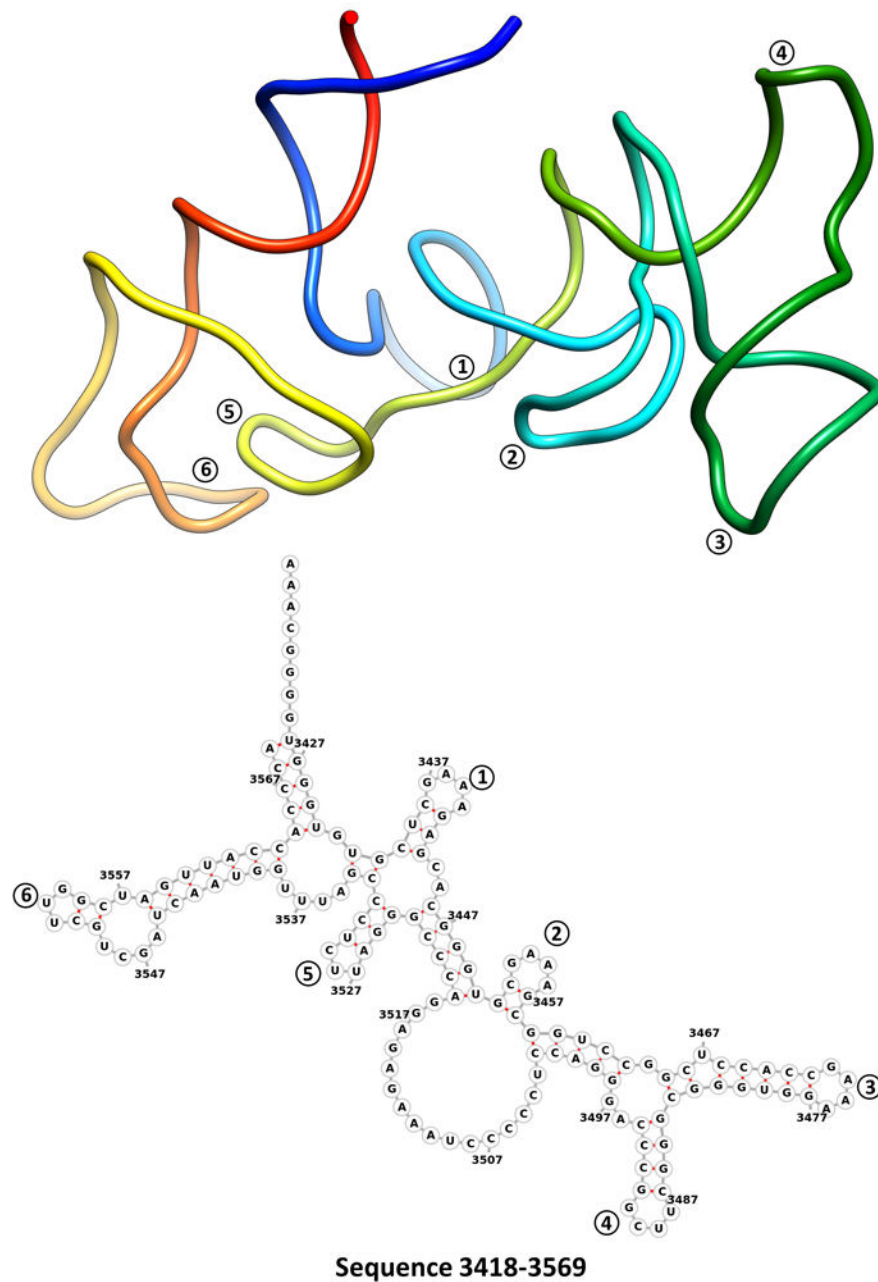
are thus not traced in our model. Overall, we were able to trace the RNA density amounting to 80% of the genome.


Sequence 1-615

Sequence 881-1290

Sequence 1711-2340

Sequence 2753-3388

Sequence 3418-3569
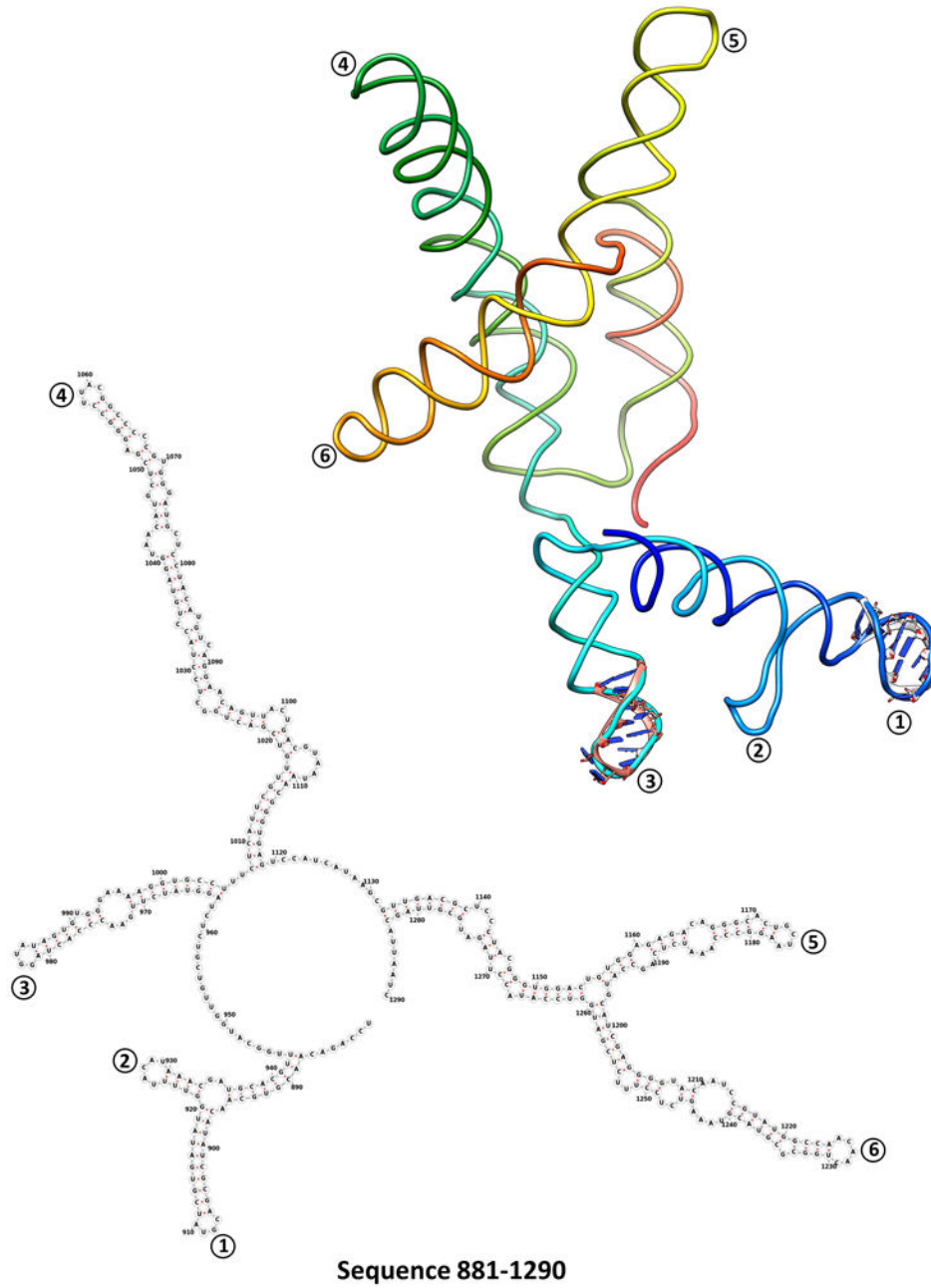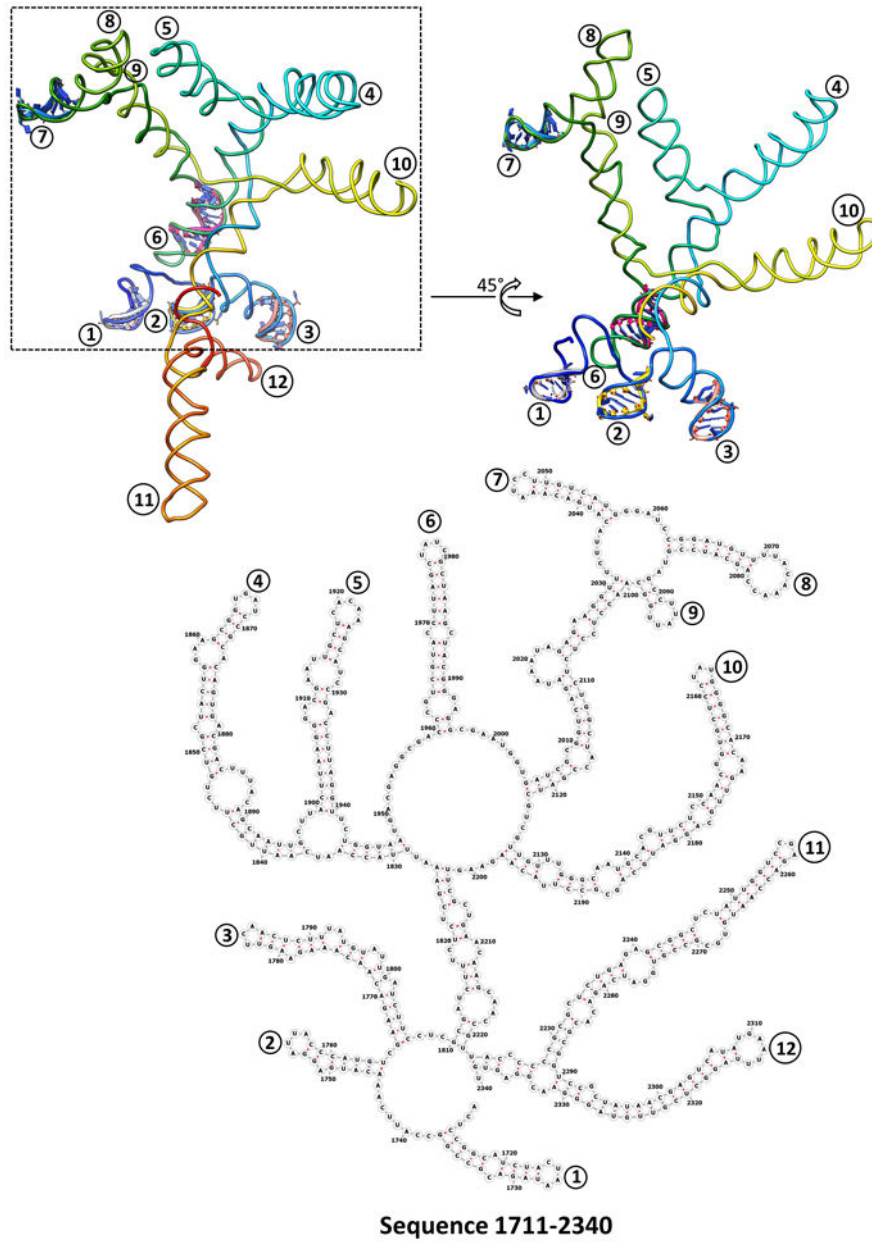
Extended Data Figures 3-7. Backbone model of the MS2 genome
Together with Fig. 2c, d, these figures show the backbone model of MS2 genome segment by segment from 5' to 3'. The backbone of each segment is rainbow-coloured (blue to red) from 5' to 3'. Atomic models of high-resolution stem-loops (ribbons) contained in the segment are also shown. Some of the base-pairings in the predicted secondary structure are modified to make it more consistent with the observed structure. Matching stem-loops in the backbone and in the predicted secondary structure are marked with the same number. Dashed boxes in some of the predicted secondary structure panels denote flexible stem-loops that are not well resolved in the cryoEM density map and thus not traceable for backbone.
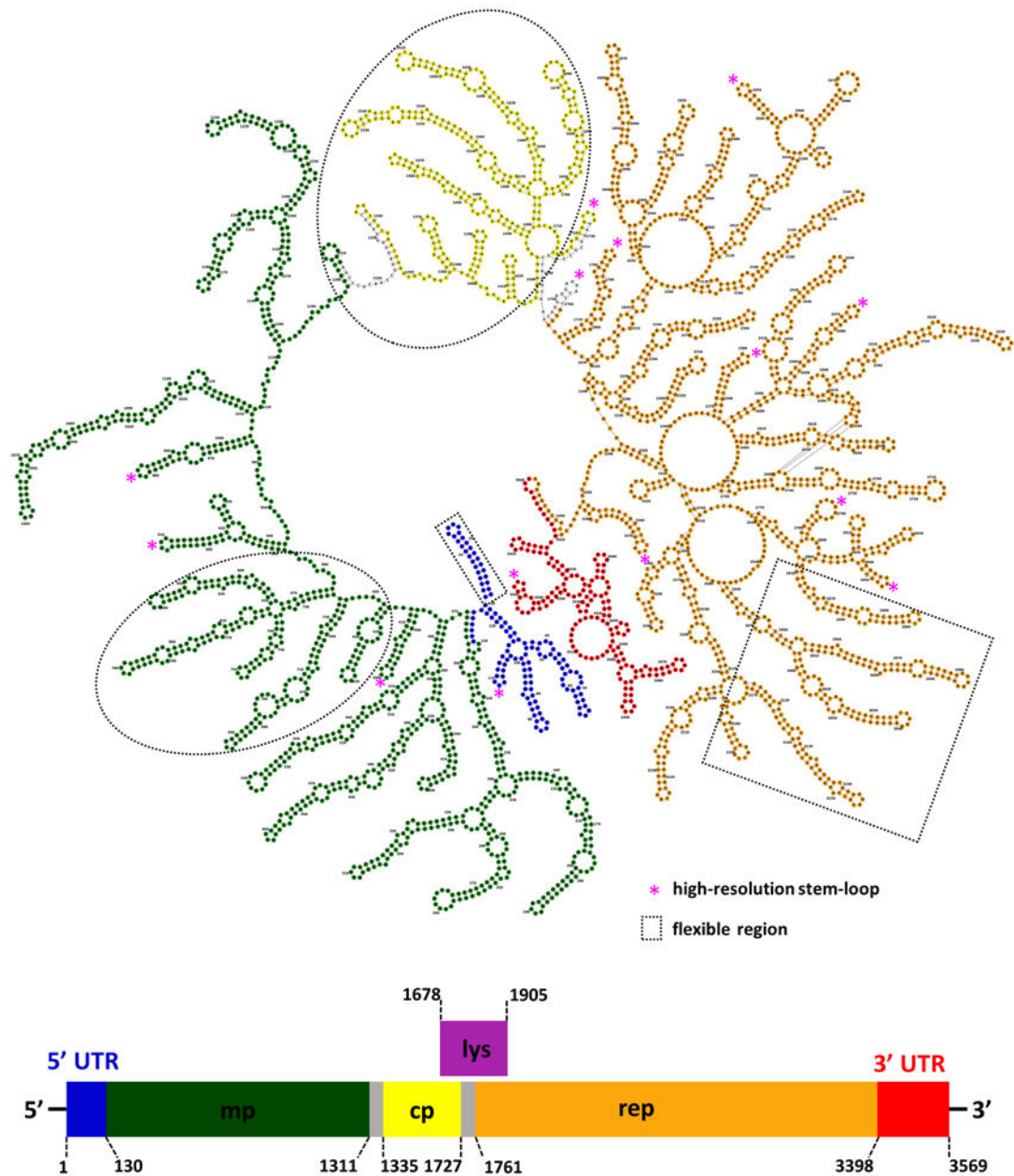
Extended Data Figure 8. Secondary structure of the MS2 genome
Secondary structures of all genome segments in Fig. 2d and Extended Data Fig. 3-7 are assembled to show the secondary structure of the entire MS2 genome. The genome sequences are coloured according to the genes encoded as depicted in the schematic diagram in the bottom, except for the lysis gene which is overlapping with the coat protein gene and the replicase gene. The star signs denote positions of the 16 high-resolution stem-loops. Segments enclosed with dotted boxes or ellipses are flexible.
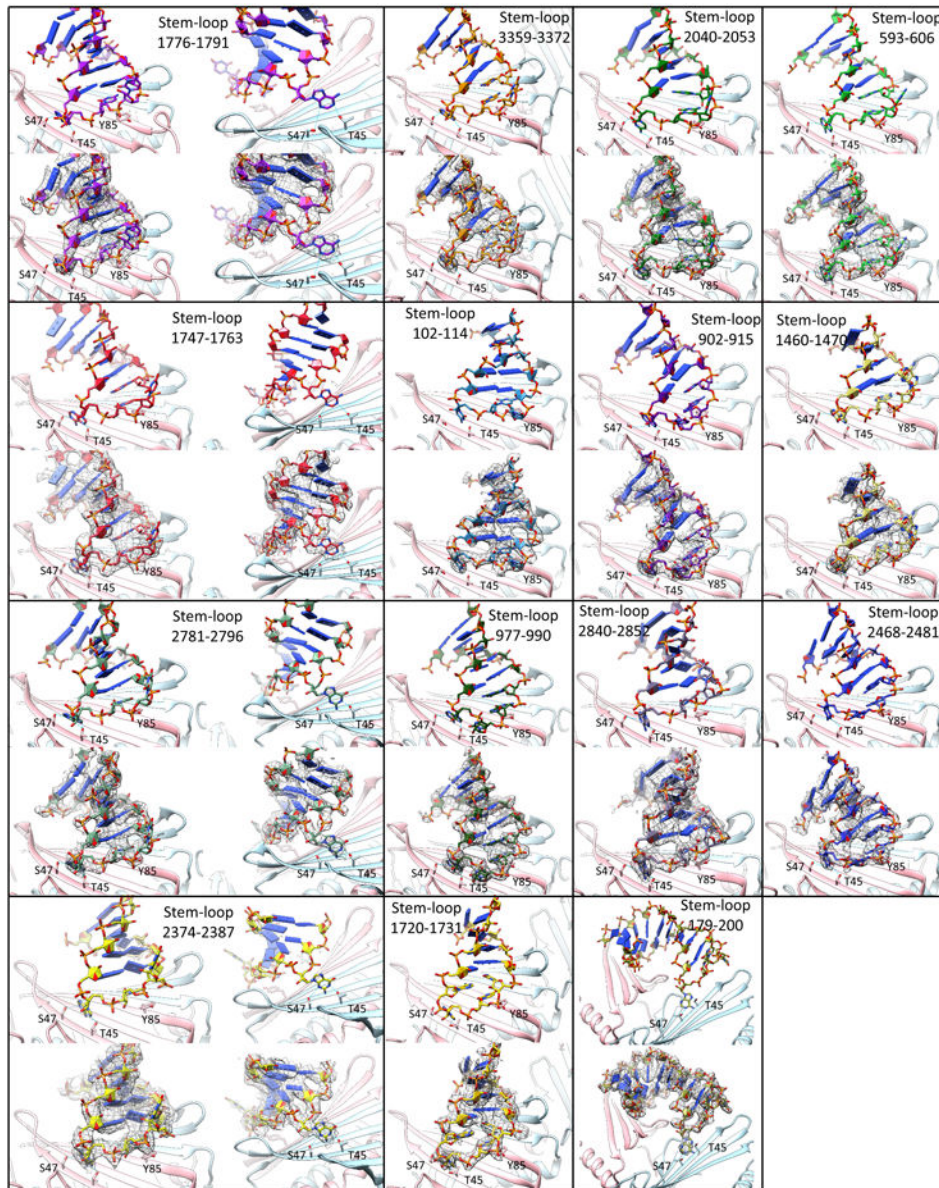
Extended Data Figure 9. CryoEM densities (mesh) and atomic models (stick) of the 15 high-resolution stem-loops that interact with CP-dimers (ribbon)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Jiang W, et al. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. Nature. 2006; 439:612–616. [PubMed: 16452981]

2. Lander GC, et al. The structure of an infectious P22 virion shows the signal for headful DNA packaging. Science. 2006; 312:1791–1795. [PubMed: 16709746]

3. Catalano, CE. Viral Genome Packaging Machines: Genetics, Structure, and Mechanism. Springer; US: 2005.

4. Basnak G, et al. Viral genomic single-stranded RNA directs the pathway toward a T=3 capsid. J Mol Biol. 2010; 395:924–936. DOI: 10.1016/j.jmb.2009.11.018 [PubMed: 19913556]

5. Sun S, Rao VB, Rossmann MG. Genome packaging in viruses. Curr Opin Struct Biol. 2010; 20:114–120. [PubMed: 20060706]

6. Perlmutter JD, Hagan MF. Mechanisms of virus assembly. Annu Rev Phys Chem. 2015; 66:217–239. [PubMed: 25532951]

7. Klug A. The tobacco mosaic virus particle: structure and assembly. Phil Trans Roy Soc Lond B. 1999; 354:531–535. [PubMed: 10212932]

8. Valentine RC, Strand M. Complexes of F-Pili and Rna Bacteriophage. Science. 1965; 148:511–513. [PubMed: 14263773]

9. Fiers W, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature. 1976; 260:500–507. [PubMed: 1264203]

10. Schmidt BF, Berkhout B, Overbeek GP, van Strien A, van Duin J. Determination of the RNA secondary structure that regulates lysis gene expression in bacteriophage MS2. J Mol Biol. 1987; 195:505–516. [PubMed: 3656423]

11. Poot RA, Tsareva NV, Boni IV, van Duin J. RNA folding kinetics regulates translation of phage MS2 maturation gene. Proc Natl Acad Sci U S A. 1997; 94:10110–10115. [PubMed: 9294171]

12. Valegard K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L. Crystal structure of an RNA bacteriophage coat protein-operator complex. Nature. 1994; 371:623–626. [PubMed: 7523953]

13. Convery MA, et al. Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. Nat Struct Biol. 1998; 5:133–139. [PubMed: 9461079]

14. Rowsell S, et al. Crystal structures of a series of RNA aptamers complexed to the same protein target. Nat Struct Biol. 1998; 5:970–975. [PubMed: 9808042]

15. Borodavka A, Tuma R, Stockley PG. Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proc Natl Acad Sci U S A. 2012; 109:15769–15774. [PubMed: 23019360]

16. Patel N, et al. Revealing the density of encoded functions in a viral RNA. Proc Natl Acad Sci U S A. 2015; 112:2227–2232. [PubMed: 25646435]

17. Rolfsson O, et al. Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. J Mol Biol. 2016; 428:431–448. [PubMed: 26608810]

18. Valegard K, Liljas L, Fridborg K, Unge T. The three-dimensional structure of the bacterial virus MS2. Nature. 1990; 345:36–41. [PubMed: 2330049]

19. Dent KC, et al. The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. Structure. 2013; 21:1225–1234. [PubMed: 23810697]

20. Golmohammadi R, Valegard K, Fridborg K, Liljas L. The refined structure of bacteriophage MS2 at 2.8 A resolution. J Mol Biol. 1993; 234:620–639. [PubMed: 8254664]

21. Dykeman EC, Stockley PG, Twarock R. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. J Mol Biol. 2013; 425:3235–3249. [PubMed: 23763992]

22. Witherell GW, Gott JM, Uhlenbeck OC. Specific interaction between RNA phage coat proteins and RNA. Proc Natl Acad Sci U S A. 1991; 40:185–220.

23. Plevka P, Tars K, Liljas L. Crystal packing of a bacteriophage MS2 coat protein mutant corresponds to octahedral particles. Protein Sci. 2008; 17:1731–1739. [PubMed: 18662904]

24. Toropova K, Stockley PG, Ranson NA. Visualising a viral RNA genome poised for release from its receptor complex. J Mol Biol. 2011; 408:408–419. [PubMed: 21376055]

25. Danziger RE, Paranchych W. Stages in phage R17 infection. 3. Energy requirements for the F-pili mediated eclipse of viral infectivity. Virology. 1970; 40:554–564. [PubMed: 4908731]

26. Clarke M, Maddera L, Harris RL, Silverman PM. F-pili dynamics by live-cell imaging. Proc Natl Acad Sci U S A. 2008; 105:17978–17981. [PubMed: 19004777]

27. Krahn PM, O'Callaghan RJ, Paranchych W. Stages in phage R17 infection. VI. Injection of A protein and RNA into the host cell. Virology. 1972; 47:628–637. [PubMed: 4551992]

28. Kozak M, Nathans D. Fate of maturation protein during infection by coliphage MS2. Nat New Biol. 1971; 234:209–211. [PubMed: 4942983]

29. Silverman PM, Valentine RC. The RNA injection step of bacteriophage f2 infection. J Gen Virol. 1969; 4:111–124. [PubMed: 4975637]

30. Shiba T, Miyake T. New type of infectious complex of E. coli RNA phage. Nature. 1975; 254:157–158. [PubMed: 1090846]

References in Methods

31. Suloway C, et al. Automated molecular microscopy: the new Leginon system. Journal of structural biology. 2005; 151:41–60. [PubMed: 15890530]

32. Li X, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nature methods. 2013; 10:584–590. [PubMed: 23644547]

33. Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of structural biology. 2003; 142:334–347. [PubMed: 12781660]

34. Kivioja T, Ravantti J, Verkhovsky A, Ukkonen E, Bamford D. Local average intensity-based method for identifying spherical particles in electron micrographs. Journal of structural biology. 2000; 131:126–134. [PubMed: 11042083]

35. Ludtke SJ, Baldwin PR, Chiu W. EMAN: semiautomated software for high-resolution single-particle reconstructions. Journal of structural biology. 1999; 128:82–97. [PubMed: 10600563]

36. Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. Journal of structural biology. 2012; 180:519–530. [PubMed: 23000701]

37. Zhang X, et al. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. Nature. 2015; 527:531–534. [PubMed: 26503045]

38. Liu H, Cheng L. Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. Science. 2015; 349:1347–1350. [PubMed: 26383954]

39. Grigorieff N. FREALIGN: high-resolution refinement of single particle structures. Journal of structural biology. 2007; 157:117–125. [PubMed: 16828314]

40. Rosenthal PB, Henderson R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. Journal of molecular biology. 2003; 333:721–745. [PubMed: 14568533]

41. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. Nature methods. 2014; 11:63–65. [PubMed: 24213166]

42. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta crystallographica Section D, Biological crystallography. 2010; 66:486–501. [PubMed: 20383002]

43. Pettersen EF, et al. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004; 25:1605–1612. [PubMed: 15264254]

44. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta crystallographica Section D, Biological crystallography. 2010; 66:213–221. [PubMed: 20124702]

45. Horn WT, et al. Structural basis of RNA binding discrimination between bacteriophages Qbeta and MS2. Structure. 2006; 14:487–495. [PubMed: 16531233]

46. Lim CY, et al. The MTE, a new core promoter element for transcription by RNA polymerase II. Genes & development. 2004; 18:1606–1617. [PubMed: 15231738]

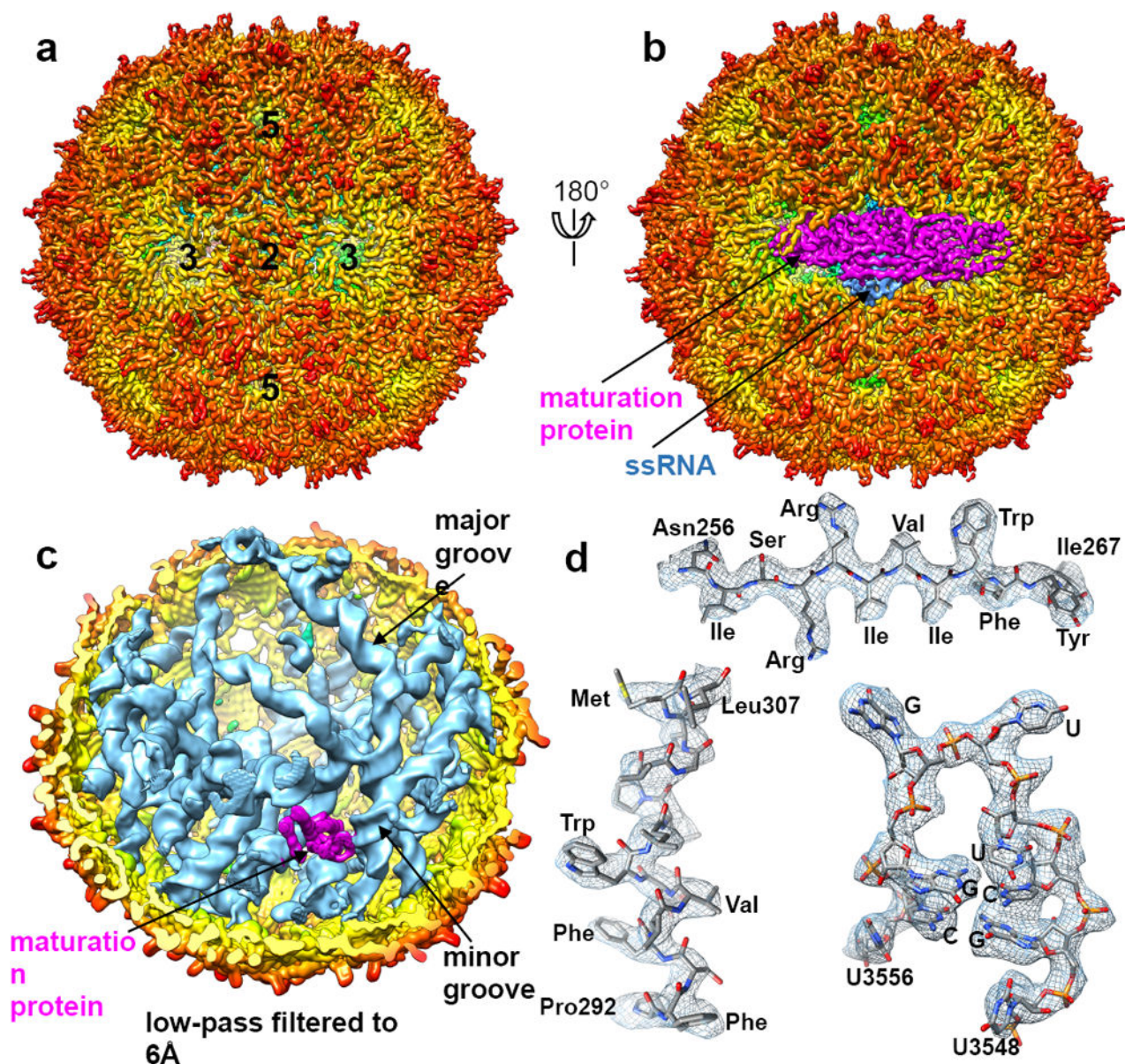47. Lorenz R, et al. ViennaRNA Package 2.0. Algorithms for molecular biology. 2011; 6:26. [PubMed: 22115189]

Figure 1. CryoEM asymmetric reconstruction of MS2 at 3.6Å resolution
a, b, Front (**a**) and back (**b**) views of the cryoEM density map along an icosahedral 2-fold symmetry axis with some 2-, 3- and 5-fold axes indicated. The capsid shell is radially coloured with MP highlighted in magenta, and ssRNA genome inside the capsid in blue. **c**, Cut-open view with half of the capsid shell removed to expose the genome. **d**, Segmented cryoEM densities (mesh) superimposed with their corresponding atomic models (sticks). Top and bottom-left panels are typical β-strand and α-helix densities from MP, respectively. The bottom-right panel is part of an MP-bound RNA stem-loop. Purines and pyrimidines are readily distinguishable.
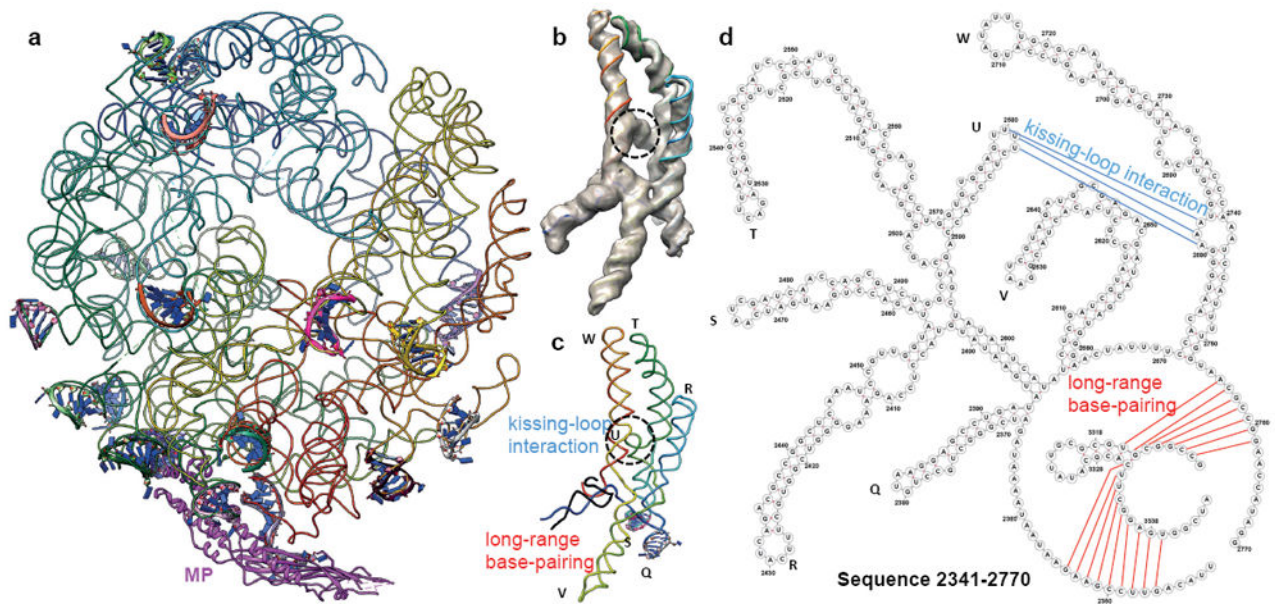
Figure 2. Modeling the ssRNA genome
a, Backbone structure of the genome (wire) and non-uniform distribution of the high-resolution stem-loops (ribbons). Backbone of the genome is rainbow-coloured (blue to red) from 5' to 3'. **b-d**, An example of tracing RNA backbone. Part of the genome density (grey in **b**) is segmented out and superimposed with its backbone model (rainbow-coloured wire, blue to red from 5' to 3') (**b** and **c**). For each of the two high-resolution stem-loops (ribbons in **c**) contained in this segment, a degenerate sequence was derived based on the resolved bases and used to search against the genome to identify sequence candidates. Each of these short sequence candidates was expanded in both directions to include ~500 bases for secondary structure prediction. The predicted secondary structure was then correlated with the backbone obtained in **b** and only one of these sequence candidates yielded correct sequence registration of individual stem-loops (indicated by numbers 1-7 in **c** and **d**). The backbone model reveals kissing-loop and long-range base-pairing interactions as indicated.
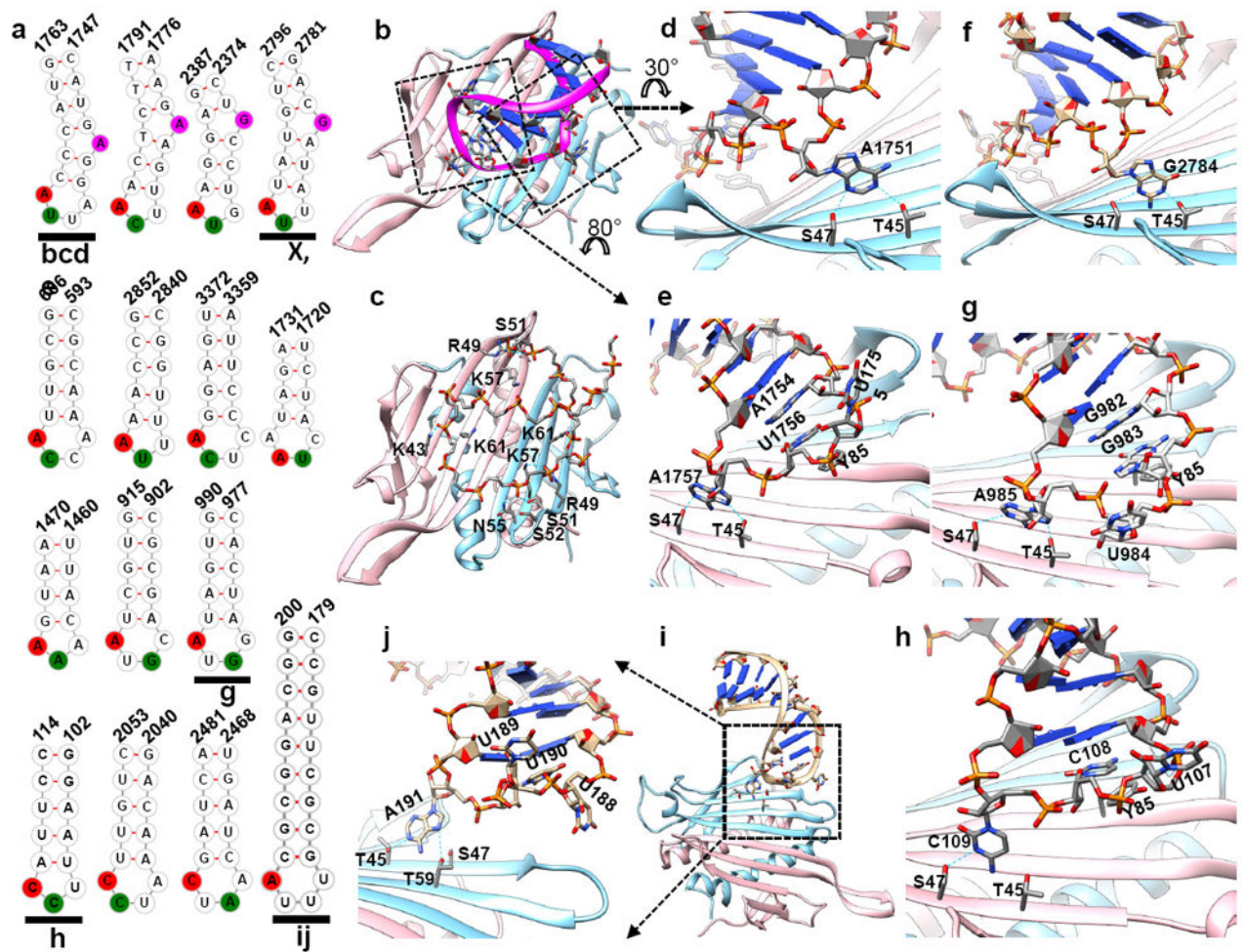
Figure 3. Conserved interaction motifs between RNA stem-loops and CP-dimers
a, Secondary structures of the RNA stem-loops with nucleotides involved in the three types of conserved interaction motifs coloured. Letters beneath some of the stem-loops identify panels in which the atomic model for that stem-loop is shown. **b-e**, Atomic model of stem-loop 1747-1763 and its interactions with a CP-dimer (pink and sky blue ribbons). In **c**, positively-charged or polar residues of the CP-dimer interacting with phosphates of the RNA backbone (sticks) are indicated. Zoom-in views of the stem (**d**) and loop (**e**) regions show the interaction motifs conserved among the 15 stem-loops. **f-j**, Accommodation of diversities in sequence or local environment. Stem-loop 2781-2796 (**f**), viewed in the same orientation as in **d**, shows that a G forms the same kind of hydrogen bonds with Thr45 and Ser47 as an A in **d**. Stem-loops 977-990 (**g**) and 102-114 (**h**), viewed in the same orientation as in **e**, show that a purine (G983 in **g**) instead of a pyrimidine (U1756 in **e**) stacks with Tyr85 and that a pyrimidine (C109 in **h**) forms only one hydrogen bond instead of a purine (A1757 in **e**) forming two hydrogen bonds with Thr45 and Ser47. Stem-loop 179-200 (**i, j**) binds to a CP-dimer from a very different angle due to steric hindrance of a neighboring stem-loop (not shown). Nonetheless, the RNA fold and one of the three interaction motifs are conserved, although hydrogen bonding with Thr59 instead of Ser47 is formed.
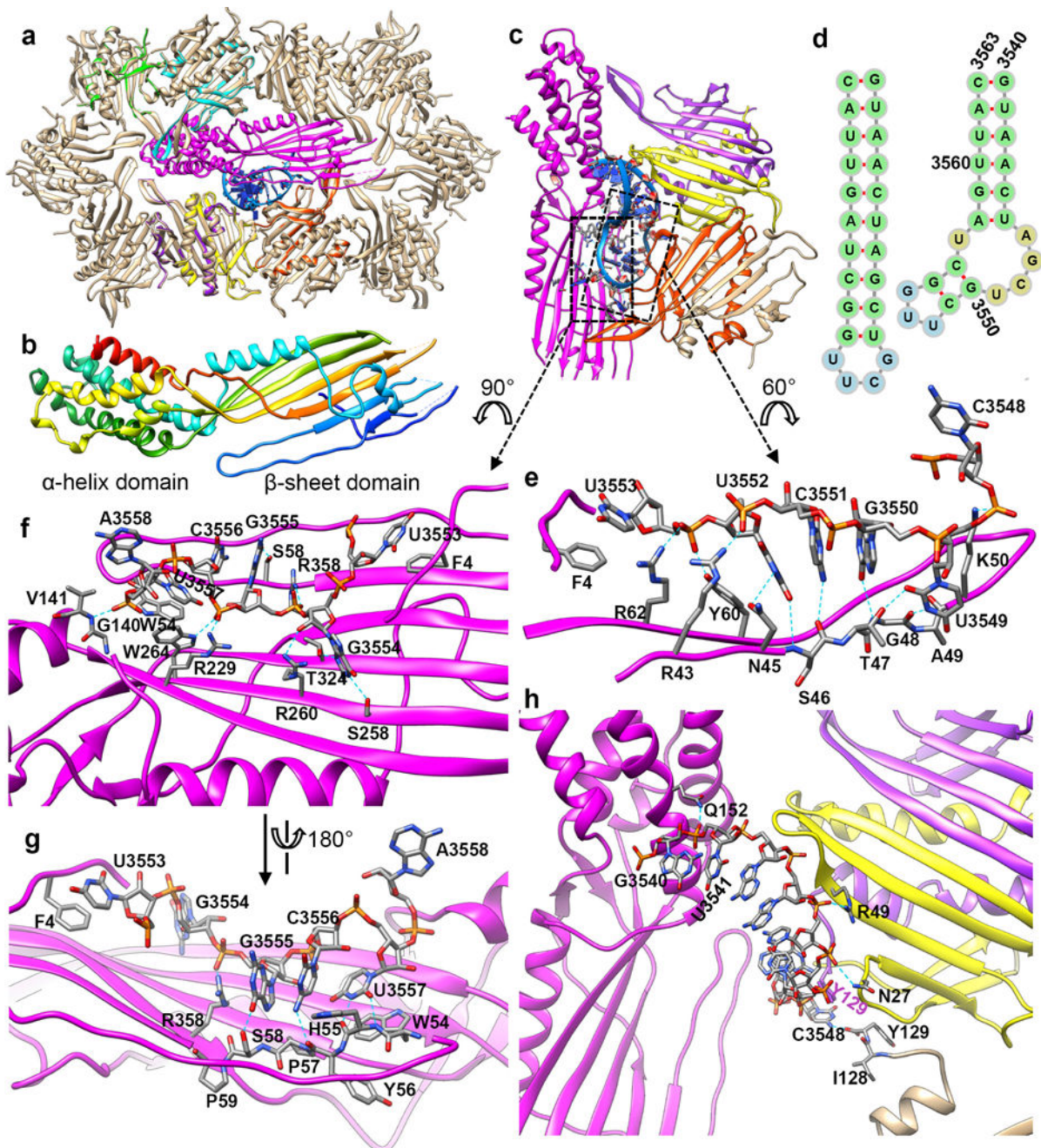
Figure 4. MP and its interactions with the 3'-end stem-loop
a, Incorporation of MP in the capsid shell. MP (magenta) replaces a CP-dimer at a 2-fold symmetry axis and induces structural changes of neighboring CPs. Atomic models of the changed CPs (coloured ribbons) are superimposed with that of CPs at other 2-fold symmetry axes (beige ribbons) which are unaffected by the MP. **b**, MP model rainbow-coloured (blue to red from N- to C-terminus). **c**, Binding of the 3'-end stem-loop to MP and neighboring CPs as viewed inside the capsid. **d**, Base-pairing in the 3'-end stem-loop as observed in our structure interacting with MP (right), or theoretically as free stem-loop (left). **e-h**, Details of

the interactions between the 3'-end stem-loop and the MP or CPs. Panel **h** is a zoom-in view of the RNA stem region, with half of the stem hidden for clarity.
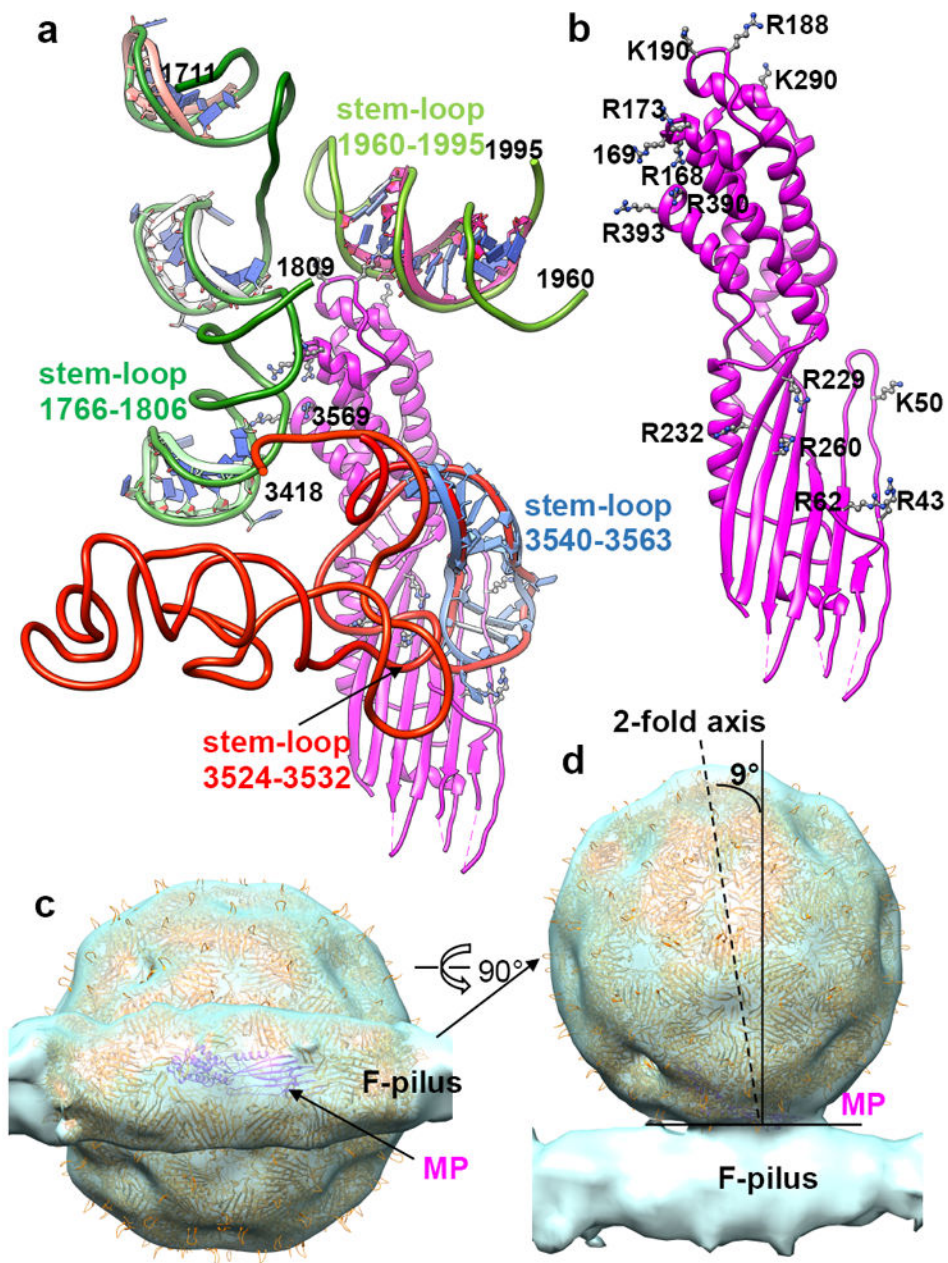
Figure 5. Binding of MP to the genome and to bacterial F-pilus

a, Overview of MP (magenta) with surrounding RNA stem-loops (wires coloured as in Fig. 2a) shown in the same orientation as in Fig. 3c. **b**, Same as **a** but without the RNA backbone model, showing the distribution of positively-charged arginine and lysine residues in MP that bind RNA stem-loops. **c, d**, Fitting our atomic models (ribbons) of the MS2 virion into a tomographic reconstruction (EMD-2365, semitransparent surface) of MS2 attached to bacterial F-pilus[19]. MP obliquely projects out from the capsid surface, resulting in a slight tilt (~9°) of the MS2 virion when attached to the F-pilus (**d**).