

ARTICLE

Received 8 Feb 2016 | Accepted 11 Jul 2016 | Published 26 Aug 2016

DOI: 10.1038/ncomms12524

OPEN

Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure *in situ*

Roman I. Koning^{1,2}, Josue Gomez-Blanco³, Inara Akopjana⁴, Javier Vargas³, Andris Kazaks⁴, Kaspars Tars⁴, José María Carazo³ & Abraham J. Koster^{1,2}

In single-stranded ribonucleic acid (RNA) viruses, virus capsid assembly and genome packaging are intertwined processes. Using cryo-electron microscopy and single particle analysis we determined the asymmetric virion structure of bacteriophage MS2, which includes 178 copies of the coat protein, a single copy of the A-protein and the RNA genome. This reveals that *in situ*, the viral RNA genome can adopt a defined conformation. The RNA forms a branched network of stem-loops that almost all allocate near the capsid inner surface, while predominantly binding to coat protein dimers that are located in one-half of the capsid. This suggests that genomic RNA is highly involved in genome packaging and virion assembly.

¹Department of Molecular Cell Biology, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, The Netherlands. ²Netherlands Centre for Electron Nanoscopy, Institute of Biology Leiden, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands. ³Biocomputing Unit, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CNB-CSIC), Darwin 3, Campus Universidad Autónoma, Cantoblanco, Madrid 28049, Spain. ⁴Biomedical Research and Study Centre, Ratsupites 1, LV-1067 Riga, Latvia. Correspondence and requests for materials should be addressed to R.I.K. (email: r.i.koning@lumc.nl).

Bacteriophage MS2 (ref. 1) is a species of the *Levivirus* genus in the *Leviviridae* family of small, positive-sense, single-stranded ribonucleic acid (RNA) bacteriophages that infect their host via adsorption to bacterial pili. The MS2 virion consists of an RNA genome, encapsidated by a $T=3$ shell, containing coat protein (CP) and a single copy of the maturation or so-called A-protein (AP), which attaches to the viral RNA and binds the host receptor. The MS2 genome is one of the smallest known, comprising just 3,569 nucleotides, and was the first genome—of any life form—to be completely sequenced². The genome encodes four proteins: maturation, coat, lysis and replicase, which are translated at different levels and time points during the bacteriophage life cycle. The RNA adopts a specific secondary structure with many double-stranded regions, which play essential roles in translational regulation and replication³. Extensive research has shown that access to ribosomal binding sites for translation is regulated by short and long range base-pairing⁴ and binding of RNA stem-loops (SLs) to CP dimers. More specifically, maturation gene translation can only take place during the first stages of synthesis of a new plus strand⁵. The start of the lysis gene is suppressed by a local hairpin and accessed occasionally by ribosomal back scanning and frame shifting of ribosomes that finished translation of the CP gene⁶. Expression of the replicase gene, and therefore replication, is controlled by base pairing with a coding region of the CP^{7,8}, and downregulated by binding of a SL containing the start codon, called the ‘translational operator’ (TR), to a CP dimer (CP)₂⁹. In addition, the RNA secondary structure is involved in replicase and AP binding and resistance to host RNAses³.

Although MS2 capsids spontaneously assemble *in vitro* from CP alone at high enough concentration¹⁰, its RNA plays an important role in virion formation *in vivo*. Encapsidation of its own genome depends on presence of the AP¹¹, which specifically binds two RNA sequences at the 3′-untranslated region (nucleotides 388–414) and the 5′-maturation region (nucleotides 3,510–3,527)¹². Second, capsid formation is highly promoted by interaction of the TR binding to CP₂ (ref. 9). Third, specific SL–CP₂ interactions form packaging signals (PSs) promote genome packaging^{13,14}.

Atomic models of the MS2 capsid¹⁵, capsid protein dimer¹⁶ and the translation suppressing 19-nucleotide RNA TR hairpin loop bound to a CP₂ (ref. 17) were determined using X-ray crystallography, revealing sequence-specific RNA–protein interactions essential for binding. The structure of aptamers showed that multiple RNA sequences can bind to MS2 CP₂ (refs. 17–22). Icosahedral cryo-electron microscopy (cryo-EM) reconstructions of the MS2 virion, including genome and AP, showed that the operator–CP₂ interaction is not unique, and that the whole genome is highly connected to the capsid via SLs^{23,24}. Asymmetric single particle²⁴ and tomography²⁵ cryo-EM reconstructions of the MS2 virion suggested that the genome within MS2 might adopt a specific tertiary structural conformation, although the resolution of these maps was too low (~4–6 nm) to resolve any recognizable tertiary RNA structures.

Following these investigations, we here determined using high-resolution single particle cryo-EM the asymmetric structure of bacteriophage MS2, to a resolution of 8.7 Å. The map outlines the AP, which replaces one CP dimer. Moreover, it shows an ordered genome that is shaped as a branched network of connected RNA stem-loops, of which the majority interacts with the inside of the capsid. The RNA–CP interactions are primarily located on one side of the capsid, which might have consequences for genome packaging and virion assembly.

Results

Asymmetric structure. The outside of our asymmetric EM map of the MS2 virion (Fig. 1a) at 8.7 Å (EMD-3403) is similar to the

previously reported capsid protein X-ray structure (protein data bank (pdb) entry: 2MS2) (ref. 15 and our icosahedrally averaged cryo-EM capsid structure at 4.1 Å (EMD-3402) (Supplementary Figs 1 and 2, Supplementary Movie 5), showing the characteristic openings in the capsid at the five- and three-fold symmetry axes and the small protrusions, formed by amino acid loops, connecting beta-sheet strands A and B, on the capsid surface (dark blue in Fig. 1a,b). More importantly, in the asymmetric virion structure the single copy of the AP is clearly resolved in the capsid (yellow, Fig. 1a–d, Supplementary Movie 1) and replaces one CC conformer-type CP₂ in the capsid, which thus contains 178 copies of the CP. The AP forms a 9 nm long ‘handle’ (~1.8 nm in diameter) extending outwards with a ~30° angle from the surface from a two-fold symmetry axis position in the capsid, extending over a neighbouring hole at a three-fold axis.

The RNA. The inside of our asymmetric MS2 map reveals that the genome within the MS2 capsid has a defined tertiary structure, forming an intricate three-dimensional (3D) branched network of interconnected SLs (Fig. 1c–e, Supplementary Movie 2). The RNA density accounts for 95% of the calculated mass of the full genome. A less ordered region in the map can explain the missing 5%. In total, 59 SL structures were discriminated in the genome, which were all connected to each other, of which the majority (53, 90%) ended near the capsid (Supplementary Fig. 3), while 6 ends were centrally located. To quantify and assess the PSs, the interactions of the RNA SL with CP₂, TR–CP₂ X-ray model (PDB entry: 1ZDZH) was fitted into the EM density map. In total, 44 SLs (83% of the SLs near the capsid and 75% of all identified SLs) ended at a CP₂ RNA binding site, 2 SLs interacted with the AP (Fig. 1d) and 7 were not located near a CP₂ interface. Close examination of these 44 binding sites showed that at 23 sites the RNA from the X-ray model fit the EM RNA density (Supplementary Fig. 4), with some of them showing in detail the interactions between nucleotides A – 4 and A – 10 in the 19-nucleotide RNA chain of the RNA hairpin loop (1ZDZH) that both make contact to Thr45 and Ser47 (not shown) located in the beta-sheets of different CP molecules (Fig. 1f). The observed amounts of SLs (59) and PSs (between 23 and 44) match earlier estimates, predicting 53 SLs and 35 PSs, that were based on the predicted RNA secondary structure, obtained by phylogenetic analysis and experimental probing of RNA in solution, and analysis of potential binding SLs¹⁴. Our results show that the majority of the many SL RNA structures that are present in the genome of MS2 bind the CP₂ in the capsid. Since these SLs all have different predicted sequences, as a result this suggests that a wide variety of different RNA sequences actually bind to CP₂ *in situ* in an overall similar conjunction as the TR, similar to what was observed with several aptamer–CP₂ interactions^{18–20}. At several sites RNA hairpin-EM densities were also observed near the capsid but in deviating conformations or sites, often rotated (7 cases) or shifted (7 cases), compared with the TR structure (data not shown). In addition, several stem structures were sideways associated to CP₂ sites. Therefore alternative, non-sequence specific, binding modes of the RNA to the CP₂ likely exist, including binding to other amino acids in the capsid.

The obtained resolution of the EM map is sufficient to visualize the double-stranded (ds) RNA SLs, including its helical nature; however, single RNA strands are not clearly resolved and therefore it was not possible to trace predicted secondary structures into the 3D map and to allocate predicted SLs in the MS2 genome into the density^{14,26}. To investigate all individual SL–CP₂ interactions in the genome to the observed SLs in the EM map a higher resolution structure would be

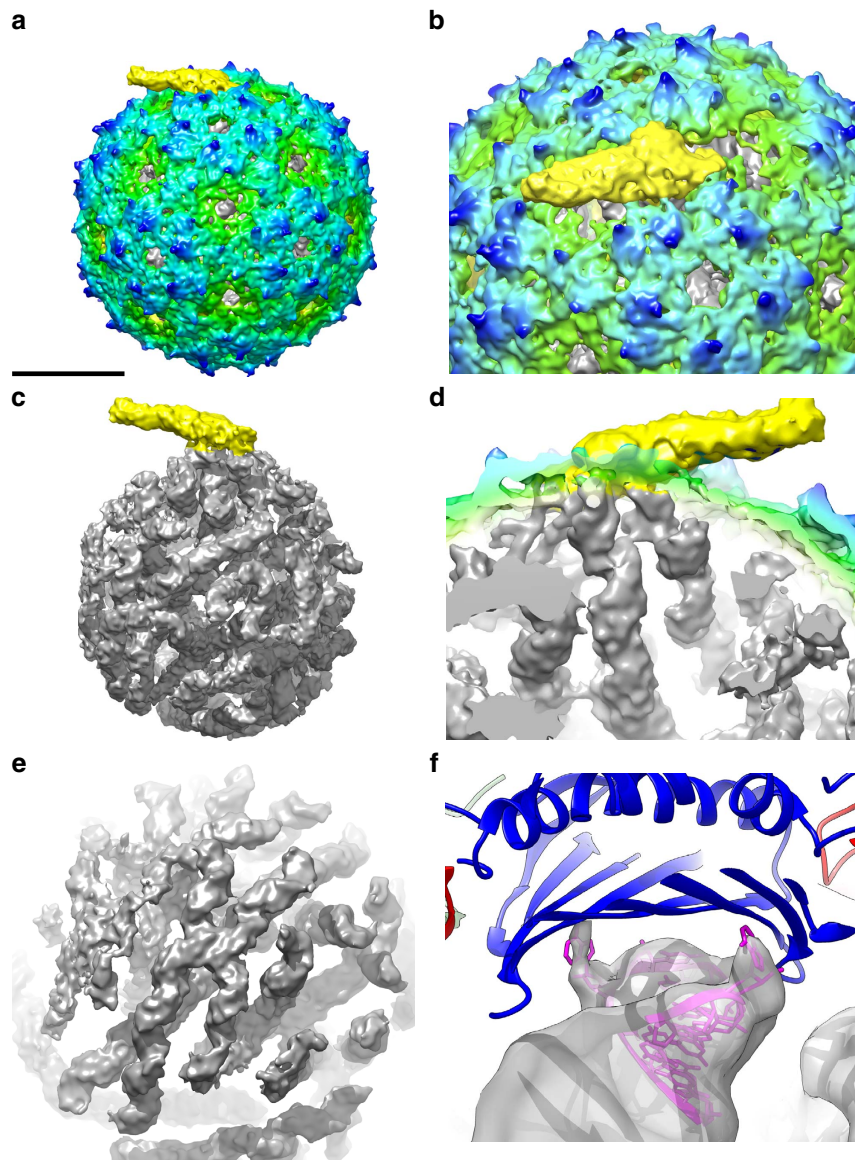


Figure 1 | The asymmetric reconstruction of bacteriophage MS2. Asymmetric structure of bacteriophage MS2 (green-blue radially coloured) shows the AP (yellow) (a), which replaces one CP dimer (b). Inside the protein capsid a structured genome (grey) is present (c) that is connected to the AP (d). The reconstruction shows the double-stranded helices in the stem loop structures (e). At some positions individual NAs connecting to the capsid are resolved, as shown by fitting of the X-ray structure of the 19-nucleotide TR (magenta) bound to the capsid (blue) (pdb:1ZDH) in the EM density (grey) (f). Scale bar is 100 Å.

required. Nevertheless, the two regions known to connect to the AP¹² could putatively be allocated in the EM map. The 3' RNA end, including nucleotides 3,510–3,527, forms a very specific repeat of short SLs that binds adjacently along the AP, while near the 5' RNA the SL formed by nucleotides 388–414 binds the AP (Fig. 2a, Supplementary Movie 3).

RNA–protein interactions. While the AP ensures specific packaging of its coding RNA and TR–CP₂ interactions promote capsid formation, the question is how the SL–CP₂ interactions that we observed here influence virion assembly. To explore a potential role of the RNA genome structure in this process we investigated the distribution of PSs, interactions of SLs with CP₂, over the capsid. From the 89 CP₂, 44 (49%) have a connecting SL, 33 (37%) has crossing (ds) RNA density, while 9 (10%) does not have any adjacent density. Notably, these 44 PSs are distributed unevenly over the capsid, being localized predominantly on one

side of the capsid (Supplementary Fig. 3). This uneven distribution was even more pronounced for the 23 RNA SLs of which the density matched the TR X-ray model (Supplementary Fig. 4). Of these, 19 (82%) were bound to dimers that were located on one of half of the capsid in only three CP₂ pentamers (Fig. 2b, Supplementary Movie 4). These pentamers were adjacent to the AP and its two binding SLs.

This uneven distribution of SL–CP₂ interactions supports a two-step encapsidation model^{13,27} in which RNA condensation proceeds full capsid formation. The role of the asymmetric distribution of RNA–protein interactions could be two-fold. Multiple SLs could increase efficient capsid formation by CP₂ recruitment from the surroundings to form the first CP₂ pentamers, adding both efficiency and localization to CP₂–CP₂ interactions that drive coat formation. Alternatively, CP₂ could induce SLs formation and condensation of the MS2 RNA, thereby inducing genome compaction during encapsidation. These two

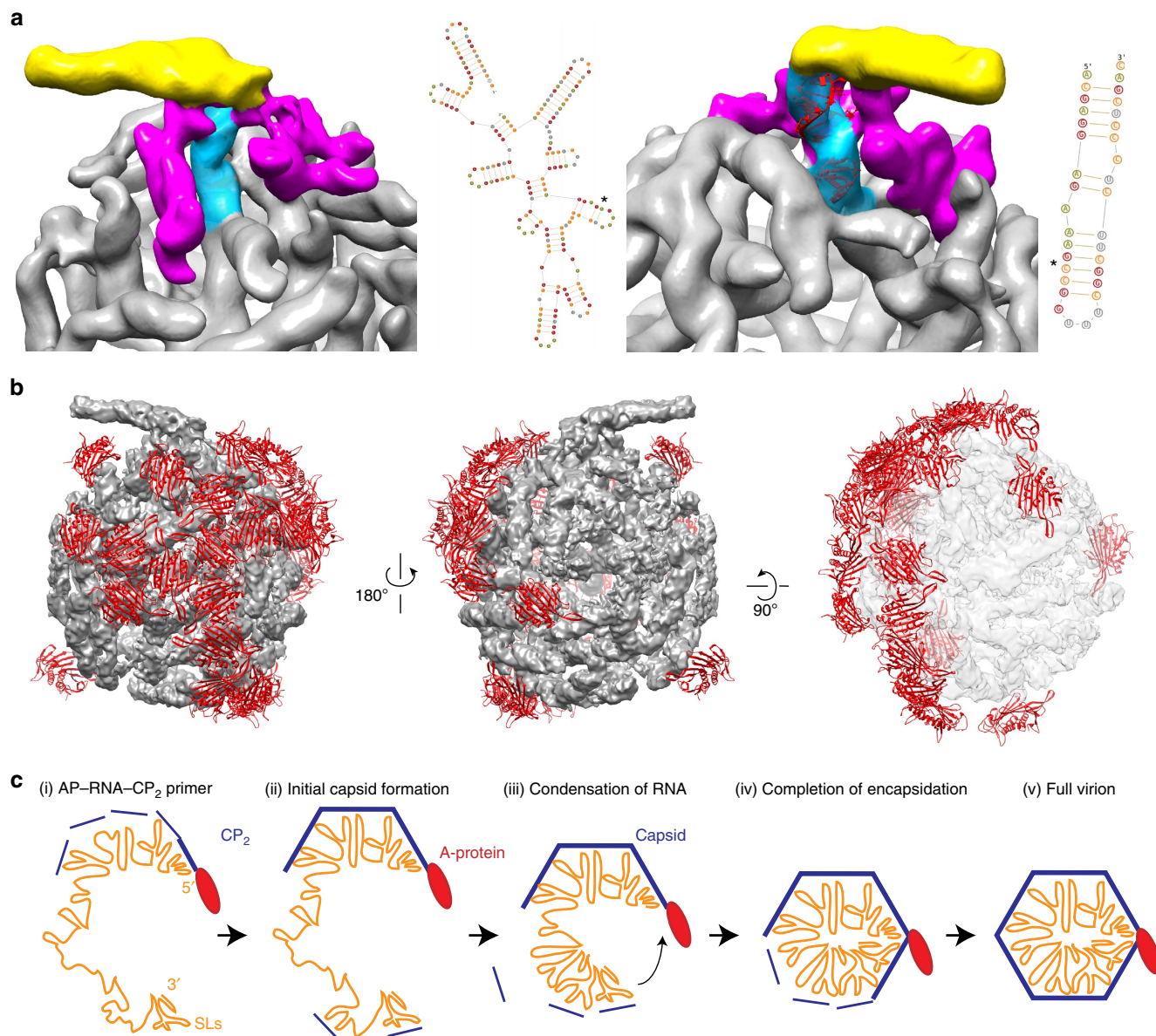


Figure 2 | Proposed RNA densities of known AP binding sites. Left: nucleotides 3,510–3,527 (magenta) and the RNA secondary structure prediction (asterisk marks binding site). Right: nucleotides 388–414 (blue) with fit in density (red) and the secondary structure of the SL (a). Asymmetric distribution of SL-CP₂ interactions; the 23 capsid dimers (red) to which the bound RNA (grey) SLs fit the TR are predominantly localized at one side of the capsid (b). Proposed model for MS2 virion assembly: AP-RNA-(CP₂)_n complex forms primer (i) to which CP₂ dimers bind and start to form the capsid (ii), CP₂ is both being recruited by existing SLs and induce new SLs in the RNA (iii) CP₂ induced refolding and binding of the 5' end AP binding site results in condensation of the RNA (iv) that enables efficient packaging and formation of a stable virion (v) (c).

possible functions are not mutually exclusive and also support the suggested formation of new SLs in the MS2 genome¹⁴.

Discussion

We reconstructed the MS2 virion, revealing the AP and the single conformation of the RNA genome *in situ*. The genome is intimately and asymmetrically linked to the capsid. The presence of a single AP, breaking the symmetry in the capsid, in MS2 is exceptional among viruses and might play an important role in the uniquely structured genome, which might not appear in other (small) viruses. Even so structures of several other viruses have shown hints of (partly) ordered (ds)RNA^{28,29}. It remains to be seen whether asymmetric cryo-EM single particle reconstructions of other viruses would reveal similar genome ordering inside

virions. The potential of cryo-EM to explore asymmetric structure determination of complete virions, including their genome, would provide unprecedented details on viral RNA structures, RNA-protein interactions and insight into viral assembly, which is valuable knowledge for drug design by targeting disruptions of viral genome folding and packaging.

Methods

Purification of phages. An overnight culture of *E. coli* strain XL1 blue was grown at 37 °C in LB medium until an OD₆₀₀ of 0.5. Calcium chloride was added to a final concentration of 2 mM and the cells were infected with phage MS2 at a multiplicity of 10 and incubated for another three hours. Lysates with phage titres of approximately 1×10^{-12} were used for purification. Cellular debris was removed by centrifugation and phage concentrated by ultrafiltration using 15 ml 100 kDa cutoff Amicon concentrators. Phage was further purified by gel-filtration

on sephacryl S500 column. Fractions were inspected for purity on coomassie-stained SDS-PAGE gel and negative staining EM.

Specimen preparation. Aliquots of purified MS2 were applied to glow-discharged holey carbon film supported by copper grids (Quantifoil R2/2) after slow discharging with negative polarity for 1 min at 30 mA using a K950X carbon coater (Emitech). Grids were vitrified by plunging into a liquid propane/ethane mixture (2:1 v/v), which was cooled by liquid nitrogen. Samples were plunged using a Leica EM GP from room temperature and blotted for 1–2 s using filter paper. After vitrification, the grids were stored in liquid nitrogen until use.

Data collection. Data acquisition was performed on a Titan Krios transmission electron microscope (FEI) operated with Cs correction at 300 keV using EPU automated single particles acquisition software (FEI). Seven frames per images were recorded on a back-thinned Falcon II detector at a nominal magnification of 59,000 \times with a sampling size of 1.14 Å per pixel.

Image processing. Image processing was performed using Scipion platform (<http://scipion.cnb.csic.es>), which is an integrative image processing framework that currently mainly uses Xmipp (<http://xmipp.cnb.csic.es/>)¹, Relion (http://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page)², Spider (<http://spider.wadsworth.org>)³ and EMAN (<http://blake.bcm.edu/emanwiki/EMAN2>)⁴ packages. Graphics were produced by UCSF Chimera (<http://www.cgl.ucsf.edu/chimera>)⁵. All movies were aligned using Optical Flow approach⁶, while contrast transfer functions (CTFs) were estimated using CTFIND3 (ref. 7), and were used to select the best quality micrographs. A total of 22,441 particles were picked automatically using Xmipp⁸, which were further screened⁹ extracted and normalized. First, an icosahedrally symmetrized reconstruction was calculated. Particles were classified using 2D reference-free Relion approach, and a subset of 18,977 particles was selected from the best classes. Then, these particles were used for Relion 3D refinement, using standard parameters for viral particles, as icosahedral symmetry and gold-standard approach. A second round of refinement was performed, now without applying any symmetry, using Xmipp projection matching. As initial model the final icosahedrally symmetrized map obtained with the previous refinement of Relion was used, filtered to 25 Å. The first four iterations were performed as global refinements decreasing the sampling search angle. Also, the whole particle set was randomly split in two halves. Each subset was refined with the same conditions as the whole set.

Data analysis and visualization. Fitting and visualization was performed using UCSF Chimera⁵. Structures of the MS2 capsids, full and missing one dimer, were both created from the X-ray structures of the MS2 capsid protein (pdb: 2MS2 (ref. 10)), and of the MS capsid protein including the operator hairpin loop with (pdb model 1ZDH¹¹). Capsids were manually aligned with the EM density and fine aligned using the fitmap command, normalized, density maps at the map resolution were created using the molmap command, which were then subtracted from the EM density using the vop command to create a map of the RNA and A-protein only (EMD-3404).

Data availability. Density maps of the icosahedral reconstruction, the asymmetric reconstruction, and the RNA + AP map, a difference map from the asymmetric reconstruction and the dimer depleted capsid structure, are available from the Electron Microscopy Data Bank with accession codes: EMD-3402, EMD-3403, and EMD-3404, respectively. The authors declare that all data supporting the findings of this study are available from the corresponding author upon request.

References

- Davis, J. E., Strauss, J. H. & Sinsheimer, R. Bacteriophage MS2: another RNA phage. *Science* **134**, 1427 (1961).
- Fiers, W. *et al.* Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**, 500–507 (1976).
- Olsthoorn, R. J. & van Duin, J. Bacteriophages with ssRNA. *Encycl. Life Sci.* <http://dx.doi.org/10.1002/9780470015902.a0000778.pub3> (2011).
- Licis, N., Van Duin, J., Balklava, Z. & Berzins, V. Long-range translational coupling in single-stranded RNA bacteriophages: an evolutionary analysis. *Nucleic Acids Res.* **26**, 3242–3246 (1998).
- Poot, R., Tsareva, N. V., Boni, I. V. & van Duin, J. RNA folding kinetics regulates translation of phage MS2 maturation gene. *Proc. Natl Acad. Sci.* **94**, 10110–10115 (1997).
- Adhin, M. R. & van Duin, J. Scanning model for translational reinitiation in eubacteria. *J. Mol. Biol.* **213**, 811–818 (1990).
- van Himbergen, J., van Geffen, B. & van Duin, J. Translational control by a long range RNA-RNA interaction; a basepair substitution analysis. *Nucleic Acids Res.* **21**, 1713–1717 (1993).
- Jou, W. M., Haegeman, G., Ysebaert, M. & Fiers, W. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature* **237**, 82–88 (1972).
- Bernardi, A. & Spahr, P.-F. Nucleotide sequence at the binding site for coat protein on RNA of bacteriophage R17. *Proc. Natl Acad. Sci.* **69**, 3033–3037 (1972).
- Sugiyama, T., Hebert, R. R. & Hartman, K. A. Ribonucleoprotein complexes formed between bacteriophage MS2 RNA and MS2 protein *in vitro*. *J. Mol. Biol.* **25**, 455–463 (1967).
- Hung, P. P., Ling, C. M. & Overby, L. R. Self-assembly of Q-beta and MS2 phage particles: possible function of initiation complexes. *Science* **166**, 1638–1640 (1969).
- Shiba, T. & Suzuki, Y. Localization of A protein in the RNA-A protein complex of RNA phage MS2. *Biochim. Biophys. Acta* **654**, 249–255 (1981).
- Borodavka, A., Tuma, R. & Stockley, P. G. A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol.* **10**, 481–489 (2013).
- Dykeman, E. C., Stockley, P. G. & Twarock, R. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J. Mol. Biol.* **425**, 3235–3249 (2013).
- Valegård, K., Liljas, L., Fridborg, K. & Unge, T. The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36–41 (1990).
- Ni, C. Z. *et al.* Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure* **3**, 255–263 (1995).
- Valegård, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**, 623–626 (1994).
- Convery, M. A. *et al.* Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nat. Struct. Biol.* **5**, 133–139 (1998).
- Horn, W. T. *et al.* The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: further challenges in the modeling of ligand-RNA interactions. *RNA* **10**, 1776–1782 (2004).
- Rowell, S. *et al.* Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nat. Struct. Biol.* **5**, 970–975 (1998).
- Valegård, K. *et al.* The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J. Mol. Biol.* **270**, 724–738 (1997).
- Van Den Worm, S. H. E. *et al.* Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. *Nucleic Acids Res.* **26**, 1345–1351 (1998).
- Toropova, K., Basnak, G., Twarock, R., Stockley, P. G. & Ranson, N. A. The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* **375**, 824–836 (2008).
- Koning, R. *et al.* Visualization by cryo-electron microscopy of genomic RNA that binds to the protein capsid inside bacteriophage MS2. *J. Mol. Biol.* **332**, 415–422 (2003).
- Dent, K. C. *et al.* The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. *Structure* **21**, 1225–1234 (2013).
- Van Den Worm, S. *RNA-protein interactions in the ss-RNA phages* (Leiden University, 2004).
- Borodavka, A., Tuma, R. & Stockley, P. G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc. Natl Acad. Sci.* **109**, 15769–15774 (2012).
- Larson, S. B. *et al.* Double-helical RNA in satellite tobacco mosaic virus. *Nature* **361**, 179–182 (1993).
- Zhang, X. *et al.* *In situ* structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* **527**, 531–534 (2015).

Acknowledgements

We thank Sjoerd van den Worm, Rene Olsthoorn, Janis Rumnieks and Christoph Diebolder for discussions and reading of the manuscript. This research has been executed with support of NeCEN, the Netherlands Centre for Electron Nanoscopy, Leiden, NL and NWO, the Netherlands Organisation for Scientific Research and is partly financed by the European Regional Development Fund of the European Commission. We acknowledge the Spanish Ministry of Economy and Competitiveness through Grants AIC-A-2011-0638, BIO2013-44647-R. This work was funded by Instruct (Grant reference PID: 1114), part of the European Strategy Forum on Research Infrastructures (ESFRI) in collaboration with the Instruct Image Processing Center I2PC. This work was supported by NanoNextNL of the Government of the Netherlands and 130 partners.

Author contributions

R.I.K. conceived the project, collected and analysed the data and wrote the manuscript, J.G.B. and J.V. performed image processing, analysed the data, and wrote the manuscript I.A. and A.K. purified the MS2, K.T., J.M.C. and A.J.K. supervised the studies.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Koning, R. I. *et al.* Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure *in situ*. *Nat. Commun.* 7:12524 doi: 10.1038/ncomms12524 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016