doi:10.1016/j.jmb.2011.02.039 J. Mol. Biol. (2011) 408, 399–407

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

COMMUNICATION

# Simple Rules for Efficient Assembly Predict the Layout of a Packaged Viral RNA

E. C. Dykeman[1], N. E. Grayson[1], K. Toropova[2], N. A. Ranson[2], P. G. Stockley[2] and R. Twarock[1]*

[1]York Centre for Complex Systems Analysis, University of York, York YO10 5DD, UK
[2]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

*Received 17 December 2010; received in revised form 15 February 2011; accepted 17 February 2011*
Available online 25 February 2011

***Edited by M. F. Summers***

***Keywords:***
ssRNA virus;
virus assembly;
genome organization;
cryo-EM;
Hamiltonian path

Single-stranded RNA (ssRNA) viruses, which include major human pathogens, package their genomes as they assemble their capsids. We show here that the organization of the viral genomes within the capsids provides intriguing insights into the highly cooperative nature of the assembly process. A recent cryo-electron microscopy structure of bacteriophage MS2, determined with only 5-fold symmetry averaging, has revealed the asymmetric distribution of its encapsidated genome. Here we show that this RNA distribution is consistent with an assembly mechanism that follows two simple rules derived from experiment: (1) the binding of the MS2 maturation protein to the RNA constrains its conformation into a loop, and (2) the capsid must be built in an energetically favorable way. These results provide a new level of insight into the factors that drive efficient assembly of ssRNA viruses *in vivo*.



A cryo-electron microscopy (cryo-EM) analysis of the wild-type MS2 virus at ~9 Å resolution with icosahedral symmetry averaging previously revealed a characteristic polyhedral distribution of RNA close to the protein layer.[1] Based on this observation, we were able to narrow down the vast number of combinatorially possible ways in which the genome could organize itself within the capsid to just over 40,500 possibilities. The remarkable efficiency of virus assembly begs the question of whether all of these organizations are equally probable or whether there are further constraints that result in the RNA taking on an organization within the particle that favors a particular RNA fold or specific capsid assembly pathway.

Addressing this question becomes possible based on the data presented in this issue by Toropova *et al.* where they describe a cryo-EM reconstruction of bacteriophage MS2 bound to its receptor.[2] The contact between the virus and its receptor, the bacterial F-pilus, is mediated by a single copy of maturation protein,[3–5] which was found to sit at a 5-fold vertex of the $T=3$ capsid. This allowed a structure of the receptor-bound virus to be determined with C5 (5-fold rotational) symmetry. Density at lower radii than the coat protein layer of the virion is organized into two distinct shells, which must correspond to maturation protein, the genomic RNA, or a combination of the two. The outer shell is similar (albeit with different averaging) to *in vitro* assembled virus-like particles that lack maturation protein[6] and, therefore, most likely represents protein-bound RNA. Icosahedral symmetry averaging of the data determined with C5 symmetry[2] results in a

*Corresponding author. E-mail address: reidun.twarock@york.ac.uk.

Present address: K. Toropova, Department of Structural Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260, USA.

Abbreviations used: ssRNA, single-stranded RNA; cryo-EM, cryo-electron microscopy; 1-D, one-dimensional.

distribution of material inside the protein layer that is very similar to the previously determined cryo-EM reconstructions of the wild-type virion and reassembled virus-like particles that have been reported using icosahedral symmetry averaging.[1,7] Since there is only one copy of maturation protein and since it is of significantly smaller mass than the RNA, this similarity in density adjacent to the coat protein layer must mostly reflect the interaction of the RNA with the coat protein subunit.

The density distribution obtained with C5 symmetry provides vital clues on how the RNA is distributed along the 5-fold axis on which the maturation protein is located. It allows us to predict, using a simple mathematical model, the asymmetric organization of the linear single-stranded RNA (ssRNA) molecule within its symmetric container. Strikingly, our results imply that only a very limited number of RNA configurations out of the over 40,500 ones identified in Ref. 1 are consistent with the 5-fold-averaged cryo-EM data, biochemical information regarding RNA interaction with maturation protein, and efficient capsid assembly. This provides new insights not only into the asymmetric organization of the genomic material but also into the sophisticated assembly strategies of the virus. For example, it shows that, besides having a role during infection, the maturation protein could also be used to vastly reduce the complexity of the assembly process by circularizing the genomic RNA.

## Cryo-EM data imply a dimer switching model for virus assembly

Icosahedrally averaged cryo-EM reconstructions of Leviviridae show that the outer shells of their genomic RNAs form distinct, cage-like structures.[1,7] An example is bacteriophage MS2, with a cage akin to the polyhedron in Fig. 1a. This polyhedron is formed from two distinct types of edges (Fig. 1b): 60 short edges, occurring in groups of five around the particle 5-fold axes, and 30 long edges, which cross underneath the 2-fold axes and connect the short edges of two neighboring 5-fold axes. Previous work

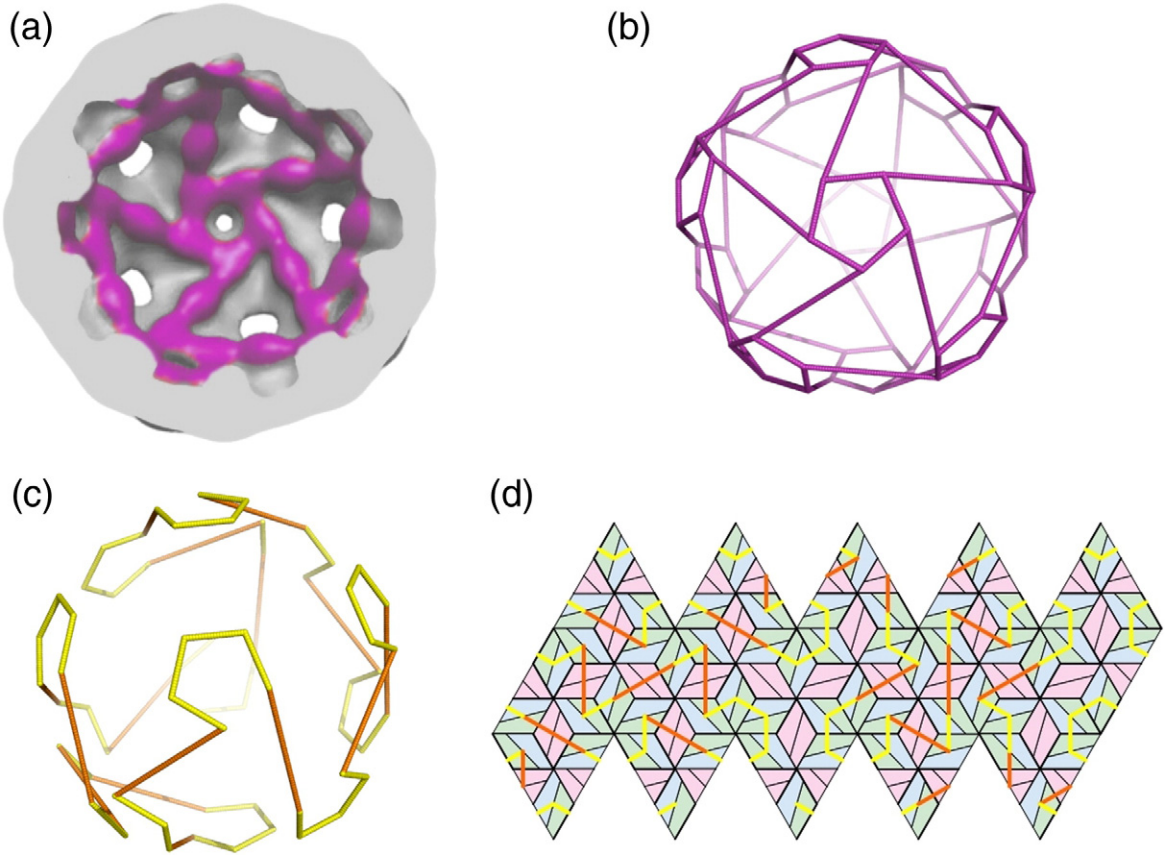


**Fig. 1.** The polyhedral cage of MS2 RNA density can be described as a Hamiltonian path. (a) A cryo-EM reconstruction of the outer RNA shell of bacteriophage MS2 (depicted in magenta) based on an image at ~17 Å resolution adapted from Van den Worm *et al.*[7] (b) A representation of this RNA shell as a polyhedral cage. (c) A three-dimensional view of a single Hamiltonian path, which meets every vertex of the polyhedron exactly once by moving along the short (yellow) and long (orange) edges of the polyhedral cage. (d) A planar net representation of the Hamiltonian path shown in (c) and its relation to the A (blue), B (green), and C (pink) quasi-equivalent subunits of the MS2 capsid.

has shown that the RNA-attributed density directly beneath the coat protein layer corresponds to stem–loops from within the viral RNA that make contact with the capsid proteins as they assemble.[9,10] MS2 capsid proteins form two types of dimer:[1,8] a symmetric and an asymmetric one, which occur 30 and 60 times, respectively, in the fully assembled capsid (see C/C- and A/B-like dimers in Ref. 1 and magenta and blue/green rhombs in the planar net in Fig. 1d, respectively). Stem–loops underneath the 60 A/B-like dimers, which are located on the vertices of the polyhedral cage in Fig. 1, convert symmetric C/C-like dimers into the required asymmetric form of dimer upon RNA–protein contact via a conformational switch.[11–14] This has led to the hypothesis of the dimer switching model for virus assembly.[1] The RNA genome provides positional information for incoming coat protein dimers through a series of conformational switching events, hence guiding the assembly of the capsid. This suggests that the linear genomic RNA molecule would ideally occupy all 60 vertices of the polyhedral cage in Fig. 1b, where each vertex corresponds to an A/B dimer contact, and extend a stem–loop to the protein layer as it does so.

## Hamiltonian paths as models of ssRNA genomic organization

We make two assumptions regarding the organization and encapsidation of the ssRNA in the infectious MS2 virus. The first assumption is that the RNA will make contact with each of the 60 A/B dimers in the capsid. The second assumption is that any excursion of the RNA from the outer to the inner shell returns to the same 5-fold axis in the outer shell. This is the most likely scenario, as it only requires that the RNA be able to fold into a string of simple structural elements, such as stem–loops, whereas the alternative would require a complicated network of long-distance interactions. This is consistent also with the observation that the inner shell, at least in maturation-protein-free capsids, becomes ordered only when the available space is sufficiently occupied by RNA. Indeed, this result has previously been attributed to RNA stem–loops being extended from the outer shell into the capsid interior, with the stem–loops being kept in place by electrostatic interactions when there is sufficient occupation (see also the discussions in Refs. 1 and 6). Based on these two assumptions, *all* possible ways in which the viral RNA can form the outer shell can be deduced from a mathematical principle called a Hamiltonian path.[15,16] The idea of a Hamiltonian path, which comes from graph theory, is simple: how can each vertex of a graph be visited once and only once by walking from one vertex to the next along the edges of a polyhedron? A Hamiltonian path naturally fits with how the MS2 viral genome interacts with its protein capsid to give the cage structure seen in cryo-EM. The genome must "visit" each of the 60 A/B coat protein dimers to effect conformational switching and then "walk" to the next A/B dimer along either one of the 30 long 2-fold edges or 60 short 5-fold edges with a segment of ssRNA. Thus, the 60 vertices of the graph of the polyhedron each represents a stem–loop interacting with an A/B coat protein dimer. Mathematics allows us to determine all possible Hamiltonian paths, and for the MS2 polyhedral cage, there are 40,678. To construct the set of all Hamiltonian paths for the MS2 polyhedron, we used a simple computer algorithm that constructs a trial path by making a series of moves on the polyhedral cage until no more moves are possible. If all vertices have been visited, then the trail path is stored as a Hamiltonian path. A three-dimensional representation of one such path is shown in Fig. 1c and in Fig. 1d as a planar net. Note that while the final organization of the RNA represented by this path is non-icosahedral, performing an icosahedral average of this (or any other) path will reproduce the polyhedral RNA cage observed experimentally (Fig. 1a and b). At first glance, the plethora of ways in which the viral RNA can realize the observed density suggests that the genome could be packaged differently in each virion. However, we show here that constraining the Hamiltonian paths to follow two simple rules, each of which can be derived from experiment, dramatically reduces the number of possible configurations and thus the complexity of virus assembly.

## The maturation protein constrains the RNA genome

The first rule incorporates biochemical information regarding the action of the viral maturation protein. Previous work by Shiba and Suzuki showed that two purine-rich sections of the viral RNA bind specifically to maturation protein.[17] These are located toward the 5′ (nucleotides 388–398) and 3′ (nucleotides 3510–3520) ends of the ~3.57-kb genome. Such bidentate binding circularizes the genomic RNA, with the high-affinity assembly initiation site stem–loop, known as TR, located approximately in the middle of the looped portion. Figure 2 illustrates the two possible ways in which this arrangement can be realized. In Fig. 2a, the 5′ and 3′ ends bind to maturation protein in a parallel configuration, while in Fig. 2b, they are antiparallel. Crucially, the formation of a closed loop of RNA (with tails of ~400 nt at the 5′ end and ~50 nt at the 3′ end) forces the genome to start and end at the same place as the maturation protein, that is, at the same 5-fold vertex. Only two types of Hamiltonian path are consistent with starting and ending in this way; one is known as a Hamiltonian cycle (Fig. 2a), and the other, a Hamiltonian pseudo-cycle (Fig. 2b).

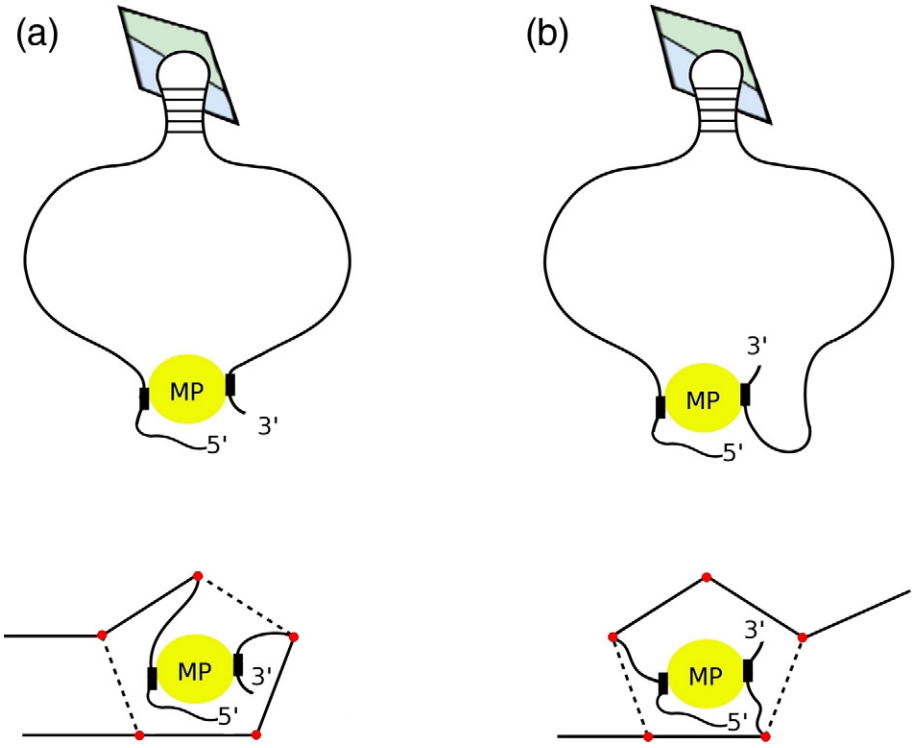


**Fig. 2.** Interaction with maturation protein constrains the RNA configuration. Maturation protein (yellow circle) binds to the MS2 genome near nucleotides 388 and 3520, resulting in tails of ~400 and ~50 nt at the 5′ and 3′ ends, respectively. The TR stem–loop, which is located approximately in the middle of the now looped RNA, is shown interacting with a coat protein A/B-like dimer, where capsid assembly would be expected to nucleate. Only two types of Hamiltonian paths are compatible with the presence of the maturation protein at a 5-fold vertex of the capsid and a closed loop of RNA. These are (a) the Hamiltonian cycle where the 5′ and 3′ ends of the genome bind in a parallel orientation and (b) the Hamiltonian pseudo-cycle where the 5′ and 3′ ends of the genome bind in an antiparallel orientation. See the lower part of the diagrams for how the 5′ and 3′ ends relate to the Hamiltonian path at the 5-fold vertex at which maturation protein is located.

Cycles have the edges of the Hamiltonian path corresponding to the 5′ and 3′ ends of the genome oriented in a parallel fashion (Fig. 2a), while pseudo-cycles have these ends antiparallel (Fig. 2b). There are only 37 cycles and 29 pseudo-cycles, that is, a total of 66 unique configurations that the genome can adopt in the outer shell when circularized by the action of the maturation protein. It is remarkable how severely the maturation protein restricts the configurations that the genome can adopt within the capsid, allowing only ~0.1% of the total possible configurations. Presumably, this brings about an accompanying reduction in the combinatorial complexity of assembly, which may contribute to the startling efficiency of virus assembly observed *in vivo*. However, the number of possible configurations can be reduced even further using an additional rule.

### Capsid assembly must be energetically favorable

The second rule simply requires that assembly of the capsid must occur in an energetically favorable way. Although the 66 paths mentioned above are all geometrically possible, not all of them are equally likely to occur due to energetic considerations. In addition to encoding an RNA configuration for the outer shell, each Hamiltonian path also provides a scenario for how the capsid is formed by sequential addition of coat protein dimers.[1] For each Hamiltonian path, the likely starting point on the path is located at an A/B dimer contact. Capsid assembly begins at this position by the binding of the TR stem–loop to a C/C-like dimer, promoting the formation of an A/B:TR complex[11] (Fig. 2). This nucleates capsid assembly, which then builds in both the 5′ and the 3′ directions of the circularized RNA. Following any Hamiltonian path from an A/B dimer chosen to be the site of capsid nucleation in the 5′ and 3′ directions gives one possible scenario for the way in which the capsid proteins and RNA could have co-assembled. More specifically, we assume that consecutive segments of the genomic sequence are recruited by the growing capsid and that assembly intermediates are stabilized by both protein–protein and protein–RNA interactions, forcing the genome to take on its packaged conformation. This model is consistent with experiments.[6,13,18]

Thus, the combined choice of the Hamiltonian path and the position in the path where nucleation begins has consequences for the types of capsid intermediates that are most likely to be formed at each stage of the assembly process.

From an earlier study of assembly intermediates in MS2 assembly,[13] we conclude that the formation of hexamers and decamers is preferred. Indeed, mass spectrometry of MS2 assembly in the presence of multiple copies of TR has revealed two dominant assembly pathways. Although the unit of capsid growth is a coat protein dimer, the only significantly populated species detected so far with stoichiometries above the dimer and below the capsid are the hexamer species (6 dimers placed symmetrically around a 3-fold axis) and the decamer species (10 dimers symmetrically around a 5-fold axis).[13,18] This makes sense from an energetic point of view because hexamers and decamers maximize the number of inter-subunit bonds with respect to other species containing the same numbers of dimers. Similarly, we assume for higher-order assembly intermediates that the assembly of the capsid (as described by the Hamiltonian path) continues to build via formation of additional complete hexamers or decamers. We therefore consider the set of Hamiltonian paths that implies a continuous formation of hexamers and decamers at the protein level as the preferred assembly scenarios. We rank these scenarios according to the number of inter-capsomer bonds formed, that is, according to the energy of bond formation. Since the RNA sequence is considered to have a relatively simple structure in terms of consecutive stem–loops, the RNA fold and its entropic contributions to capsid formation would be expected to be comparatively similar in each scenario, and we therefore concentrate on the energy of inter-capsomer bond formation as a means of distinguishing between these paths.

Figure 3 shows an example of the different possible Hamiltonian paths (or equivalently RNA organizations) that can occur for capsid intermediates which contain two, three, and four completed decamers. The number of inter-capsomer contacts in each assembly intermediate is indicated at the top right, while the arrows show how the intermediates with two or three completed decamers assemble into higher-order intermediates with three or four completed decamers. The leftmost column corresponds to an assembly scenario where the intermediates with two, three, and four completed decamers all have the maximum possible number of inter-capsomer contacts, while both the assembly scenarios in the two other columns have one intermediate that contains one less contact. We consider as energetically favorable those assembly scenarios that continuously maximize the number of inter-subunit contacts during assembly, and there is precisely one path of this type. We have tested and confirmed that it fulfills the relative distribution of densities observed in the experiment (see discussion around Fig. 5 for details on how the density distribution has been computed). In order to determine a suitable "energy cutoff," we then interrogated the next best paths according to the number of inter-capsomer contacts made during the formation of the capsid. There are two paths that deviate from the maximum number of inter-capsomer contacts only once during assembly. These two paths collectively also fulfill the density criterion, as well as when averaged together with the first. We then determined the next best paths, which deviate from the maximum number of contacts twice during assembly. There are five of these; however, they collectively no longer fulfill the density criterion and also do not fulfill it when averaged in combination with the three previously determined paths. We therefore chose our cutoff at this point and consider only the first three paths as possible solutions.

Figure 4 shows the three corresponding Hamiltonian paths together with a breakdown of the different components that define the structures of their associated protein containers. Interestingly, these three paths are distinguished in the following way: on the assumption that assembly starts in the middle of the Hamiltonian paths consistent with the location of the TR stem–loop, dimer addition along the 5′ and 3′ ends of the ssRNA molecule occurs such that the dimers added along each respective direction build up one of the two hemispheres of the protein container independently. For all other Hamiltonian paths, dimers added along the 5′ and 3′ ends both contribute to the formation of several hexamers and decamers in the capsid and therefore require dimer addition in the 5′ and 3′ directions to occur in a highly coordinated fashion to optimize hexamer and decamer completion. The lack of a need for coordination of dimer addition in the 5′ and 3′ directions in the case of our three solutions may therefore make the assembly process more robust.

## Experimental validation of the model

In order to validate this assembly model and its prediction for the asymmetric organization of the RNA in the outer shell, we compared the density for protein-bound RNA from the 5-fold-averaged cryo-EM reconstruction[2] with that predicted by the three Hamiltonian paths that satisfy our two simple rules. Our predictions were compared against a one-dimensional (1-D) projection of the cryo-EM data, which was masked to include only the density within the outer shell. The density in each two-dimensional layer perpendicular to the 5-fold axis was summed to give a single data point along the 1-D projection line. Each such data point is hence proportional to the amount of density present in

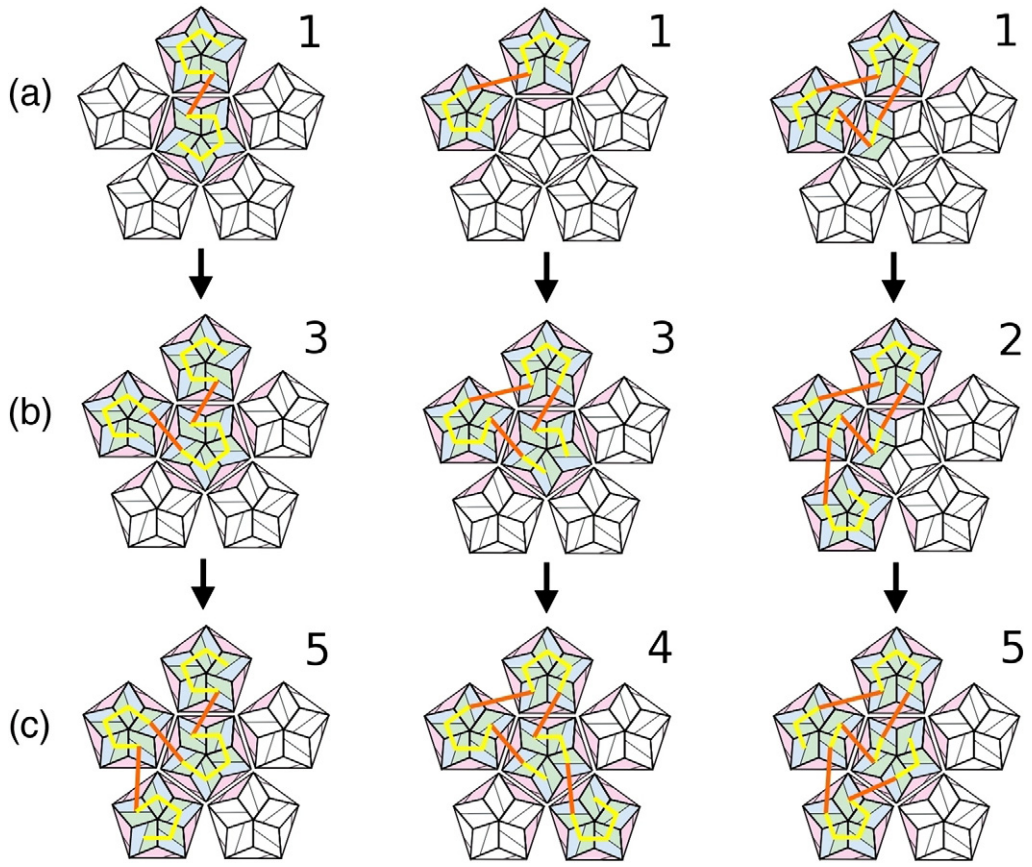


**Fig. 3.** The Hamiltonian paths with highly connected assembly intermediates. We demonstrate the formation of early intermediates with two, three, and four completed decamers. The number of inter-capsomer contacts for each intermediate is shown at the top right of each illustration. A/B and C/C dimers are color-coded as in Fig. 1d. (a) shows the intermediates with at least two completed decamers on each of the paths. Note that the two leftmost intermediates each exhibits two completed decamers but no completed hexamers (i.e., six rhomb shapes of alternating color surrounding the corner where three pentagonal patches meet), while the rightmost intermediate also has a completed hexamer. These continue assembling to intermediates with three or four completed decamers (and further hexamers) as illustrated in (b) and (c). Only the assembly scenario in the leftmost column maximizes the number of inter-capsomer contacts in (a), (b), and (c).

that two-dimensional layer. When data points for all layers in the capsid are combined, they produce a 1-D plot that shows how the relative amount of density (ascribed to RNA in the outer shell) changes when moving along the 5-fold axis. Three distinct (lower, middle, and upper) density peaks are observed in the experimental projection.

In order to obtain a theoretical prediction for the density in the outer shell, we employed the following technique. For a given Hamiltonian path, only certain edges of the polyhedron (shown in Fig. 1b) are occupied by RNA. In order to compute a 1-D plot for comparison with the experimental data, we required an estimate of how much density an occupied edge of the polyhedron would contribute to a cryo-EM reconstruction. Lacking a direct way of computing this quantity, we made a simple qualitative estimate based on the minimum number of nucleotides needed for the RNA to traverse the two different edges of the polyhedron. These distances were calculated from X-ray data (Protein Data Bank ID 1zdh[19]) and implied that an occupied long edge of the polyhedron contributes roughly 4.5 times as much density as an occupied short edge; that is, the different types of edges contribute to the overall density in a ratio of 9:2. Using this ratio, we computed the theoretical profiles for each Hamiltonian path by calculating the number of short and long edges that are occupied within the volume defined by the radial range corresponding to each of the three peaks seen in the experiment. In particular, the density within each such volume was calculated as 2 times the number of occupied short edges plus 9 times the number of occupied long edges contained in this volume. We display the result as three Gaussians with maxima corresponding to the three density values thus obtained, normalized such that the middle peak

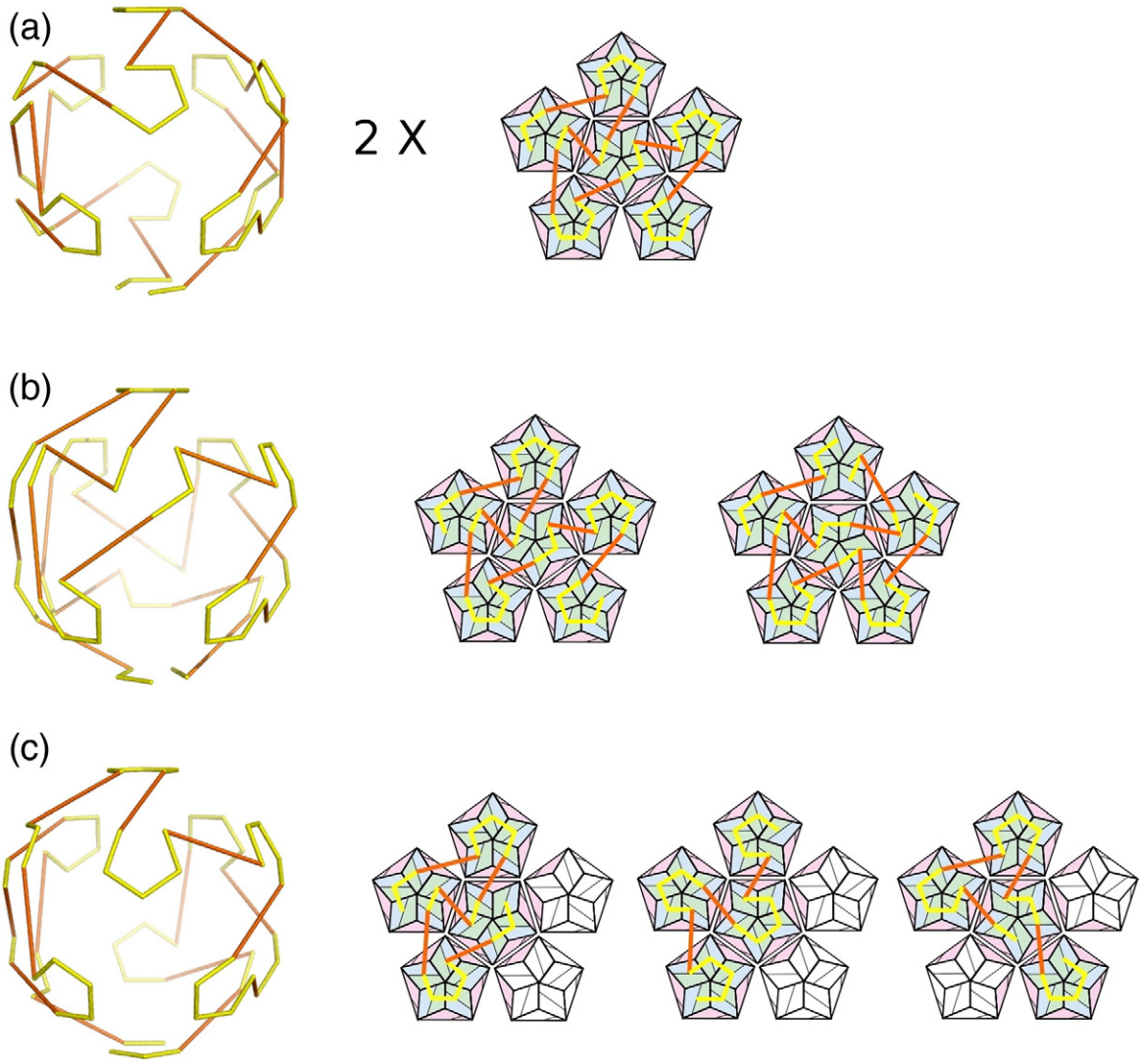


**Fig. 4.** The three predicted Hamiltonian paths for the MS2 system. The different components of each path are shown on the right. (a) A Hamiltonian cycle that builds each hemisphere of the capsid in an identical fashion. (b) A Hamiltonian pseudo-cycle that builds one hemisphere identical with the path in (a) but builds the second hemisphere slightly different. (c) A Hamiltonian pseudo-cycle that uses all three of the different motifs illustrated in Fig. 3c to construct the capsid.

coincides with the experimental value. Note that our measure is qualitative and, by construction, only has significance for the relative heights of the peaks.

Figure 5a shows the relative intensities of the peaks for different scenarios: when averaged over all 40,678 Hamiltonian paths, when averaged over only the 66 paths consistent with the presence of maturation protein, and when averaged over the 3 paths obtained after inclusion of the additional energetic argument. There is a striking agreement between the theoretical predictions and the experimental results. Note that the largest deviation is with the average of all 40,678 paths, that is the prediction in the absence of any rules, while the ensemble of three paths fulfilling both rules displays the best fit. The intensities for each of these three paths are shown individually in Fig. 5b. Figure 5c indicates the peak heights for the three individual paths in comparison with the experiment in a bar graph. The best fit is with the unique Hamiltonian path (path 3) that constructs the capsid by maximizing the number of inter-capsomer contacts as each individual capsomer is completed, that is, the one that is most energetically favorable according to our energy model. This suggests that the corresponding asymmetric RNA configuration occurs with the highest frequency in the ensemble of three favorable paths.

## Conclusion

Using two simple rules that are based on (1) the known interactions of MS2 genomic RNA with its

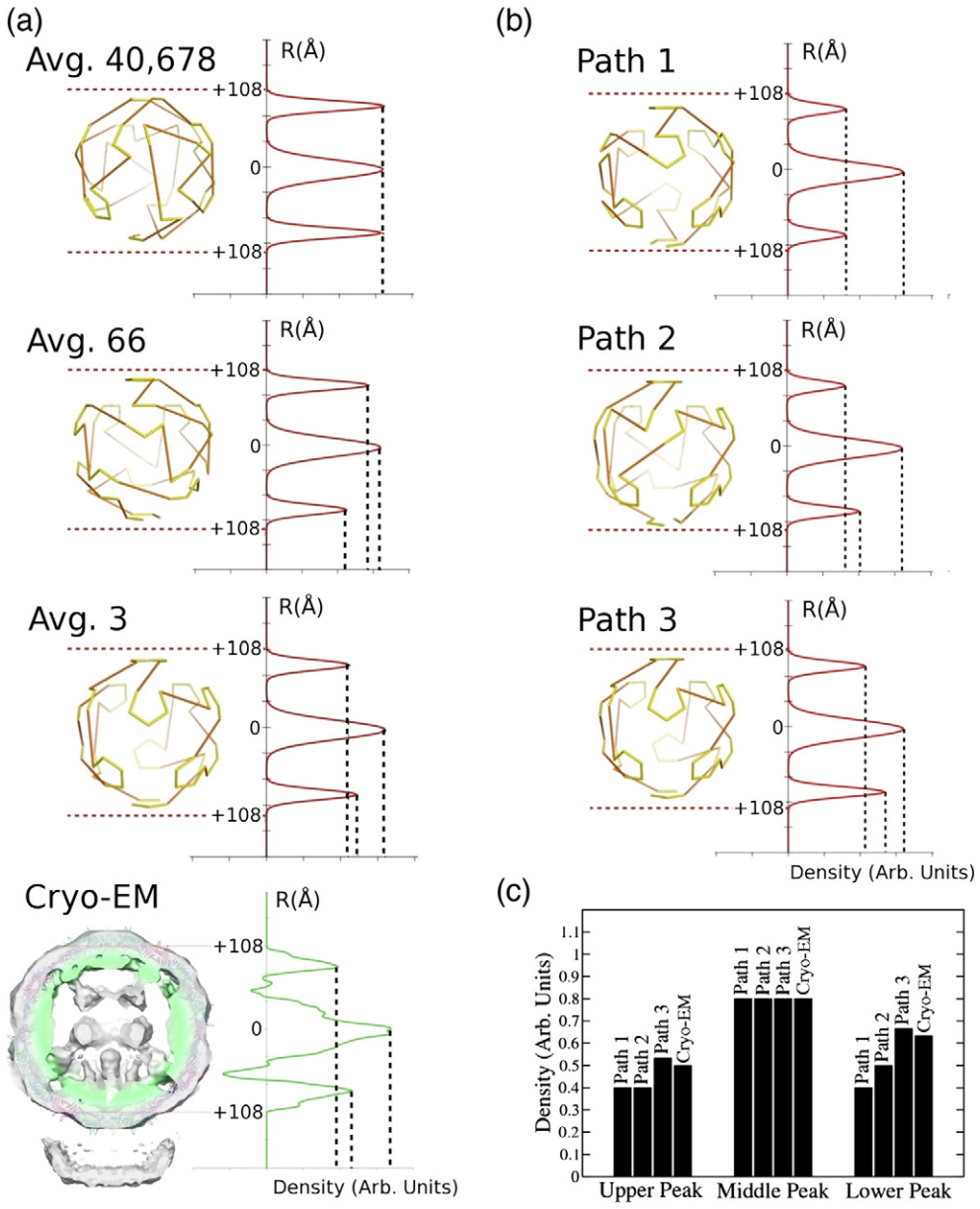


**Fig. 5.** Comparison of the prediction with the experiment. The figure shows a comparison of the RNA density in contact with the MS2 protein layer. The density profiles are shown as 1-D projections. The vertical axis of the projection gives the radial distance $R$ (away from the center of the capsid) in angstrom, while the horizontal axis indicates the amount of density at that radial distance in arbitrary units. All theoretical density profiles are normalized so that the middle peak has the same maximum value as the middle peak in the experimental data. Density profiles are shown in (a) for the following: average of all possible 40,678 Hamiltonian paths, average of all 66 paths consistent with RNA interaction with maturation protein, average of the 3 favorable paths, and the C5 cryo-EM reconstruction. (b) shows the density profiles for the three favorable paths individually. Path 1 corresponds to the Hamiltonian cycle in Fig. 4a, path 2 corresponds to the Hamiltonian pseudo-cycle in Fig. 4b, and path 3 corresponds to the Hamiltonian pseudo-cycle in Fig. 4c. (c) shows a bar graph containing a detailed comparison of the relative peak heights of the three favorable paths and the cryo-EM data. Path 3, which is most favorable according to our model, is the closest fit with the cryo-EM data.

maturation protein and (2) a requirement that the MS2 capsid be built in an energetically favorable manner, we have been able to predict features of the asymmetric distribution of RNA within a virus capsid. The implication must be that virus assembly *in vivo* is a consequence of the components following simple, local assembly rules. Our analysis relies upon a simple mathematical principle

(Hamiltonian paths) that can be used to completely describe the different ways in which a linear, asymmetric RNA molecule can repeatedly interact with an icosahedrally symmetric capsid. It shows that only three different configurations are likely for the linear ssRNA sequence as it resides in the infectious virus. This suggests that, by repeating a simple construction pattern during assembly, (i.e., repeated assembly of hexamers and decamers) ssRNA viruses ensure that they will form closed capsids efficiently in which genome encapsidation is all but guaranteed. These results allow us to understand the factors controlling the asymmetric distribution of the viral genome and are a major step toward understanding the mechanisms that underlie the formation of these particles. For example, they may serve in the creation of atomic models of a complete virion.

## Acknowledgements

R.T. acknowledges a Research Leadership award from the Leverhulme Trust that provides funding for E.C.D. and N.E.G. K.T. thanks the Leeds University Interdisciplinary Institute for Bionanosciences for PhD studentship support. N.A.R. and P.G.S. are grateful to The Leverhulme Trust for project grant support that funded part of this work.

## References

1. Toropova, K., Basnak, G., Twarock, R., Stockley, P. G. & Ranson, N. A. (2008). The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly. *J. Mol. Biol.* **375**, 824–836.
2. Toropova, K., Stockley, P. G. & Ranson, N. A. (2010). Visualising a viral RNA genome poised for release from its receptor complex. *J. Mol. Biol.* doi: 10.1016/j.jmb.2011.02.040.
3. Steitz, J. A. (1968). Identification of the A protein as a structural component of bacteriophage R17. *J. Mol. Biol.* **33**, 923–936.
4. Steitz, J. A. (1968). Isolation of the A protein from bacteriophage R17. *J. Mol. Biol.* **33**, 937–945.
5. Roberts, J. W. & Steitz, J. E. A. (1967). The reconstitution of infective bacteriophage R17. *Proc. Natl Acad. Sci.* **58**, 1416–1421.
6. Rolfsson, Ó., Toropova, K., Ranson, N. A. & Stockley, P. G. (2010). Mutually-induced conformational switching of RNA and coat protein underpins efficient assembly of a viral capsid. *J. Mol. Biol.* **401**, 309–322.
7. Van den Worm, S. H., Koning, R. I., Warmenhoven, H. J., Koerten, H. K. & van Duin, J. (2006). Cryo electron microscopy reconstructions of the *Leviviridae* unveil the densest icosahedral RNA packing possible. *J. Mol. Biol.* **363**, 858–865.
8. Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1990). The three-dimensional structure of the bacterial virus MS2. *Nature*, **345**, 36–41.
9. Lago, H., Parrott, A. M., Moss, T., Stonehouse, N. J. & Stockley, P. G. (2001). Probing the kinetics of formation of the bacteriohage MS2 translational operator complex: identification of a protein conformer unable to bind RNA. *J. Mol. Biol.* **305**, 1131–1144.
10. Carey, J. & Uhlenbeck, O. C. (1983). Kinetic and thermodynamic characterization of the R17 coat protein–ribonucleic acid interaction. *Biochemistry*, **22**, 2610–2615.
11. Stockley, P. G., Rolfsson, O., Thompson, G. S., Basnak, G., Francese, S., Stonehouse, N. J. *et al.* (2007). A simple, RNA-mediated allosteric switch controls the pathway to formation of a *T*=3 viral capsid. *J. Mol. Biol.* **369**, 541–552.
12. Basnak, G., Morton, V. L., Rolfsson, O., Stonehouse, N. J., Ashcroft, A. E. & Stockley, P. G. (2010). Viral genomic ssRNA directs the pathway towards a $T=3$ capsid. *J. Mol. Biol*, **395**, 924–936.
13. Morton, V. L., Dykeman, E. C., Stonehouse, N. J., Ashcroft, A. E., Twarock, R. & Stockley, P. G. (2010). The impact of viral RNA on assembly pathway selection. *J. Mol. Biol.* **401**, 298–308.
14. Dykeman, E. C., Stockley, P. G. & Twarock, R. (2010). Dynamic allostery controls coat protein conformer switching during MS2 phage assembly. *J. Mol. Biol.* **395**, 916–923.
15. Hamilton, W. R. (1858). An account of the Icosian calculus. *Proc. R. Ir. Acad.* **6**, 415–416.
16. Rudnick, J. & Bruinsma, R. (2005). Icosahedral packing of RNA viral genomes. *Phys. Rev. Lett.* **94**, 038101.
17. Shiba, T. & Suzuki, Y. (1981). Localization of A protein in the RNA–A protein complex of RNA phage MS2. *Biochim. Biophys. Acta*, **654**, 249–255.
18. Knapman, T. W., Morton, V. L., Stonehouse, N. J., Stockley, P. G. & Ashcroft, A. E. (2010). Determining the topology of virus assembly intermediates using ion mobility spectrometry – mass spectrometry. *Rapid Commun. Mass Spectrom.* **24**, 3033–3042.
19. Valegård, K, Murray, J. B, Stonehouse, N. J, van den Worm, S., Stockley, P. G. & Liljas, L. (1997). 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.