



How and why RNA genomes are (partially) ordered in viral capsids

Charles M Knobler¹ and William M Gelbart^{1,2,3}

There is a long and productive progression of X-ray crystallographic and electron microscopy studies establishing the structures of the spherical/icosahedral and cylindrical/helical capsids of a wide range of virus particles. This is because of the high degree of order – down to the Angstrom scale – in the secondary/tertiary/quaternary structure of the proteins making up the capsids. In stark contradistinction, very little is known about the structure of DNA or RNA genomes *inside these capsids*. This is because of the relatively large extent of disorder in the confined DNA or RNA, due to several fundamental reasons: topological defects in the DNA case, and secondary/tertiary structural disorder in the RNA case. In this article we discuss the range of partial order associated with the encapsidated genomes of single-stranded RNA viruses, focusing on the contrast between mono-partite and multi-partite viruses and on the effects of sequence-specific and non-specific interactions between RNA and capsid proteins.

Addresses

¹ Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA 90095, United States

² Molecular Biology Institute, UCLA, United States

³ California NanoSystems Institute, UCLA, United States

Corresponding author: Knobler, Charles M (gelbart@chem.ucla.edu)

Current Opinion in Virology 2022, 52:203–210

This review comes from a themed issue on **Virus structure and expression**

Edited by José R Castón and Adam Zlotnic

For complete overview about the section, refer [Virus structure & Expression \(2022\)](#)

Available online 24th December 2021

<https://doi.org/10.1016/j.coviro.2021.11.014>

1879-6257/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction: ssRNA versus dsDNA phenotypes

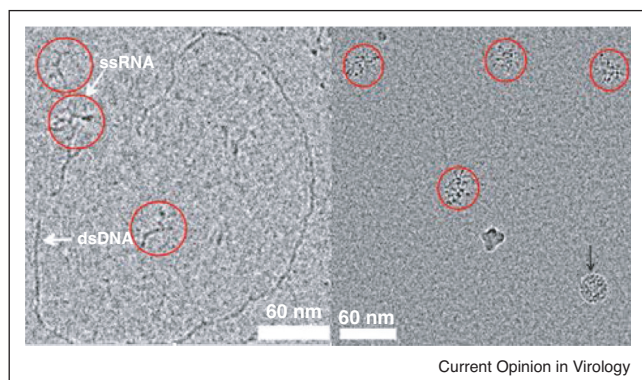
Viruses are unique among evolving organisms in that many of them have RNA rather than DNA as their genetic material. Further, the largest class of viruses are those with single-stranded (ss) RNA as their genomes while those with double-stranded (ds, duplex) DNA genomes constitute the second largest class. And it is the fundamentally different natures of single-stranded

versus double-stranded nucleic acids that determine many of the qualitative differences in the life cycles of these two predominant classes of viruses. In particular, their genome packaging and delivery steps involve contrasting physical phenomena as direct consequences of the fact that ssRNA behaves like a compact, *flexible*, effectively *branched* polymer [1,2,3^{*}] whereas dsDNA behaves like an extended, *stiff* and hard-to-compress, *linear* polymer [4–6]. These two biomolecule *phenotypes* differ greatly for the same *genotype* (gene coding), as is illustrated by the cryo- electron micrographs [7] shown in [Figure 1](#). Each of the red circles in the left image encloses a ssRNA molecule made by *in vitro* transcription of the dsDNA molecule that runs counter-clockwise from the middle of the left edge of the image on the right, around the bottom and up the right edge to the middle of the top, and then down a bit to the left. Significantly, the ssRNA contains exactly the same genetic information as the dsDNA, but is conspicuously more compact and hence spontaneously packageable into a virus particle in the presence of capsid protein – as happens in the cytoplasm of infected cells and *in vitro* under the right pH and salt conditions [8–10]. For the corresponding dsDNA form of the gene to be packaged, on the other hand, a large amount of work (compared to thermally available energy) has to be done in order to crowd and bend the DNA upon itself, accounting for why dsDNA viruses need to encode a motor protein strong enough to push their genome into a *pre-formed* capsid, often building up pressures (stored energy densities) as large 50 atm [11^{*},12^{*},13^{*},14–17].

The imperatives of disorder in packaged dsDNA and ssRNA

In the case of dsDNA, where (in contrast to ssRNA) there is no issue of largely unknown secondary and tertiary structure (see below), the impediment to ordering of the confined genome is the fundamental topological fact that a long chain cannot be accommodated at high density in a spherical volume without many regions of non-hexagonal packaging and of high-curvature [18]. More explicitly, optimally hexagonal (‘close’) packing of the DNA is required by the strong confinement, that is, the packing of the self-repelling genome at almost crystalline densities (volume fractions on the order of 0.5 [19]). If the genome consisted of many duplexes with a very special distribution of lengths, it could be organized in the capsid with essentially perfect hexagonal order. But it is instead a single duplex whose length is hundreds of times greater than the radius of the capsid. Accordingly, the confined DNA is unable to achieve the idealized cases of ‘spool’

Figure 1



ssRNA and dsDNA gene phenotypes.

Cryoelectron-microscopy images [7] of ssRNA and dsDNA, and of the relative sizes of ssRNA genes and the viral capsids that protect them. LEFT: Each circled molecule is a 2117nt-long ssRNA, while the dsDNA is 2117 base pairs long – with DNA sequence encoding the ssRNA, that is, the two molecules contain identical genetic information, but in ssRNA and dsDNA form, respectively. RIGHT: Here the red circles highlight copies of the 2774-nt-long ssRNA molecule that comprises the RNA helicase gene of the plant virus cowpea chlorotic mottle virus (CCMV), which is packaged by CCMV capsid protein; the black downward arrow in the lower right highlights one of these virus particles, a capsid protein shell containing one copy of the 2774-nt helicase gene.

chain arrangement [20,21] because topological defects are unavoidable, a result of the continuous nature of the chain and its need to avoid bending on a scale significantly smaller than the DNA persistence length (50 nm [4–6]). And there is necessarily a large number (ensemble) of these disordered configurations, each consistent with their packaging constraints and displaying a different set and distribution of topological defects, for example, non-parallel ‘crossings’/contacts of neighboring portions of the chain and/or radii of curvature much smaller than the persistence length.

In the case of ssRNA, on the other hand, each gene is *naturally* compacted [1,2,3*] by the large extent of self-complementary base-pairing – secondary structure formation – which ‘gathers in’ distant nucleotides along the chain. For short (<50–100 nt-long, e.g. transfer and enzymatic) RNAs, there are only a few low-energy secondary structures, often with a dominant ‘ground state’ that determines the activity of the RNA. But for viral RNAs, that is, sequences long enough (thousands of nts) to encode at least one gene, there are hundreds of thermally accessible secondary structures associated with each RNA molecule, requiring that it be described as a statistical object. The preponderant majority of these structures are those that render the RNA effectively branched and result in it having a significantly smaller radius of gyration – size – than it would have without the benefit of

intramolecular base-pairing [1,2,3*]. Additionally, viral gene sequences have evolved to exploit synonymous mutations to further minimize the 3D size of the molecule [22], and in particular to bring it down to the size dictated by the packing of the capsid protein. One expects then, to zeroth approximation, that the disorder associated with the large ensemble of native (‘unconfined/free-state’) secondary structures will carry over into the encapsidated state of the RNA, with weak ‘corrections’/ordering arising from interactions between the RNA and capsid protein.

Confinement of ssRNA genomes

As is clear from the right-hand image of Figure 1, when equilibrated in its ‘free state’ before addition of capsid protein, each encapsidated molecule that ends up being packaged by itself (e.g. the 2774nt-long 2nd gene of CCMV) is only slightly larger than the shell of protein that confines it. In this sense the RNA is only ‘weakly confined’ and we expect that its native secondary and tertiary structures are largely conserved upon encapsidation. Consistent with this, an asymmetric reconstruction of brome mosaic virus (BMV), a bromovirus closely related to CCMV, discussed below as an example of ‘Weak Ordering’ of packaged RNA, shows that only a small fraction of the virion-confined RNA is ordered. A larger degree of RNA ordering is found in the cases of bean pod mottle virus (BPMV), satellite tobacco mosaic virus (STMV) and Pariacoto virus (PaV), mentioned in the same section. Substantively more RNA ordering is found for the bacterial virus MS2 (discussed in the ‘Substantial Ordering’ section), because of the ‘extra’ constraints introduced by specific RNA–protein interactions and by co-localization of the RNA ends at a symmetry-breaking protein in an otherwise icosahedral capsid, resulting in a majority fraction of the RNA being ordered.

In this article we discuss the necessarily limited extent to which ssRNA genomes are ordered inside their protective capsids, how this is related to the nature of the spontaneous co-self-assembly of RNA and protein, and how it can be determined by high-resolution structural studies using X-ray crystallography and electron microscopy – much as has been done for individual protein subunits. Indeed, structural virology has a history [23] that is as old as that of protein crystallography: as early as the mid/late 1930s X-ray diffraction was reported from concentrated solutions of tobacco mosaic virus (TMV), a cylindrical-capsid positive-sense ssRNA virus [24]. This was a time when it was not even clear that the TMV virus particles contained both protein and RNA, and twenty years before RNA was understood to be the ‘messenger’ between DNA and protein sequence. But by the early 1940s these X-ray studies were able to provide important information about the sizes and shape of the TMV virions (18nm-by-300 nm cylinders), soon corroborated by low-resolution electron

microscopy imaging [25]. And in the mid-1950s the helical nature of these particles [26] and the icosahedral symmetry of spherical viruses [27] were established, followed by the first atomic-resolution X-ray crystallographic studies of icosahedral RNA viruses in the late 1970s and early 1980s [28,29]. While this work showed the detailed organization of the capsid proteins in the shell surrounding the genome, no strong electron density – let alone structure – could be ascribed to the RNA. The presence of a shell of RNA inside the capsid was identified by small-angle neutron scattering [30] but it was only a decade later that a viral structure was determined in which order associated with the RNA could be identified [31]. In the ensuing forty years a thousand icosahedral virus structures have been placed in the Protein Data Bank [32] (with the number increasing rapidly upon the advent of cryoelectron reconstructions), and several protein structural ‘folds’ – like the ‘jelly-roll’ [33] – have been shown to be common to groups of viruses. But despite these significant advances in our knowledge of viral capsid structure, allowing us to deduce the position of essentially every amino acid residue, we still know little about the structure of the RNA within them – as has been put succinctly, ‘[encapsidated] viral RNAs are the dark matter of structural virology’ [34].

The case of perfect genome ordering

As an outlier we have the singular case of TMV – and a small handful of other cylindrical viruses, for example, Ebola [35] – in which the ssRNA is *perfectly* ordered in its virion. This comes about because of the unique way in which the RNA is packaged into its capsid *protein*. More explicitly, instead of being a hollow (icosahedrally symmetric) *sphere*, the TMV capsid shell is a hollow (helically symmetric) *cylinder*, with a thickness of 7 nm and inner and outer radii of 2 and 9 nm. But the overwhelmingly important difference between this virion and virtually all others is that the RNA is not contained in the hollow interior of the shell but is instead embedded *in the protein itself*. Remarkably, as reported in classic back-to-back papers [36] by Rosalind Franklin and Donald Caspar in which they compared the radial density profiles of TMV cylinders with and without RNA, the RNA is buried 2–4 nm into the shell. Further, its string of nucleotides has the full helical symmetry of the capsid, and is in this way stripped of all its native secondary/tertiary structure and is perfectly ordered.

An inescapable consequence of perfect order for the RNA in its TMV-packaged state is the huge increase in its free energy from its ‘free state’, equilibrated in physiological solution. The packaged RNA – with its single helical, hydrogen-bond-free, configuration – has none of the benefit of the self-complementary base-pairings or tertiary-structure interactions it enjoys in its free state, nor of the free-state entropy arising from the many configurations associated with each of the large number of

thermally accessible secondary structures. Of course, what offsets these free energy gains is the strong stabilization of the ssRNA helix by the capsid proteins in which it is embedded, three nucleotides per subunit. Clearly these RNA-capsid interactions are huge compared to those present in the usual case of RNA confined in the *hollow interior* of the majority of spherical viruses and of cylindrical viruses as well.

Partial genome ordering

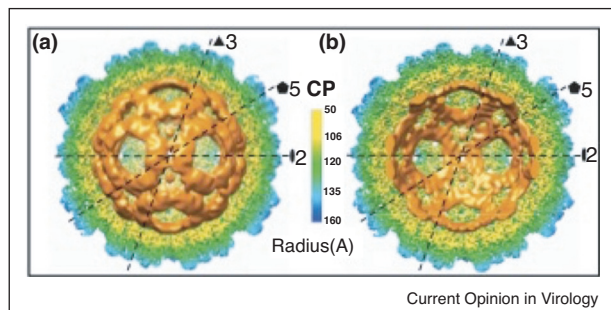
In principle the 3D structure of *partially* ordered RNA in a viral capsid, that is, the structures present in sufficiently large subsets of the RNA configurations, can be determined by asymmetric reconstruction of cryoelectron microscopy images. Among the few such studies that have been carried out, two reconstructions, that of the plant virus BMV and that of the bacteriophage MS2, provide interesting contrasts even as their capsids are closely identical in diameter, have $T=3$ triangulation numbers, and package RNAs ≈ 3000 nt in length.

Weak RNA ordering: BMV

Plant viruses are unique in that a significant fraction (30–40%) of them are – like BMV – *multipartite*, that is, their genes are not all contained in the same virion [37]. In particular, the BMV genome consists of three single-stranded positive-sense RNAs: RNA1 (≈ 3200 nt) and RNA2 (≈ 2800 nt) which code for the replicase and helicase proteins, and RNA3 (≈ 2100 nt) which codes for the movement and capsid proteins. These gene molecules are packaged into identical capsids, one containing RNA1, one containing RNA2 and a third containing RNA3 and a subgenomic RNA4 (≈ 700 nt) that provides redundant coding for the CP. Transfer of the BMV genetic information to a cell therefore requires infection by no less than three types of capsids, each containing ≈ 3000 nts.

The capsid shells, composed of 180 copies of a ≈ 20 kDa protein, are indistinguishable. This fact, and the near equivalence of the mass of the RNA they contain, make it nearly impossible to isolate one virion type from another by physical means such as sedimentation. Separation by electrophoresis is also precluded because the electrophoretic mobility of BMV is independent of the charge of the encapsidated cargo [38]. As a result, any measurements reported for purified BMV represent an average of those properties over the three virus types, which substantively limits the resolution of structural studies of the RNA. This limitation has been overcome by Chakravarthy *et al.* [39**] who produced BMV virions with unique RNA content by agrobacterial infection in plants. Using this strategy, BMV virions containing only RNA3 + RNA4 were synthesized and their structure determined by cryoEM reconstructions [40**].

Figure 2

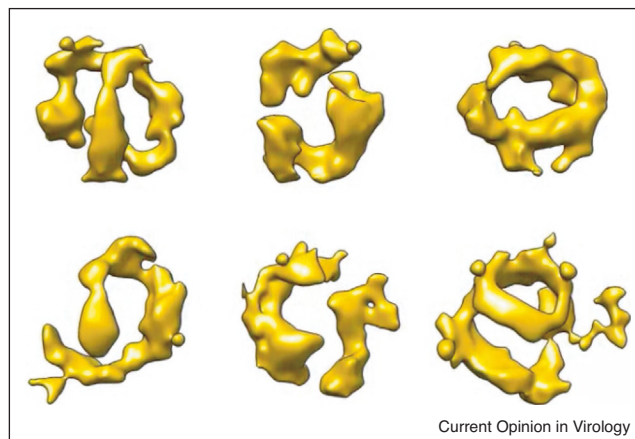


Average icosahedral order of encapsidated BMV RNA. Symmetric reconstruction of BMV illustrating how the RNA (orange) appears as rings of density around each of the capsid pentamers. **(a)**: Entire virus. **(b)**: Back half, showing the capsid interior [40*].

Figure 2 shows BMV reconstructions in which icosahedral symmetry has been imposed. This symmetry imposition necessitates that any RNA density resolved adhere to icosahedral symmetry, and as such, it cannot resolve any asymmetric RNA density. The RNA density (orange) is seen only at the interior surface of the capsid and, notably, the only RNA density resolved is found to be interacting with the positively charged N-termini of the CP, a result of the electrostatic attraction between the protein N-termini and the negatively charged phosphate backbone of the RNA. More specifically, the RNA interacts with the N-termini of the hexameric CP subunits, which extend outward from the threefold symmetry axes along the twofold symmetry axes at the 'edges' of the pentamers (the fivefold symmetry axes), resulting in the formation of RNA 'rings around the pentamers'. The fact that the interior of the particle appears empty indicates that much of the RNA is disordered and hence does not survive icosahedral averaging; it is estimated that less than 50% of the total RNA density is resolved in the icosahedrally symmetric (orange) structure shown here. We note further that no features like double-helix grooves are discernible, and certainly fitting to nucleotide models is not possible. The fact that the resolved RNA density lacks a particular structure implies that the CP and RNA are interacting through non-specific interactions that do not depend on the presence of a unique or predominant RNA configuration, and that this density represents instead an ensemble of RNA secondary/tertiary structures.

Surprisingly, attempts at carrying out an asymmetric reconstruction – both with and without capsid subtraction – produced little new information about the structure; in other words, the symmetric and asymmetric reconstructions of BMV are nearly identical, suggesting that the RNA density is not sufficiently ordered, and

Figure 3



Absence of a dominant RNA configuration in BMV virion. Asymmetric reconstruction of the internal density associated with the N-termini CP and RNA at the twofold axis. Shown are six of twenty similar but distinct representative three-dimensional RNA conformations, indicative of the lack of a specific organization [40*].

therefore, cannot be reconstructed. As such, the internal RNA genome was examined using 'subparticle reconstruction' which focuses on the CP N termini and the RNA at each of the three symmetry axes. In each instance there is no unique configuration, as shown for example in Figure 3. The structures generated through the subparticle analysis at the twofold axis have been divided into 20 classes, and an asymmetric reconstruction has been carried out separately for each of them, thereby allowing for only the signal in this subparticle region to contribute to the final reconstruction. The RNA density for six representative classes is shown, from which it is clear that there is no unique or dominant configuration. Further, the generated classes were almost equally populated, suggesting strong diversity in structure across the reconstructions. A similar result was found for subparticle reconstructions at the threefold and fivefold axes. These cryoEM results make clear that in BMV there is no unique RNA configuration, but rather an ensemble of conformations associated with the packaged genome.

This lack of preference for a particular configuration of the encapsidated RNA3 + RNA4 is consistent with the fact that RNA1 and RNA2 – involving very different sequences and ensembles of secondary and tertiary structures from RNA3 + RNA4 – need also to be accommodated in/package by identical capsid shells. Accordingly, there is evolutionary pressure on the virus for its capsid protein to bind and package RNA using non-specific mechanisms, for example, the nucleotide-sequence-independent electrostatic interactions associated with

phosphate charge on the RNA backbone and the cationic residues in the CP N-termini that end up lining the inside of the capsid.

Bean Pod Mottle Virus is the first virus in which structured RNA was identified. It has a bi-partite genome consisting of 3500 nt and 5800 nt ssRNA positive-sense RNAs packaged into separate but identical $T = 3$ capsids identified on the basis of sedimentation as, respectively, the M (medium) and B (bottom) components; 20 % of the capsids formed *in vivo* are empty and constitute the light (L) component. Chen, et al. [31], using x-ray crystallography, solved the structure of the M component and found that about 20% of the RNA has a single-stranded helical structure with base stacking in the form of 33 nt arranged in trefoil rings around the 20 three-fold axes. From analysis of the nucleotide sequence of this component it was concluded [41] that there were “no reiterated structures which might obviously account for the appearance of ordered RNA” in its three-dimensional structure; Lin, et al. [42], however, noted that there are numerous pentameric sequences of the form APuPyPyPy dispersed through the genome, but without any regular order, which may interact with the protein and function as packaging signals. Interestingly, a study [42] of the B component showed an RNA structure that, except for subtle differences, is virtually identical to that of the M component. In the case of *single-molecule-genome* ssRNA viruses like STMV [43] and two-molecule-genome-but-*single-capsid* Pariacoto virus (PaV) [44], there is more organization of the RNA than in the multipartite bromoviruses, specifically involving the ordering of duplexes near the inner capsid surface. (The ‘satellite’ virus STMV has a 1058-nt-long RNA genome packaged in a $T = 1$ 60-subunit protein shell, and Pariacoto has a pair of 3011-nt and 1311-nt RNA genes packaged together in a $T = 3$ 180-subunit shell.) But the ordered RNA portions still encompass only a fraction of the RNA, with disordered connections between the duplexes that are not resolvable in reconstructions. There are, unfortunately, very few studies in which the secondary structures before and after packaging have been probed. One example [45^{*}] is that of STMV, where differences in the secondary structures of ‘free’ and ‘in virio’ forms of the genome show that the packaged form involves more branching of the RNA and larger, more-flexible, single-stranded loops between duplexes, consistent with the susceptibility of these helices to being icosahedrally ordered through interaction with the capsid edges. Indeed, X-ray crystallographic studies show that a significant fraction of the packaged RNA is organized into icosahedrally ordered double-helical segments, though it is important to note that the imposition of icosahedral symmetry obfuscates any non-icosahedrally ordered density [43]. This reorganization of the RNA upon packaging is also consistent with cryo-EM imaging showing that STMV is unusually extended when free in solution [46]. Similarly, icosahedrally averaged X-ray crystallographic

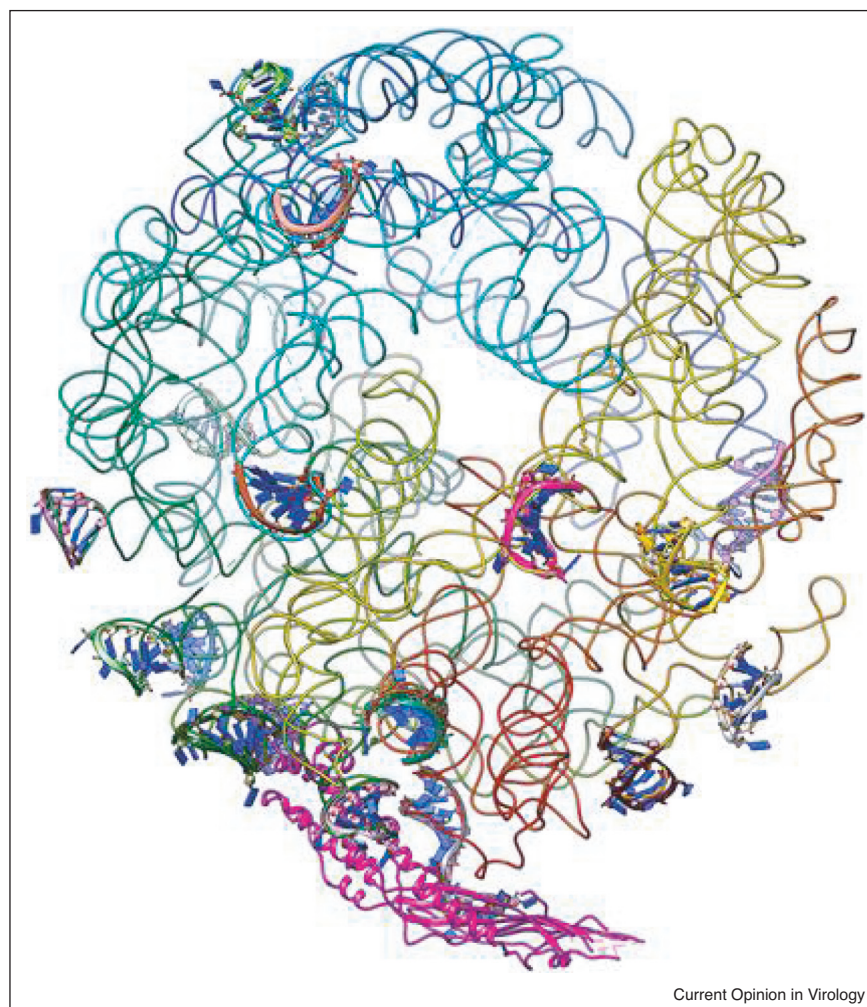
and cryoEM reconstructions of Pariacoto virus show that about a third of the RNA is organized into a dodecahedral cage of co-linear duplexes at the inner surface of the capsid – again, this cage structure may be the result of imposing symmetry on the RNA density resolved [44]; in this case, as with BMV, the cationic N-termini of CPs penetrate the capsid interior and interact nonspecifically with the anionic RNA. Clearly it is important to elucidate further the changes in the ensemble of secondary structures undergone as these ssRNA viral genomes are packaged through interactions with their capsid proteins and become partially ordered.

Substantial RNA ordering: MS2

In contrast, asymmetric reconstructions of MS2 can be made without recourse to subtraction of the icosahedral capsid contribution. In particular, two recent studies, one at a resolution of 8.7 Å [47^{*}] and the other at 4.4 Å [48^{**}] reveal a high degree of ordering of the ssRNA genome in this virion. More explicitly, unlike for BMV, when the data sets in the higher-resolution study [48^{**}] were divided into 10 classes and subjected to refinement, the resulting RNA density maps were almost identical, consistent with a single dominant structure. A significant portion of the genome could be resolved to sufficient resolution that the electron density shows prominent major and minor grooves, indicating the presence of dsRNA. In addition, as many as fifty stem loops can be identified, most of which contact the capsid at their tip, and sixteen of these stem loops were sufficiently resolved to permit individual nucleotides to be identified, which allowed the backbone of 80% of the genome to be traced and a model of its structure to be constructed: see Figure 4. Further, several long-range base-pairs and kissing-loop interactions were determined.

Figure 4 points up the asymmetry of the MS2 RNA structure in a very particular way, as seen in the conspicuously higher density (and number of highly ordered stem loops) proximal to the maturation protein, and the extended distribution of resolved RNA stem loops across the interior of the particle. The ordering of the structure arises from two main sources that go beyond the simple *confinement* of the genome. First, there are specific interactions – ‘packaging signals [PSs]’ – associated with the stem loops that interact strongly with the CP (in particular, three conserved motifs of CP-RNA interactions were determined from the cryoEM structure). These structures and their role in genome packaging have been extensively investigated for MS2 and argued to both drive the self-assembly process and determine the unique RNA configuration arising from a special distribution of the PSs throughout the genome [49,50^{*}]. The second extra layer of constraint on the genome organization in MS2 arises from the presence of a maturation protein that replaces two of the CPs at one of the icosahedral twofold axes, breaking the capsid symmetry,

Figure 4



Reconstruction of the dominant RNA configuration in MS2 virion. Backbone structure of the MS2 RNA and its maturation protein (magenta) [48**]. The blue-to-red coloring indicates 5' to 3' direction. The 16 high-resolution stem loops are indicated by ribbons.

and extending out of the capsid for binding host cells and initiating infection. Although in the asymmetric reconstruction the protein appears to bind only to the 3' end of the RNA (see Figure 4), it has been demonstrated that there is also binding at the 5' end, effectively circularizing the RNA [51] (with a similar situation for the related phage Q β [52]). These results make clear how and why, in contrast to BMV where the RNA is uniformly distributed at the capsid surface and is described by an ensemble of secondary/tertiary structures, a single dominant configuration is present within the MS2 virion. In other words, a majority fraction of the MS2 ssRNA genome transitions from an ensemble of secondary/tertiary structures in solution to essentially a single configuration after encapsidation.

Conclusions

The ordering of the RNA genomes in BMV and MS2 represent two extreme scenarios – an ensemble of structures in BMV, and a predominant ordered structure in MS2. It appears that the order in MS2 is due primarily to the constraining influence of the maturation protein and its effective circularizing of the RNA and ‘tying up’ of its ends at a particular point on the inside surface of the capsid. In this connection it would be interesting to compare the genome order in MS2 with that of an *in vitro* reconstituted capsid that does not include the maturation protein.

It is likely that partial order associated with ensembles of RNA structures is the more common scenario for single-stranded genomes. The next several years are likely to

feature many new high-resolution efforts to identify and characterize the nature and extent of genome order in RNA viruses, with the challenge of relating this order (and lack thereof) to the genome packaging and delivery. These studies will further elucidate, and be informed by, the role of ‘packaging signals’ – (sequence-) specific RNA structures that bind capsid protein with high affinity and direct the assembly of and stabilization of capsids.

Conflict of interest statement

Nothing declared.

Acknowledgements

We are pleased to acknowledge the helpful comments of Dr Christian Beren in the preparation of this article. Our work on genome packaging in RNA viruses and virus-like particles has been supported by the Genetic Mechanisms Program of the Molecular and Cellular Biosciences (MCB) Division of the National Science Foundation, most recently by grants MCB 1716975 and 2103700.

Acknowledgements

We are pleased to acknowledge the helpful comments of Dr Christian Beren in the preparation of this article. Our work on genome packaging in RNA viruses and virus-like particles has been supported by the Genetic Mechanisms Program of the Molecular and Cellular Biosciences (MCB) Division of the National Science Foundation, most recently by grants MCB 1716975 and 2103700.

References

- Yoffe A, Prinsen AM, Gopal A, Knobler CM, Gelbart WM, Ben-Shaul A: **Predicting the sizes of large RNA molecules.** *Proc Natl Acad Sci U S A* 2008, **105**:16153-16158.
- Fang LT, Gelbart WM, Ben-Shaul A: **The size of RNA as an ideal branched polymer.** *J Chem Phys* 2011, **135**:155105-15511-5.
- Gopal A, Zhou ZH, Knobler CM, Gelbart WM: **Visualizing large RNA molecules in solution.** *RNA* 2012, **18**:284-299
This work presents cryo-EM images of viral-genome-length single-stranded RNA molecules in their native state in aqueous solution, and analyzes their sizes and shapes in terms of the branching patterns associated with their ensembles of secondary structures.
- Bednar J, Furrer P, Katrich V, Stasiak A, Dubochet J, Stasiak A: **Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA.** *J Mol Biol* 1995, **254**:579-594.
- Mantelli S, Muller P, Harlepp S, Maaloum M: **Conformational analysis and estimation of the persistence length of DNA using atomic force microscopy in solution.** *Soft Matter* 2011, **7**:3412-3416.
- Guilbaud S, Salomé L, Destainville N, Manghi M, Tardin C: **Dependence of DNA persistence length on ionic strength and ion type.** *Phys Rev Lett* 2019, **122**:028102/1-6.
- Gopal A, Knobler CM, Gelbart WM: unpublished.
- Bancroft JD: **The self-assembly of spherical plant viruses.** *Adv Virus Res* 1970, **16**:99-134.
- Zlotnick A, Porterfield JZ, Wang JC-Y: **To build a virus on a nucleic acid substrate.** *Biophys J* 2013, **104**:1595-1604.
- Garmann RF, Comas-Garcia M, Knobler CM, Gelbart WM: **Physical principles in the self-assembly of a simple spherical virus.** *Acc Chem Res* 2015, **49**:48-55.
- Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C: **The bacteriophage phi29 portal motor can package DNA against a large internal force.** *Nature* 2001, **413**:748-752
The authors report the first single-molecule measurements of the forces exerted by viral-encoded motor proteins in packaging double-stranded DNA into pre-formed capsids, for the particular case of bacteriophage phi29. The forces are shown to increase from zero to tens of piconewtons as the confinement of the genome proceeds to completion.
- Kindt J, Tzili S, Ben-Shaul A, Gelbart WM: **DNA packaging and ejection forces in bacteriophage.** *Proc Natl Acad Sci U S A* 2001, **98**:13671-13674
In parallel with the experimental results reported in Ref. [11], the authors here report the first theoretical prediction of the forces and pressures built up in a viral capsid as dsDNA is packaged into it. The pressures become as large as tens of atmospheres upon completion, due to a combination of strong bending and self-repulsion of the confined DNA.
- Evilevitch A, Lavelle L, Knobler CM, Raspaud E, Gelbart WM: **Osmotic pressure inhibition of DNA ejection from phage.** *Proc Natl Acad Sci U S A* 2003, **100**:9292-9295
The first measurement of the pressure in a dsDNA virus, for the particular case of bacteriophage lambda. The pressure is found to be as large as tens of atmospheres, and to depend on the concentrations of mono- and multi-valent cations that moderate the self-repulsion of the confined DNA.
- Tzili S, Kindt J, Gelbart WM, Ben-Shaul A: **Forces and pressures in DNA packaging and release from viral capsids.** *Biophys J* 2003, **84**:1616-1627.
- Purohit PK, Kondev J, Phillips R: **Mechanics of DNA packaging in viruses.** *Proc Natl Acad Sci U S A* 2003, **100**:3173-3178.
- Gelbart WM, Knobler CM: **Pressurized viruses.** *Science* 2009, **323**:1682-1683.
- Bauer DW, Huffman JB, Homa FL, Evilevitch A: **Herpes virus genome, the pressure is on.** *J Am Chem Soc* 2013, **135**:11216-11221.
- Klug WS, Feldmann MT, Ortiz M: **Three-dimensional director-field predictions of viral DNA packing arrangements.** *Comput Mech* 2005, **35**:146-152.
- Purohit P, Inamdar MM, Grayson PD, Squires TM, Kondev J, Phillips R: **Forces during bacteriophage DNA packaging and ejection.** *Biophys J* 2005, **88**:851-866.
- Cerritelli M, Cheng N, Rosenberg AH, McPherson CE, Booy FP, Steven AC: **Encapsidated conformation of bacteriophage T7 DNA.** *Cell* 1997, **91**:271-280.
- Lander GC, Tang L, Casjens SR, Gilcrease EB, Prevelige P, Poliakov A, Potter CS, Carragher B, Johnson JE: **The structure of an infectious P22 virion shows the signal for headful DNA packaging.** *Science* 2006, **312**:1791-1795.
- Tubiana L, Božič AL, Micheletti C, Podgornik R: **Synonymous mutations reduce genome compactness in icosahedral ssRNA viruses.** *Biophys J* 2015, **108**:194-202.
- Rossmann MG: **Structure of viruses: a short history.** *Q Rev Biophys* 2013, **46**:133-180.
- Bernal JD, Fankuchen I: **Structure types of protein crystals from virus-infected plants.** *Nature* 1937, **139**:923-924.
- Kausche GA, Pfankuch E, Ruska H: **Die Sichtbarmachung von pflanzlichem Virus in Übermikroskop.** *Naturwiss* 1939, **27**:292-299.
- Franklin RE, Holmes KC: **Tobacco mosaic virus: application of the method of isomorphous replacement to the determination of the helical parameters and radial density distribution.** *Acta Cryst* 1958, **11**:213-220.
- Caspar DLD: **Structure of bushy stunt virus.** *Nature* 1956, **177**:475-476.
- Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G: **Tomato bushy stunt virus at 2.9 Å resolution.** *Nature* 1978, **276**:368-373.
- Abad-Zapatero C, Abdel-Meguid S, Johnson JE, Leslie A, Rayment I, Rossmann MG, Suck D, Tsukihara T: **Structure of southern bean mosaic virus at 2.8 Å resolution.** *Nature* 1980, **286**:33-39.

30. Chauvin C, Witz J, Jacrot B: **Structure of the tomato bushy stunt virus: a model for protein-RNA interactions.** *J Mol Biol* 1978, **124**:641-665.
 31. Chen Z, Stauffacher C, Li Y, Schmidt T, Bomu W, Kamer G, Shanks M, Lomonosoff G, Johnson JE: **Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution.** *Science* 1989, **245**:154-159.
 32. Johnson JE, Olson AJ: **Icosahedral virus structures and the protein data bank.** *J Biol Chem Rev* 2021, **296**:1000554.
 33. Hogle JM, Chow M, Filman DJ: **Three dimensional structure of poliovirus at 2.9 Å resolution.** *Science* 1985, **229**:1358-1365.
 34. Schroeder SJ, Stone JW, Bleckley S, Gibbons T, Mathews DM: **Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints.** *Biophys J* 2011, **101**:167-175.
 35. Sugita Y, Matsunami H, Kawaoka Y, Noda T, Wolf M: **Cryo-EM structure of the Ebola virus nucleoprotein-RNA complex at 3.6 Å resolution.** *Nature* 2018, **563**:137-140.
 36. (a). Caspar DLD: **Structure of tobacco mosaic virus: radial density distribution in the tobacco mosaic virus particle.** *Nature* 1956, **177**:928;
(b). Franklin RE: **Structure of tobacco mosaic virus: location of the ribonucleic acid in the tobacco mosaic virus particle.** *Nature* 1956, **177**:928-930.
 37. Sicard A, Michalakos Y, Gutiérrez S, Blanc S: **The strange lifestyle of multipartite viruses.** *PLoS Pathog* 2016, **12**:e1005819.
 38. Duran-Meza AL, Villagrana-Escareño MV, Ruiz-García J, Knobler CM, Gelbart WM: **Controlling the surface charge of simple viruses.** *PLoS One* 2021, **16**:e0255820.
 39. Chakravarthy A, Reddy VS, Rao ALN: **Unravelling the stability and capsid dynamics of the three virions of brome mosaic virus assembled autonomously in vivo.** *J Virol* 2020, **94**:e01794-19
- This work reports the first *in vivo* (in planta) preparation of pure virions of the multipartite plant virus BMV, thereby permitting rigorous investigations of the structure and properties of virus particles all of which contain the same RNA.
40. Beren C, Cui Y, Chakravarty A, Yang X, Rao ALN, Knobler CM, Hong Zhou Z, Gelbart WM: **Genome organization and interaction with capsid protein in a multipartite RNA virus.** *Proc Natl Acad Sci U S A* 2020, **117**:10673-10680
- The authors present the first asymmetric reconstruction of pure virions of a multipartite virus – in particular, BMV particles containing RNA3 and RNA4 molecules – and show that, even though there is some icosahedrally symmetric RNA density just inside the positively charged capsid interior, the majority of the RNA is disordered and associated with an ensemble of secondary structures and overall configurations.
41. Lin T, Cavarelli J, Johnson JE: **Evidence for assembly-dependent folding of protein and RNA in an icosahedral virus.** *Virol* 2003, **31**:26-33.
 42. Macfarlane SA, Shanks M, Davies JW, Zlotnick A, Lomonosoff GP: **Analysis of the Nucleotide Sequence of Bean Pod Mottle Virus Middle Component RNA.** *Virology* 1991, **183**:405-409.
 43. Larson SB, Day JS, McPherson A: **Satellite tobacco mosaic virus refined to 1.4 Å resolution.** *Acta Crystallogr D Biol* 2014, **70**:2316-2330.
 44. Tang L, Johnson KN, Ball LA, Lin T, Yeager M, Johnson JE: **The structure of Pariacoto virus reveals a dodecahedral cage of duplex RNA.** *Nat Struct Biol* 2001, **8**:77-83.
 45. Larman BC, Dethoff EA, Weeks KM: **Packaged and free satellite tobacco mosaic virus (STMV) RNA genomes adopt distinct conformational states.** *Biochemistry* 2017, **56**:2175-2183
- This is one of the few chemical probing studies reporting on the differences between the secondary structures of the 'free' (solution-equilibrated) and 'in virio' (encapsidated) states of ssRNA genomes, in this case STMV. The changes in secondary structure are correlated with enhancement of RNA-capsid interactions involving RNA duplexes.
46. Garmann RF, Gopal A, Athavale SS, Knobler CM, Gelbart WM, Harvey SC: **Visualizing the global secondary structure of a viral RNA genome with cryo-electron microscopy.** *RNA* 2015, **21**:877-886.
 47. Koning RI, Gomez-Blanco J, Akopjana I, Vargas J, Kazaks A, Tars K, Carazo JM, Koster AJ: **Asymmetric cryo-EM structure of phage MS2 reveals genome structure *in situ*.** *Nat Commun* 2016, **7**:12524
- The authors report the first asymmetric reconstruction of a ssRNA virus, MS2, discovering that a significant fraction of the RNA genome can be resolved.
48. Dai X, Li Z, Lai M, Shu S, Du Y, Hong Zhou Z, Sun R: ***In situ* structures of the genome and genome-delivery apparatus in an ssRNA virus.** *Nature* 2017, **541**:112-116
- This work reports the asymmetric reconstruction of MS2 at sufficient resolution to assign 80% of the RNA backbone configuration. The dominant role of the symmetry-breaking maturation protein in the capsid is highlighted, through its interaction with key stem loops in the RNA genome.
49. Toropova K, Basnak G, Twarock R, Stockley PG, Ranson NA: **The three-dimensional structure of genomic RNA in bacteriophage MS2: implications for assembly.** *J Mol Biol* 2008, **375**:824-836.
 50. Stockley PG, Twarock R, Bakker SE, Barker AM, Borodavka A, Dykeman E, Ford RJ, Pearson AR, Phillips SEV, Ranson NA, Tuma R: **Packaging signals in single-stranded RNA viruses: nature's alternative to purely electrostatic assembly mechanism.** *J Biol Phys* 2013, **39**:277-287
- The authors present general arguments and specific examples – involving novel experimental and the theoretical techniques – demonstrating the role of RNA-sequence-and-secondary-structure-dependent 'packaging signals' in orchestrating the formation and stabilization of capsids by ssRNA viral genomes.
51. Shibu T, Suzuki Y: **Localization of a protein in the RNA-a protein complex of RNA phage MS2.** *Biochim Biophys Acta* 1981, **654**:249-255.
 52. Gorzelnik KV, Cui A, Reed CA, Jakana J, Young R, Zhang JX: **Asymmetric cryo-EM structure of the canonical *Allolevivirus* Qβ reveals a single maturation protein and the genomic ssRNA *in situ*.** *Proc Natl Acad Sci U S A* 2016, **113**:11519-11524.