

## The icosahedral RNA virus as a grotto: organizing the genome into stalagmites and stalactites

Stephen C. Harvey · Yingying Zeng ·  
Christine E. Heitsch

Received: 28 January 2013 / Accepted: 8 March 2013 / Published online: 27 April 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** There are two important problems in the assembly of small, icosahedral RNA viruses. First, how does the capsid protein select the viral RNA for packaging, when there are so many other candidate RNA molecules available? Second, what is the mechanism of assembly? With regard to the first question, there are a number of cases where a particular RNA sequence or structure—often one or more stem-loops—either promotes assembly or is required for assembly, but there are others where specific packaging signals are apparently not required. With regard to the assembly pathway, in those cases where stem-loops are involved, the first step is generally believed to be binding of the capsid proteins to these “fingers” of the RNA secondary structure. In the mature virus, the core of the RNA would then occupy the center of the viral particle, and the stem-loops would reach outward, towards the capsid, like stalagmites reaching up from the floor of a grotto towards the ceiling. Those viruses whose assembly does not depend on protein binding to stem-loops could have a different structure, with the core of the RNA lying just under the capsid, and the fingers reaching down into the interior of the virus, like stalactites. We review the literature on these alternative structures, focusing on RNA selectivity and the assembly mechanism, and we propose experiments aimed at determining, in a given virus, which of the two structures actually occurs.

**Keywords** RNA virus · RNA secondary structure · Virus assembly · Icosahedral viruses

---

S. C. Harvey (✉) · Y. Zeng  
School of Biology, Georgia Institute of Technology, Atlanta, GA 30332, USA  
e-mail: steve.harvey@biology.gatech.edu

C. E. Heitsch  
School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

## 1 Introduction

Viruses were among the first biological objects to be investigated by physicists, beginning with the founding of the “Phage School” by Max Delbrück shortly after World War II. Viruses often combine great biological complexity with apparent physical simplicity, and the main goal of that early work was the determination of the principles of viral structure. In recent years, this work has expanded into investigations of the mechanisms of viral assembly [1–3]. Such understanding should offer novel opportunities for pharmacological intervention, and it should facilitate the design of a wide range of useful nanoparticles. As a consequence, these processes are currently under intense experimental and computational investigation.

Small, icosahedral RNA viruses are among the most popular objects of current studies of structure and assembly. These viruses contain a single-stranded RNA (ssRNA) genome surrounded by a protein shell (the capsid). Some important human pathogens are small ssRNA viruses (e.g., hepatitis E, rubella, rhinovirus). Assembly of ssRNA viruses is spontaneous, at least in the cells that they infect (some viruses can be assembled *in vitro*, but others cannot).

At present, we do not understand how the formation of the mature virus depends on the secondary and tertiary structure of the RNA molecule that is packaged. We do not know to what extent these structures vary from particle to particle in an ensemble of mature viruses, we do not know whether these structures change (whether the RNA refolds) as the capsid proteins bind to the RNA, and we do not know what pathway(s) the RNA follows in achieving its final packaged structure.

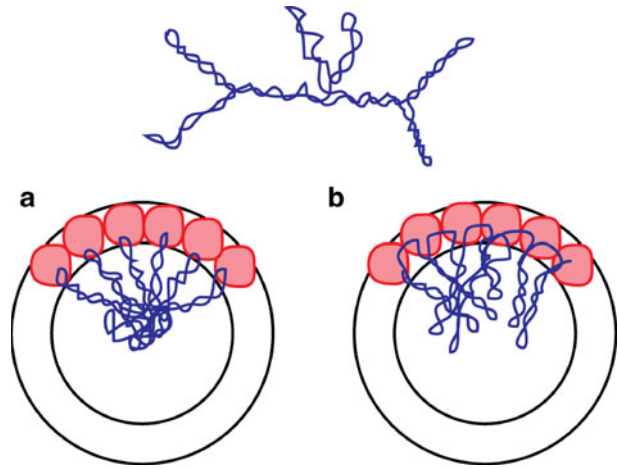
For several ssRNA viruses, studies using X-ray crystallography have revealed the capsid structure at atomic resolution. This work takes advantage of the icosahedral symmetry of the virus because the presence of multiple copies of identical proteins in identical conformations enhances the signal-to-noise ratio. Examples of particular interest to us in this work include satellite tobacco mosaic virus (STMV) [4, 5], Pariacoto virus (PaV) [6], and MS2 [7]. For STMV and PaV, the crystal structures revealed some of the RNA density, but the icosahedral averaging precluded identification of the sequences associated with the density. The MS2 crystal structure did not reveal any information on the structure of the RNA, but the crystal structure was later determined for a virus-like particle obtained by soaking an RNA stem-loop into previously crystallized MS2 capsids [8].

The crystallographic studies have been supplemented by single-particle reconstructions using cryo-electron microscopy (cryo-EM) [6, 9, 10]. Some of these studies use the native virus, while others target virus-like particles obtained by assembling the capsid proteins around RNA molecules other than those of the native genomes.

The combination of X-ray crystallography and cryo-EM has provided glimpses of the RNA organization inside these viruses. Unfortunately, the RNA genome does not possess a highly symmetric sequence, so the icosahedral averaging blurs all features of the RNA structure except those that do have approximate icosahedral symmetry. The genomes of STMV and PaV both contain multiple double-helical regions that vary in sequence and in structural details, but that are visible in the structural studies. We have developed detailed all-atom models for the genomic RNA of these two viruses, which, when added to the previously determined capsid structures, provide the first complete models for any viruses.

The two models that we developed [11, 12] are based on two fundamentally different assumptions about how the RNA is arranged in the virus. As we will soon explain, the assumptions lead to a “stalactite model” for PaV and a “stalagmite model” for STMV (Fig. 1).

**Fig. 1** When an RNA molecule (*top*) is packaged into a virus, it may be organized in either of two ways. **a** The body of the RNA occupies the centermost part of the virus, and the stem-loops reach outward toward the capsid, much like stalagmites in a grotto. **b** The body of the RNA is associated with the capsid proteins, and the stem-loops reach inward toward the center of the virus, resembling stalactites in a grotto



In this article, we review previous work on these two viruses and others, asking whether the two types of structures are equally probable in the world of small, icosahedral ssRNA viruses. We also examine the probable implications for viral assembly for the stalactite and stalagmite models.

## 2 The structure of the Pariacoto virus: a stalactite model

Pariacoto virus (PaV) is a  $T = 3$ , non-enveloped, icosahedral virus. A member of the Nodaviridae family, it was originally isolated from the Southern Armyworm, *Spodoptera eridania*, the larva of a Peruvian moth [13]. The PaV genome consists of two positive-sense ssRNAs [14], one of which (RNA1, 3011 nucleotides) codes for protein A, the catalytic subunit for the RNA-dependent RNA replicase, while the other (RNA2, 1311 nucleotides) codes for the capsid precursor protein, designated  $\alpha$ . The relatively small size (30 nm diameter) and the ease with which it can be produced in various cell lines [15] make PaV and other members of the Nodaviridae family relatively easy to characterize at the molecular level [16–18].

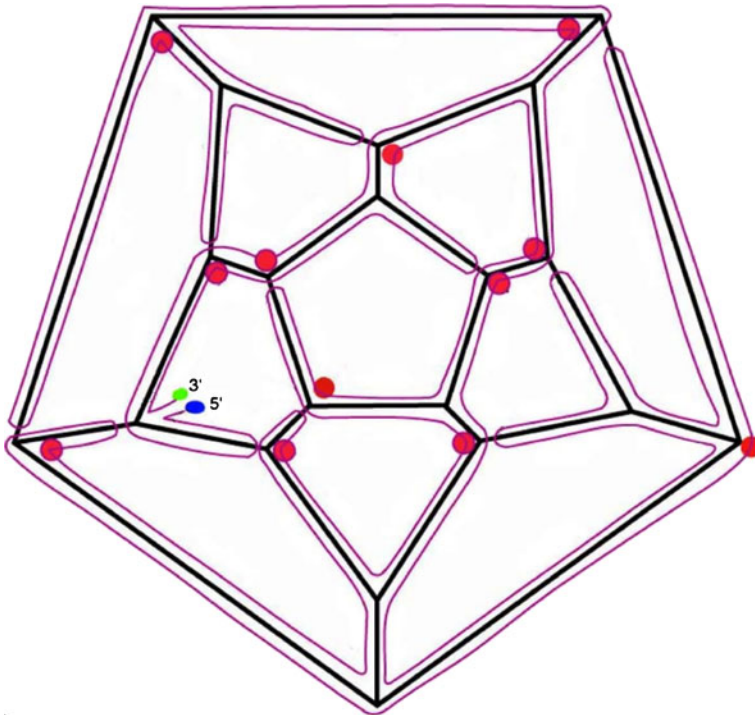
The crystal structure of PaV [6] revealed an asymmetric unit containing three quasi-equivalent protein subunits (A, B, and C). It also contains 30 segments of double-helical RNA (each 25 base pairs long) just inside the capsid. The sequence does not contain 30 copies of any sequence of that length, so the 30 RNA duplexes certainly have a variety of sequences, and probably some differences in structure, e.g., mismatches and bulges. As a consequence, the double-helical structure seen in the crystal structure represents an average of these duplexes. Taken together, the 1,500 nucleotides in the 30 duplexes represent about 35% of the total genome—but no particular part of the RNA can be assigned to any part of these duplexes.

The RNA duplexes are perpendicular to the twofold crystallographic axes, and they lie on the 30 edges of a dodecahedron, forming a “dodecahedral cage” (note that the dodecahedron and the icosahedron form a pair of complementary Platonic solids: the former has 12 faces and 20 vertices, while the latter has 20 faces and 12 vertices; each has 30 edges, but the edges of the icosahedron are perpendicular to those of the dodecahedron, and vice versa).

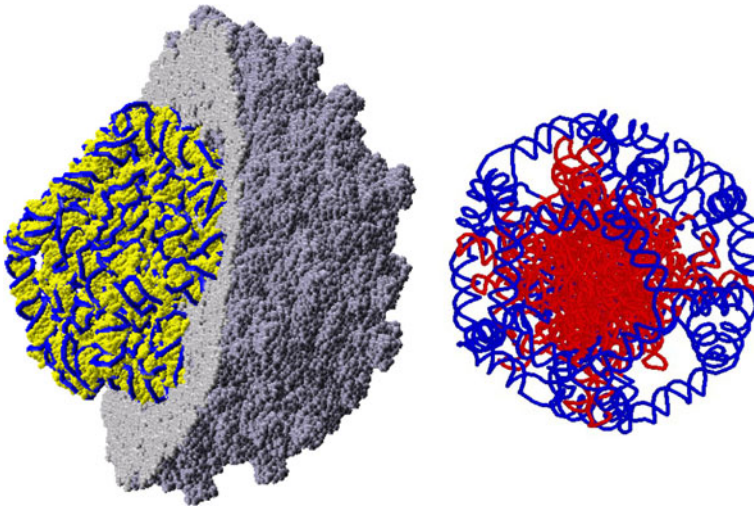
We were challenged to develop an all-atom model for PaV by Jack Johnson, Liang Tang, and Annette Schneemann, who had done the original crystallographic and cryo-EM studies on PaV [6]. Their examination of the density maps led them to conclude that the remaining RNA (the 65% not seen in the crystal structure) must occupy the centermost region of the virus, and that it is connected to the dodecahedral cage at the vertices. This led us to jointly postulate a stalactite conformation (Fig. 1b) for the RNA.

The RNA secondary structure is not known, nor is it known whether all base pairing in RNA1 and RNA2 is intramolecular, or whether there are base pairs between RNA1 and RNA2. In the absence of such information, we chose to represent the bipartite genome as one single strand. With this assumption, and in the absence of data on the RNA secondary structure, we had to postulate a secondary structure that is consistent with the crystal structure (30 anti-parallel duplexes whose axes define the dodecahedral cage) and with the stalactite model. It is easily shown that there are many possible ways to satisfy these constraints. We chose a representative solution [9], as shown in Fig. 2.

To build a three-dimensional model, we modeled the stalactites with pieces taken from the 23 S ribosomal RNA of *E. coli*, and we invented a sequence compatible with the secondary structure for those parts of the RNA on the dodecahedral cage and the vertices.



**Fig. 2** Organization of a hypothetical secondary structure for PaV RNA, in the stalactite conformation [9]. This is just one solution for mapping the RNA onto the dodecahedral cage in a fashion that is consistent with the data from X-ray crystallography and cryo-electron microscopy. The 5' and 3' ends are indicated. Each edge of the dodecahedral cage contains an antiparallel RNA double helix. Red circles indicate three- and four-way junctions at those vertices where RNA stalactites drop down into the interior of the virus to connect with that part of the RNA not on the dodecahedral cage



**Fig. 3** All-atom stalactite model for Pariacoto virus (PaV) [11]. On the *left* is a view with half of the capsid (*grey*) cut away to reveal the RNA genome (*yellow*, with the backbone highlighted in *blue*). On the *right* is a representation of the RNA backbone: the *blue regions* are those parts in contact with the capsid, while the stalactites are colored in *red*

The final all-atom model [11] contains every residue of the genome and every residue of the capsid proteins, so it was the first complete model of any virus-like particle (we say “virus-like particle” because it contains an RNA that is the same size as the genomic RNA, but that does not have the genomic sequence). The three-dimensional model is shown in Fig. 3.

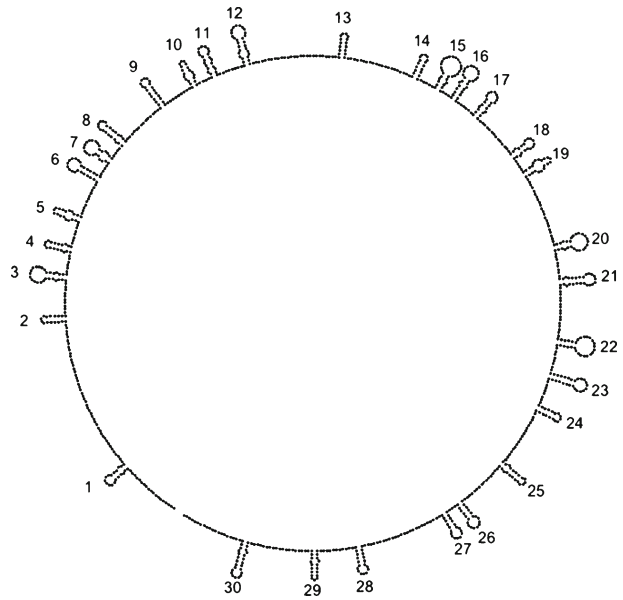
### 3 The structure of satellite tobacco mosaic virus: a stalagmite model

Satellite tobacco mosaic virus (STMV) is a  $T = 1$  icosahedral virus with a diameter of 17 nm and a single-stranded RNA genome containing 1,058 nucleotides [19]. The 1.8 Å X-ray structure of this virus showed 30 RNA double helices, each one centered on a twofold symmetry axis [4]. Each helix has nine base pairs, plus an unpaired nucleotide at the 3' end of each strand and 57% of the genome is visible in the crystal structure [4]. Since the viral particle can enter the crystal lattice in any of 60 orientations, the RNA density is averaged among these orientations, so the crystal structure provides no information on the sequence of each double helix.

Larson and McPherson proposed that the RNA double helices represent 30 hairpin loops, and that they are connected by single-stranded linkers [5]. Subsequent atomic force microscopic studies on RNA released from the capsid are consistent with this linear arrangement of structural domains [20]. Their model for the organization of the genome in the mature virus suggests an efficient pathway for viral assembly. It is a stalagmite model (Fig. 1a).

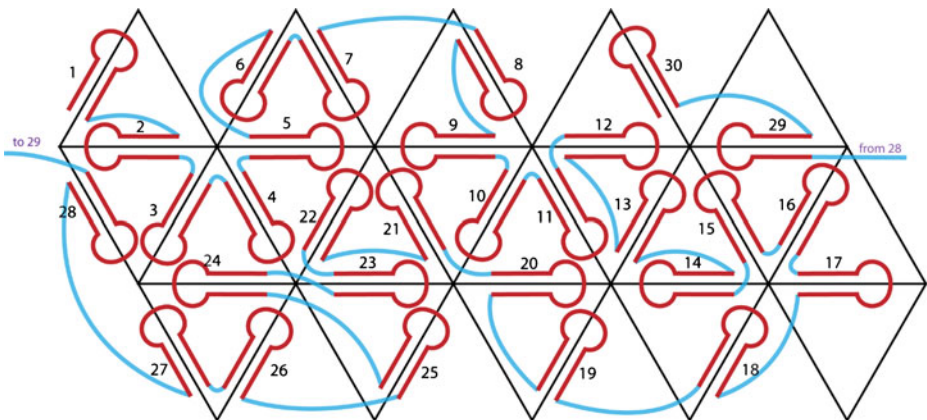
At the time the Larson-McPherson model was built, there were no data to permit assignment of specific regions of the genomic RNA to each of the stem-loops. Schroeder et al. later did chemical probing of the RNA inside the virus; they combined the resulting

**Fig. 4** The secondary structure of STMV RNA proposed by Schroeder et al. [21], based on chemical probing and the assumption of Larson and McPherson that the RNA is organized into a series of stem-loops

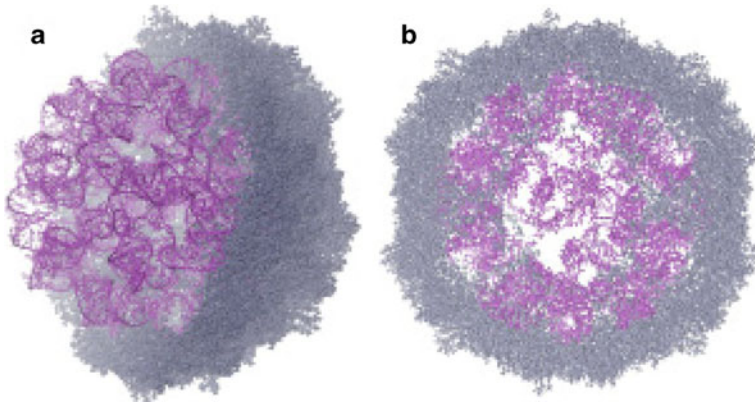


data with helix search algorithms to identify possible combinations of stem-loops, allowing them to propose a genomic secondary structure consistent with the organization originally proposed by Larson and McPherson; they emphasized that this is a representative secondary structure from a large ensemble of possible structures [21]. The model of Schroeder et al. is shown in Fig. 4.

Figure 5 shows how the 30 stem-loops in the Schroeder secondary structure model can be organized so that each edge of the icosahedral structure is covered by one hairpin loop [12]. We arranged these according to the lengths of the connections between successive



**Fig. 5** A model for the organization of the 30 stem-loops of the STMV genomic RNA in STMV, following the proposal of Larson and McPherson [5], and incorporating the secondary structure proposed by Schroeder et al. [21]



**Fig. 6** An all-atom model for STMV [12], with the capsid in *grey* and the RNA genome in *violet*. On the *left* is a cutaway view showing half the capsid and the complete RNA genome, while on the *right* is a 50 Å slice through the center of the virus

stem-loops. With linkers of less than eight nucleotides, the neighboring hairpins adopt a tail-to-tail conformation, while head-to-tail or head-to-head conformations can be used when the linker is longer than 13 nucleotides. Of course, the arrangement shown in Fig. 5 is just one out of a very large ensemble of possible arrangements.

We built and refined a model for the STMV RNA, which we combined with the the X-ray structure of the capsid [4]. The capsid proteins have 147 residues, but the 12 amino-terminal amino acids are not seen in the crystal structure, so we also modeled those. The final model [12] is shown in Fig. 6. We believe that this is the first complete all-atom model for any virus that incorporates the actual sequence of the genomic RNA.

(For completeness, we should note that we have recently determined the secondary structure of the STMV genome by chemical probing of an *in vitro* transcript [22], finding a very extended structure that is quite different from the Schroeder model; we do not yet know if this structure can be packaged into the virus without refolding. We have argued that if the RNA *in vitro* does have the structure proposed by Schroeder et al., then it is either packaged co-transcriptionally, or there is extensive refolding upon packaging.)

#### 4 MS2: a recent addition

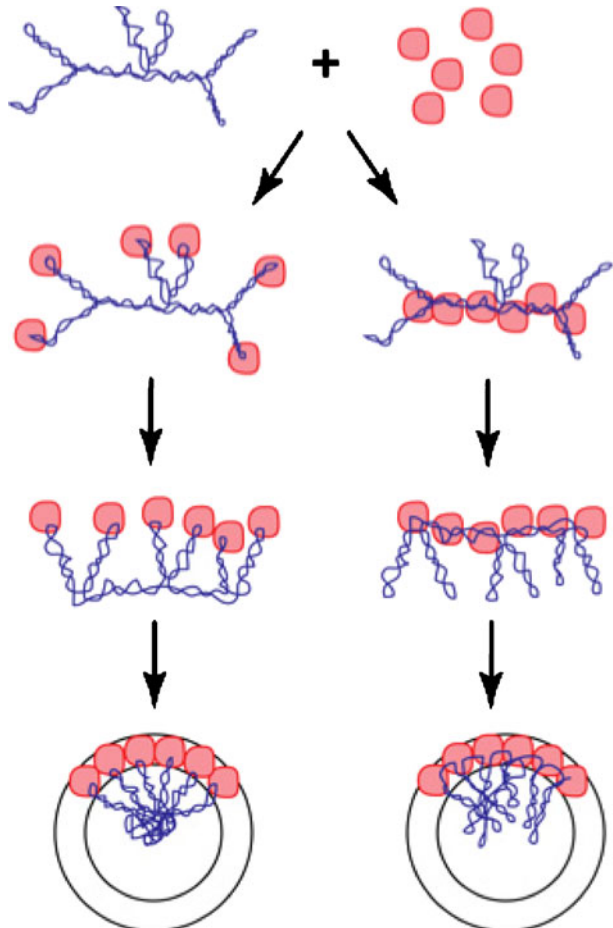
MS2 is a  $T = 3$  icosahedral single-stranded RNA bacteriophage with a genome of 3,569 nucleotides. The crystal structure of the wild-type virion has been determined at 2.8 Å [7, 23]. Capsid assembly can be triggered *in vitro* by an RNA stem-loop containing only 19 nucleotides (the TR stem-loop) [24, 25], and the structure of a virus-like particle assembled from capsid proteins and 60 copies of the TR stem-loop has been determined by cryo-EM [10]. Instead of TR, one can also trigger assembly by using RNA aptamers selected for their binding affinity to the capsid protein [26]. One of these (F5) can be modified to give tighter binding by replacing a single adenosine with 2'-deoxy-2-aminopurine; the 2.8 Å crystal structure of this aptamer soaked into pre-crystallized recombinant capsids revealed the details of the RNA–protein interaction [8].

The combination of these experimental results clearly shows that MS2 genomic RNA is organized into a stalagmite structure inside the mature virus (we note here that we have developed an all-atom model for MS2, in collaboration with the research groups of Peter Stockley (University of Leeds) and Reidun Twarock (University of York), and we expect to submit that model for publication in the near future).

## 5 Discussion and conclusions

Structural studies have produced two classes of models for the organization of RNA genomes inside small icosahedral viruses. The combination of data from X-ray crystallography and cryo-EM on PaV [6] led us to suggest that the genome has a stalactite structure [9]. In contrast, the crystal structure of STMV [4] was interpreted as favoring the organization of the genome into a stalagmite structure [5]. To date, there are no concrete data to prove that either of these is correct, but we have shown that it is possible to build all-atom models where the genome is organized into either the stalactite structure [11] or the stalagmite structure [12].

**Fig. 7** Alternative pathways for the formation of small icosahedral RNA viruses. RNA (blue) and capsid proteins (red) can interact in either of two ways. In the left-hand pathway, the proteins bind to the tips of RNA stem-loops. Protein–protein interactions lead toward the formation of the mature virus with the RNA in the stalagmite conformation. Alternatively, capsid proteins might bind predominantly to the body of the RNA (right-hand pathway), leading to the stalactite conformation (only a small fragment of the RNA and a fraction of the capsid proteins are shown here, for graphical clarity. In the mature virus, the capsid proteins would form an icosahedral or spherical structure)





**Table 1** Experimental search for alternative packaging pathways

	<i>RNA stalagmites:</i> Proteins bind to stem-loops	<i>RNA stalactites:</i> Proteins bind to RNA core
Nucleate with specific sequences?	YES	NO
Nucleate with pieces of RNA containing no base pairs?	Possibly	Possibly
Nucleate with double-stranded RNA?	NO	YES

Observations are highlighted in *red*, and predictions are highlighted in *blue*

The one case where the genome organization is known for certain is MS2, where the combination of X-ray crystallography [8] and cryo-EM [10] clearly demonstrates the stalagmite structure. Assembly requires the existence of specific stem-loop structures, which is consistent with the proposed assembly pathway for stalagmites and inconsistent with the pathway for stalactite structures (Fig. 7 and Table 1).

Other viruses are also known to require specific packaging signals, including brome mosaic virus (BMV) [27], which is another T = 3 single-stranded RNA virus (with a tetrapartite genome). In contrast, cowpea chlorotic mottle virus does not require a specific packaging signal [28], even though CCMV and BMV are structurally and genetically very similar [29]. It will be interesting to see whether the stalagmite pattern turns out to be universal in cases that do have packaging signals and to determine the organization of RNA genomes in those viruses that do not require such signals.

The question of how the RNA genome is organized is of critical importance for elucidating the assembly pathway. As shown in Fig. 7, the assembly pathways are necessarily quite different for viruses whose genomes have the stalactite conformation in the mature virus vs. those with the stalagmite conformation. If the capsid proteins bind to the stem-loops of the RNA (left-hand pathway), then the body of the RNA will lie in toward the center of the mature virus, with the stem-loops reaching out toward the capsid-like stalagmites. On the other hand, if the capsid proteins bind to the body of the RNA, holding it just under the capsid (right-hand pathway), then the stem-loops will hang downward from the capsid toward the center of the virus, like stalactites.

It should be possible to experimentally determine which of the two conformations is obtained by the RNA genome in a given icosahedral virus. Table 1 outlines the kinds of experiments that should be able to distinguish between the stalactite and stalagmite conformations.

**Acknowledgements** This work was first presented at the Workshop on Physical Virology at the Abdus Salam International Centre for Theoretical Physics in Trieste, Italy (September, 2012). SCH is grateful to ICTP and the organizers for providing travel support to that meeting. Use of the word “grotto” in the title was inspired by the Grotta Gigante, near Trieste. Supported by grants from the US National Institutes of Health (GM 070785 to SCH and GM 083621 to CEH).

## References

1. Knobler, C.M., Gelbart, W.M.: Physical chemistry of DNA viruses. *Annu. Rev. Phys. Chem.* **60**, 367–383 (2009)
2. Jardine, P.J., Anderson, D.L.: DNA packaging in double-stranded DNA phages. In: Calendar R. (ed.) *The Bacteriophages*, 2nd edn., pp. 49–65. Oxford University Press, Oxford (2006)

3. Johnson, J.E., Chiu, W.: DNA packaging and delivery machines in tailed bacteriophage. *Curr. Opin. Struct. Biol.* **17**, 237–243 (2007)
4. Larson, S.B., Day, J., Greenwood, A., McPherson, A.: Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J. Mol. Biol.* **277**, 37–59 (1998)
5. Larson, S.B., McPherson, A.: Satellite tobacco mosaic virus RNA: structure and implications for assembly. *Curr. Opin. Struct. Biol.* **11**, 59–65 (2001)
6. Tang, L., Johnson, K.N., Ball, L.A., Lin, T., Yeager, M., Johnson, J.E.: The structure of Pariaicoto virus reveals a dodecahedral cage of duplex RNA. *Nat. Struct. Biol.* **8**, 77–83 (2001)
7. Valegard, K., Liljas, L., Fridborg, K., Unge, T.: The three-dimensional structure of the bacterial virus MS2. *Nature* **345**, 36–41 (1990)
8. Horn, W.T., Convery, M.A., Stonehouse, N.J., Adams, C.J., Liljas, L., Phillips, S.E., Stockley, P.G.: 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)
9. Tihova, M., Dryden, K.A., Le, T.V., Harvey, S.C., Johnson, J.E., Yeager, M., Schneemann, A.: Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *J. Virol.* **78**, 2897–2905 (2004)
10. 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)
11. Devkota, B., Petrov, A.S., Lemieux, S., Boz, M.B., Tang, L., Schneemann, A., Johnson, J.E., Harvey, S.C.: Structural and electrostatic characterization of Pariaicoto virus: Implications for viral assembly. *Biopolymers* **91**, 530–538 (2009)
12. Zeng, Y., Larson, S.B., Heitsch, C.E., McPherson, A., Harvey, S.C.: A model for the structure of satellite tobacco mosaic virus. *J. Struct. Biol.* **180**, 110–116 (2012)
13. Zeddiam, J.L., Rodriguez, J.L., Ravallec, M., Lagnaoui, A.: A Noda-like virus isolated from the sweet-potato pest *Spodoptera eridania* (Cramer) (Lep.; Noctuidae). *J. Invertebr. Pathol.* **74**, 267–274 (1999)
14. Krishna, N.K., Schneemann, A.: Formation of an RNA heterodimer upon heating of nodavirus particles. *J. Virol.* **73**, 1699–1703 (1999)
15. Schneemann, A., Zhong, W., Gallagher, T.M., Rueckert, R.R.: Maturation cleavage required for infectivity of a nodavirus. *J. Virol.* **66**, 6728–6734 (1992)
16. Fisher, A.J., McKinney, B.R., Wery, J.P., Johnson, J.E.: Crystallization and preliminary data analysis of Flock House virus. *Acta Crystallogr. B.* **48**(Pt 4), 515–520 (1992)
17. Mori, K., Nakai, T., Muroga, K., Arimoto, M., Mushiake, K., Furusawa, I.: Properties of a new virus belonging to Nodaviridae found in larval striped jack (*Pseudocaranx dentex*) with nervous necrosis. *Virology* **187**, 368–371 (1992)
18. Reinganum, C., Bashiruddin, J.B., Cross, G.F.: Boolarra virus: a member of the Nodaviridae isolated from *Oncopera intricoides* (Lepidoptera: Hepialidae). *Intervirology* **24**, 10–17 (1985)
19. Mirkov, T.E., Mathews, D.M., Du Plessis, D.H., Dodds, J.A.: Nucleotide sequence and translation of satellite tobacco mosaic virus RNA. *Virology* **170**, 139–146 (1989)
20. Kuznetsov, Y.G., Dowell, J.J., Gavira, J.A., Ng, J.D., McPherson, A.: Biophysical and atomic force microscopy characterization of the RNA from satellite tobacco mosaic virus. *Nucleic Acids Res.* **38**, 8284–8294 (2010)
21. Schroeder, S.J., Stone, J.W., Bleckley, S., Gibbons, T., Mathews, D.M.: Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints. *Biophys. J.* **101**, 167–175 (2011)
22. Athavale, S.S., Gossett, J.J., Bowman, J.C., Hud, N.V., Williams, L.D., Harvey, S.C.: Secondary structure of the free RNA of satellite tobacco mosaic virus. *PLoS One* **8**, e54384 (2013)
23. Golmohammadi, R., Valegard, K., Fridborg, K., Liljas, L.: The refined structure of bacteriophage MS2 at 2.8 Å resolution. *J. Mol. Biol.* **234**, 620–639 (1993)
24. Beckett, D., Uhlenbeck, O.C.: Ribonucleoprotein complexes of R17 coat protein and a translational operator analog. *J. Mol. Biol.* **204**, 927–938 (1988)
25. Beckett, D., Wu, H.N., Uhlenbeck, O.C.: Roles of operator and non-operator RNA sequences in bacteriophage R17 capsid assembly. *J. Mol. Biol.* **204**, 939–947 (1988)
26. Hira, I., Spingola, M., Peabody, D., Ellington, A.D.: The limits of specificity: an experimental analysis with RNA aptamers to MS2 coat protein variants. *Mol. Divers.* **4**, 75–89 (1998)
27. Choi, Y.G., Rao, A.L.: Packaging of brome mosaic virus RNA3 is mediated through a bipartite signal. *J. Virol.* **77**, 9750–9757 (2003)
28. Annamalai, P., Rao, A.L.: Dispensability of 3′ tRNA-like sequence for packaging cowpea chlorotic mottle virus genomic RNAs. *Virology* **332**, 650–658 (2005)
29. Osman, F., Choi, Y.G., Grantham, G.L., Rao, A.L.: Molecular studies on bromovirus capsid protein. *Virology* **251**, 438–448 (1998)