UC Irvine

Title
Satellite tobacco mosaic virus RNA: structure and implications for assembly

Authors
Larson, Steven B
McPherson, Alexander

Copyright Information

Peer reviewed

Satellite tobacco mosaic virus RNA: structure and implications for assembly

Steven B Larson and Alexander McPherson*

The initial appearance of 45% of the single-stranded RNA of satellite tobacco mosaic virus in electron density maps suggested the entire RNA conformation could be delineated. Subsequent work has localized nearly 80% of the RNA as stem-loop elements. Connection of the stem-loops in the most efficient manner produces a persuasive model for the encapsidated RNA. This arrangement has significant implications for virus assembly and for the essential role of RNA.

Addresses
Department of Molecular Biology and Biochemistry, 560 Steinhaus Hall, University of California Irvine, Irvine, CA 92697-3900, USA
*e-mail: amcphers@uci.edu

Abbreviations
ss   single-stranded
STMV   satellite tobacco mosaic virus

Introduction

The structure of satellite tobacco mosaic virus (STMV), a T = 1 icosahedral particle of 17 nm diameter, has a single-stranded (ss) RNA genome of 1058 nucleotides [1]. Secondary structure prediction [2,3] indicates possible base pairing approaching 70%, a feature shared by other satellite viruses and satellite RNAs [4–6]. Its 1.8 Å X-ray structure [7–9] revealed double-helical segments of RNA centered on all dyad axes of the virion, 30 in all, In addition, an isolated nucleotide having an occupancy of half that of the helical nucleotides was observed some distance from the ends of each helix, apparently an ordered component of an otherwise disordered region of the RNA. The helical segments of encapsidated RNA, which reflect the icosahedrally averaged structure, comprise nine base pairs, plus additional nucleotides stacked on the helix ends. Temperature factors increased steeply from the central base pair, suggesting that the helix ends were disordered because of the diverse paths assumed by single strands of RNA emerging from helical segments and/or that some helices were less than nine base pairs.

The 'free nucleotide' could not be included at either end of a helix by brief extension of its helix geometry, but its fixed position and substantial occupancy implied that it was a component of some repetitive, but variable, substructure of the RNA. It was ultimately interpreted as arising from distal nucleotides of single-stranded loops that close one end of each helical segment [7], thereby yielding half occupancy. More recent studies using virions containing partially

digested RNA (A McPherson et al., unpublished data) essentially confirm this assignment.

In sum, the RNA visible by X-ray diffraction, shown in Figure 1, implied that nonidentical stem-loop elements occurred at every edge of the icosahedron. A similar arrangement of stem-loop RNA elements was earlier proposed for southern bean mosaic virus [10] and for bean pod mottle virus [11]. The distribution of stem-loops, based on the stem sizes deduced from crystallography, and the loop size, based on modeling [12], which requires at least nine nucleotides to create a stereochemically acceptable model, would consume a minimum of 80% of the 1058 nucleotides.

The structure of STMV RNA must, when free in the host cell, accommodate both a TMV replicase recognition site at its 3′ terminus [13,14] and a histidinylatable tRNA-like structure as well [15]. Other discrete three-dimensional domains are also, of necessity, formed during RNA replication, translation, intercellular transport and other physiological processes [16–18]. An inescapable conclusion is that the secondary and tertiary structure of RNA when encapsidated must differ from that when free in the cell [17,18]. There is insufficient space inside the capsid, given the icosahedral distribution of stem-loops, to accommodate other significant structural domains. Furthermore, there would be insufficient nucleotides to create such structures after all of the stem-loops and connecting links suggested by the X-ray structure were made. Finally, the ensemble of helices encapsidated within the virus is incompatible with physiological processes requiring linear nucleic acid. Thus, the RNA must exist in at least two, and probably many, conformations during the life cycle of the virus. It probably exhibits a conformationally fluid character as it fulfills its physiological obligations.

Results from X-ray diffraction, together with data from genetics, biochemistry and physical chemistry, provide a foundation for a model of the entire RNA molecule encapsidated within STMV. This model, in turn, has important implications for assembly of the virus. The model and its consequences are examined in this review.

Potential ssRNA secondary structural motifs

The distribution of stem-loops from the X-ray structure could be compatible with numerous RNA secondary patterns. Their distinguishing characteristic is the degree to which long-range pairing interactions are allowed. In the extreme case, illustrated in Figure 2a, the structure is highly self-involved and intricate, with virtually every part of the structure interdependent upon every other. Nucleotide stretches hundreds of bases apart, due to the complicated

Figure 1
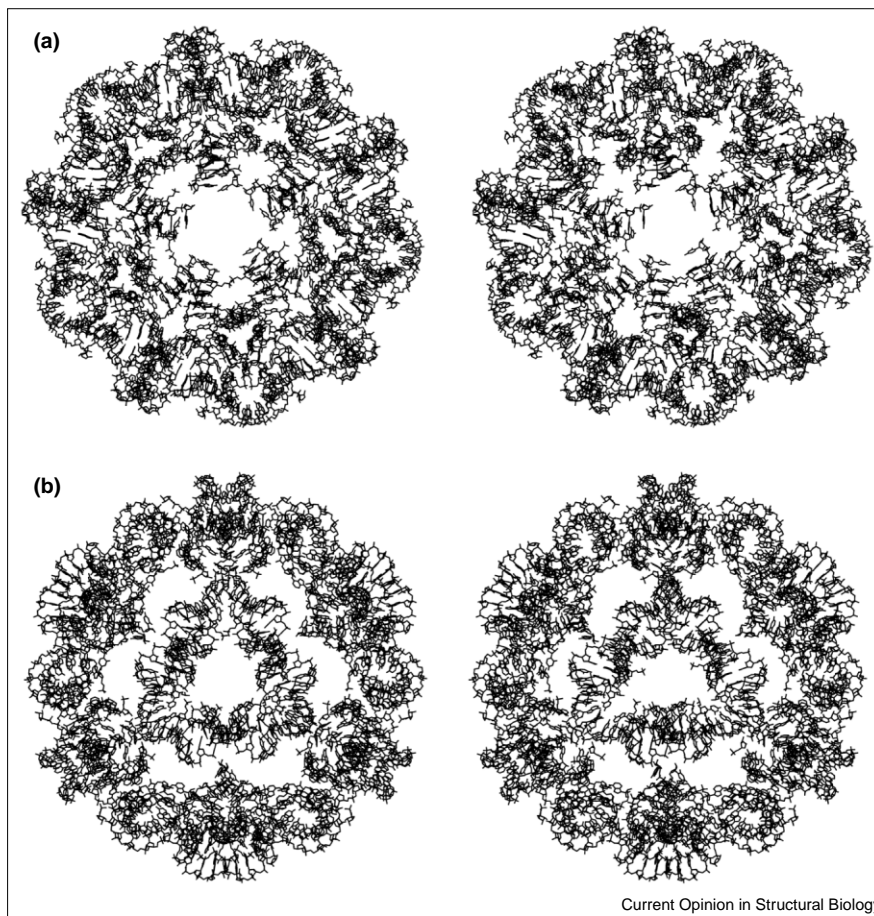
X-ray diffraction analysis of crystals of STMV [7–9] reveals icosahedrally disposed double-helical segments of RNA, each composed of nine base pairs plus stacked bases at either 3′ terminus. In addition, a single, well-defined but isolated nucleotide with about half occupancy is bound by the coat proteins. A hemisphere of the RNA core is seen as stereo views **(a)** along a fivefold axis and **(b)** along a threefold axis. The RNA core has a diameter of about 105 Å and a 60 Å diameter void at its center.

Current Opinion in Structural Biology

folding back of the polynucleotide upon itself, might pair to generate double-stranded regions. This is the kind of structure predicted on the basis of energy considerations and maximization of base pairing [2,3]. It is implicit, however, that the RNA molecule exists in its entirety when it folds into its three-dimensional structure, that is, synthesis is complete. This is, however, probably not true for an RNA daughter strand emerging from a replication complex [16], for which packaging is initiated before synthesis is complete.

Alternatively, the RNA may form a series of local secondary structures, stem-loop elements, dependent upon the pairing of stretches of bases relatively close to one another in the nucleotide sequence, as shown in Figure 2b. This pattern, essentially that proposed by Fresco *et al.* [19], is inherently simple, imposes no rigorous constraints on the overall conformation of the RNA, provides fluidity between states and does not require the RNA be entirely synthesized before folding initiates.

From purely architectural and topological principles, the folding possibility in Figure 2a is unlikely to produce a three-dimensional arrangement consistent with that observed by X-ray diffraction. Not only would it be difficult

to reconcile a rigorously defined and organized set of unique and complex structural elements with the constraints of the encapsidated conformation, but also such an arrangement would be intrinsically unfavorable from a mechanical standpoint. It would exemplify, as pointed out by Crane [20], what engineers term an 'over-determined structure', a brittle structure — one exhibiting so many internal interdependencies, like a framework with too many struts and welds, that it cannot tolerate stress. It is too rigid. An 'over-determined structure' in the case of encapsidated STMV RNA seems particularly improbable as the structure must be fluid and assume other conformations when released from the virion. Furthermore, many non-wildtype strains of the virus exist [21–23] that include multiple point mutations and even some substantial deletions [21,24]. Were the conformation of the encapsidated RNA rigorously defined and closely dependent on precise long-distance pairing arrangements, it probably could not tolerate these mutations.

Reconciliation of secondary structure with the crystallographic constraints is equivalent to placing contiguous RNA stem-loop elements on the 30 twofold symmetry axes of the net shown in Figure 2c, which, when folded by

Figure 2

Possible motifs for a long RNA molecule of roughly 1000 bases that is capable, by folding back upon itself, of substantial intramolecular base pairing. **(a)** Extensive base pairing between regions far from one another in the sequence produces a rigorously defined and highly constrained conformation. **(b)** The RNA exhibits the most unconstrained conformation, reflecting a sequence of strictly local stem-loop substructures. This allows the restructuring of conformations in a fluid manner. Although (a) provides the lowest energy conformation, it also produces the most rigid three-dimensional structure. On the other hand, (b) may yield the least total base pairing, but it also provides the greatest flexibility in terms of three-dimensional structure. **(c)** A two-dimensional grid of 20 triangles that, when folded so that adjacent edges meet, produces a three-dimensional T=1 icosahedron. The twofold, threefold and fivefold symmetry elements of the icosahedron are indicated. **(d)** The most efficient way of placing linearly connected RNA helical segments at all twofold axes of a T=1 icosahedron can be considered in terms of the most efficient way (shortest overall path) of connecting the twofold axis positions on this lattice. A series of linearly connected RNA stem-loop structures, like the RNA molecule in (b), can readily follow many similar, but equally efficient, paths among the 30 twofold axis positions of the icosahedral diagram. The simplest configuration is shown here and produces the shortest overall path. It begins at one fivefold apex at the lower right of the diagram and terminates at the most distant fivefold axis at the upper left.


Current Opinion in Structural Biology
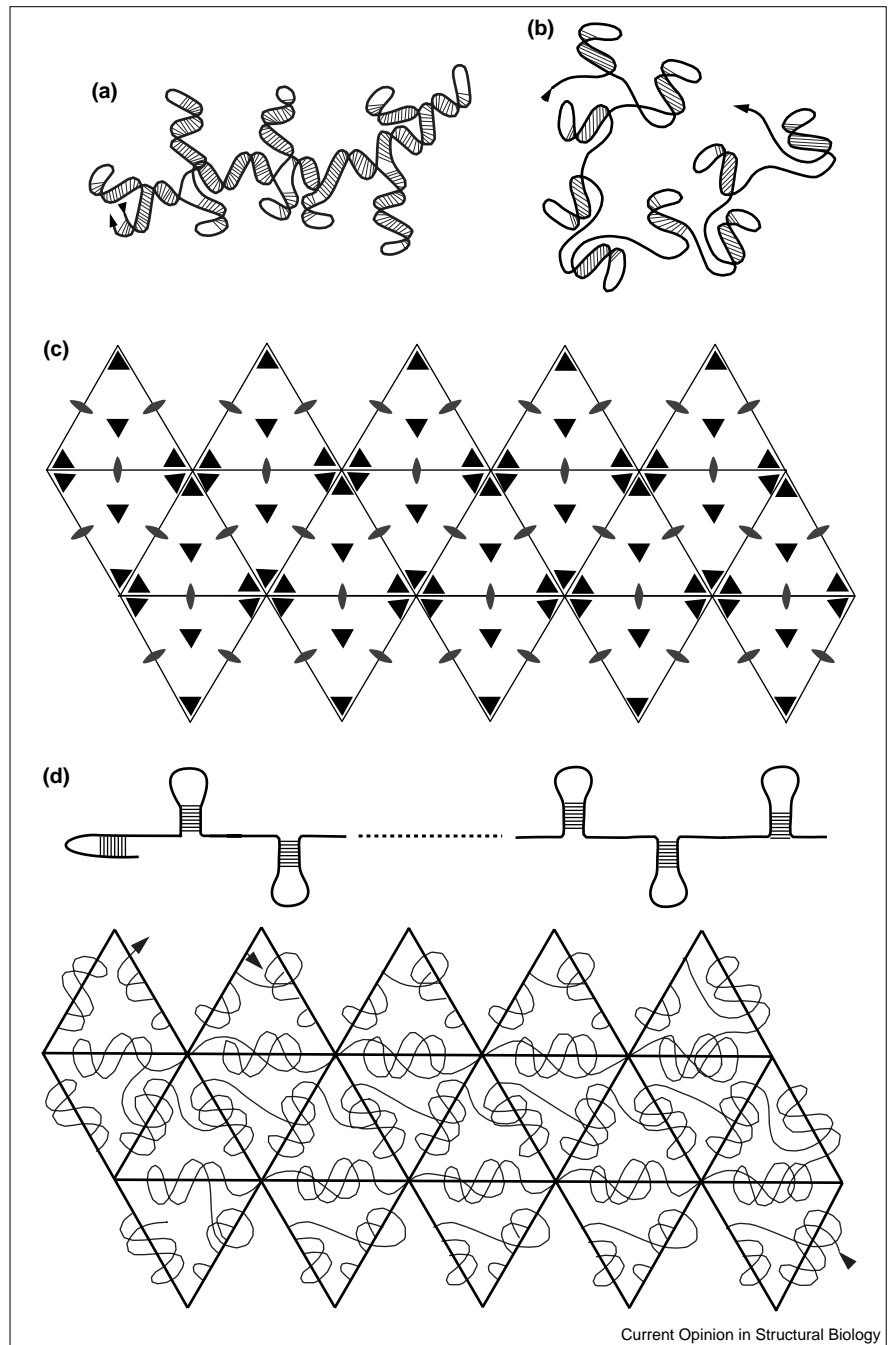
joining all edges, produces a T = 1 icosahedron. It is difficult, from a topological perspective, to envision how a highly interdependent and rigorously defined entity, a nonrepetitive arrangement of helical segments and single-stranded regions, could be efficiently deployed in a manner consistent with the icosahedral grid.

The RNA motif consisting of a series of stem-loop substructures can, however, be made to conform to the encapsidated RNA in a straightforward and uncomplicated way that places little constraint on the overall structure of

the RNA [17–19]. For STMV, such a pattern meets the criterion of Ocam's razor and, as shown in Figure 2d, provides a plausible means of placing 30 helices at the twofold axes of the virus and generating both the loop elements and the connecting links. The RNA could configure itself in an uninterrupted manner as a single daughter strand emerging from the replication complex. Devoid of long-range interdependencies, at least before packaging, it would be relatively insensitive to modest mutation and evolution, and it is a structure compatible with the hypotheses of earlier investigators of RNA structure [17–19,25–27]. It would

Figure 3
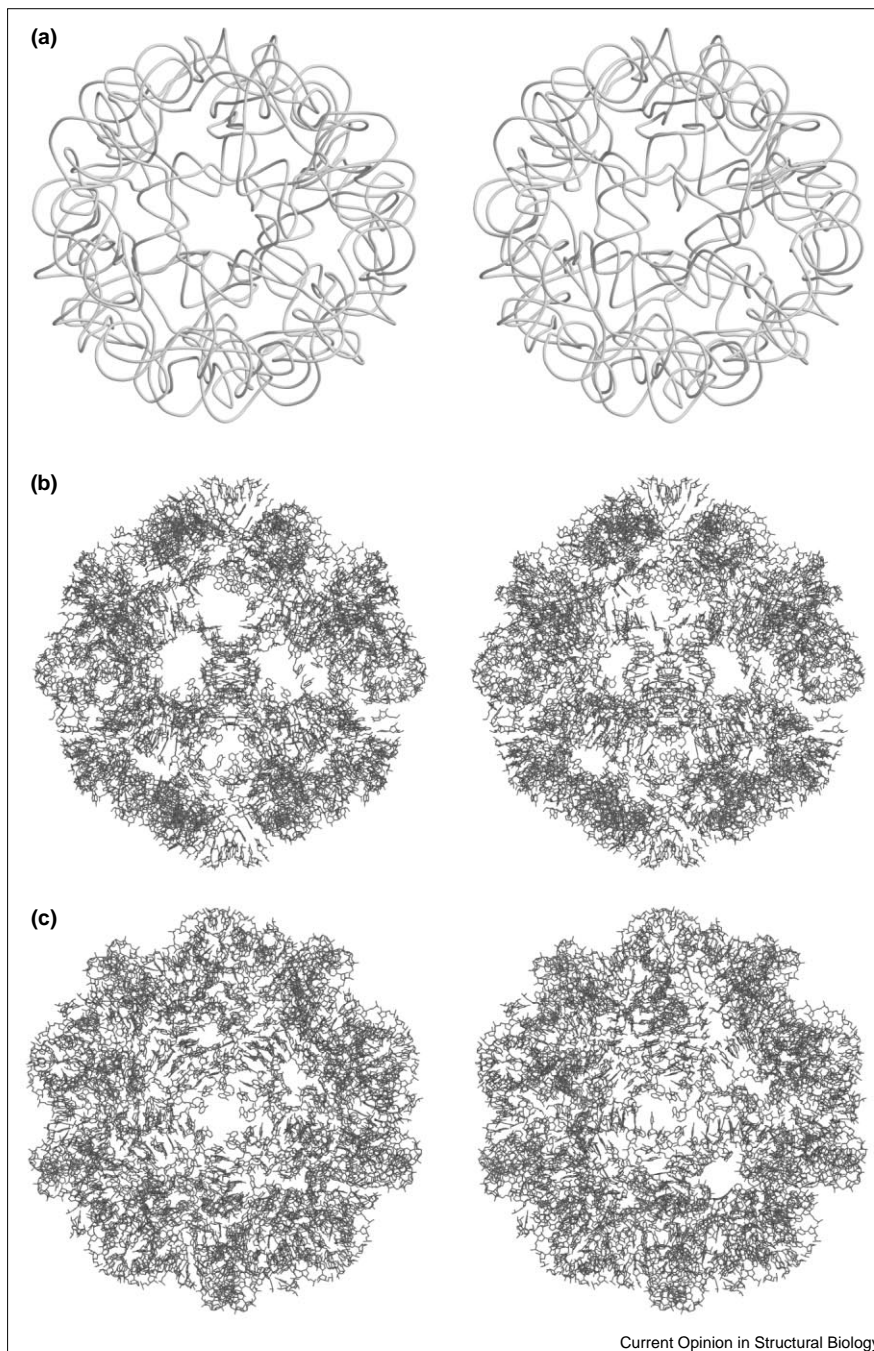
Conformation of encapsidated STMV RNA. **(a)** Stereo diagram, roughly along a fivefold axis, of the course of the RNA chain, represented as a continuous line. **(b,c)** Stereo diagrams, viewed along the twofold and fivefold axes, respectively, of the entire encapsidated RNA molecule modeled by connecting together, in a linear manner, all of the 30 stem-loop elements according to the path in Figure 2d. The model utilizes virtually all the nucleotides and places the stems and 'free nucleotides' in positions consistent with the X-ray diffraction structure.

exhibit conformational fluidity and the capability to transform in a cooperative manner into alternative conformations. Furthermore, the RNA has the attractive property that, as it is synthesized, it would present a series of defined binding sites for the coat protein, which could direct, in an ordered manner, assembly of the viral capsid.

An idealized model of the RNA

The RNA distribution seen in electron density maps is a consensus of what pertains throughout the virion [28]. Every helix need not be the same length, regularity may vary from segment to segment, loops may be different sizes and conformations, and the paired bases of each helical stem are, of course, different. The minimum number of helix loop elements consistent with the X-ray structure would consume 870 of 1058 nucleotides in the RNA, leaving at most 188 nucleotides to make single-stranded linkages. Economy in the expenditure of nucleotides, therefore, is of considerable weight and short paths between free ends of helical stems would be preferred. Few diversions between helices could be permitted and deployment of helix stem elements in an efficient and

Figure 4
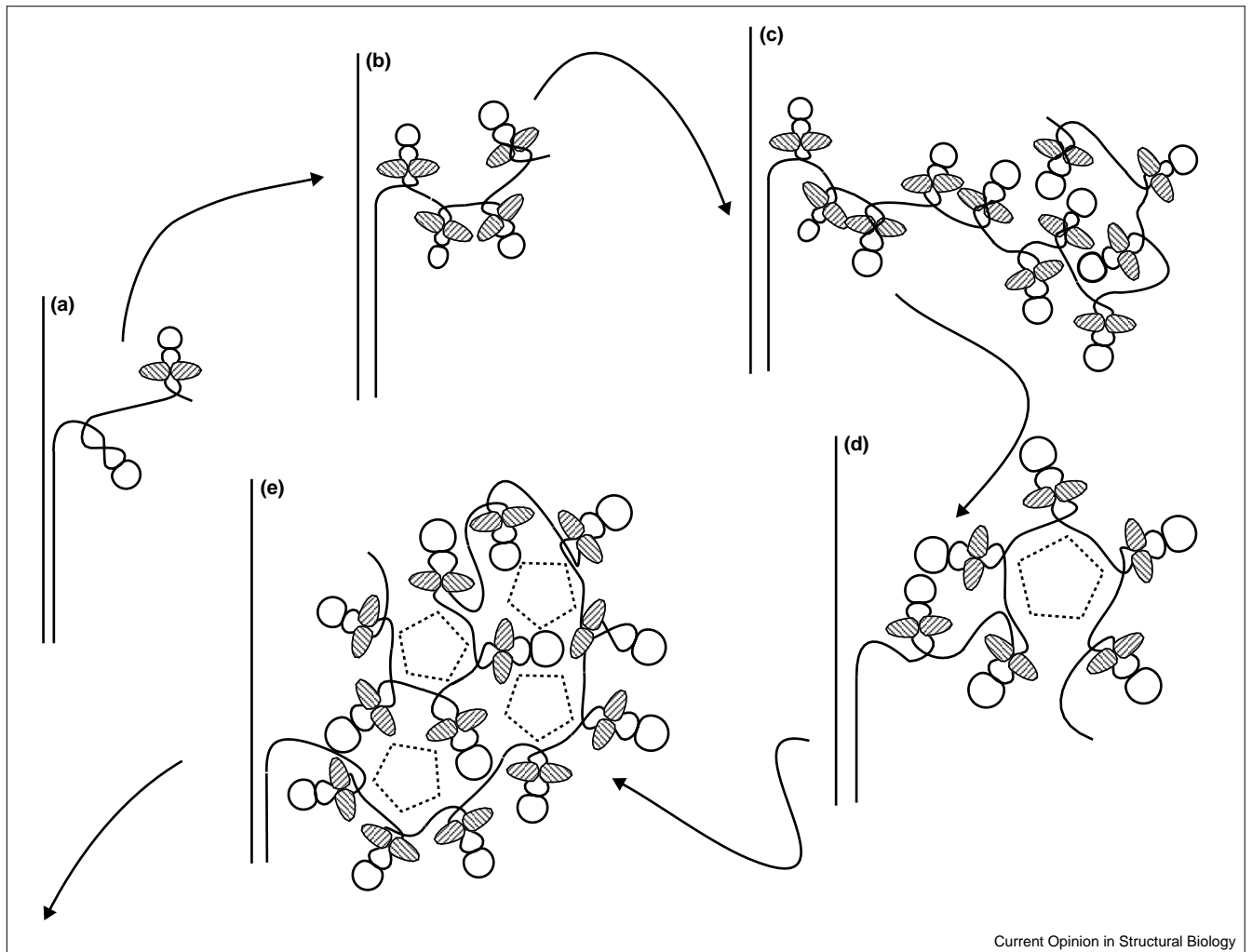
Schematic diagram illustrating a path of assembly consistent with the structural model. RNA strands emerging from the replicative complex exhibit local, otherwise transient, stem-loop structures. **(a)** Although metastable stem-loop substructures might rapidly unravel in the absence of protein, permitting restructuring to the physiological conformation, they are secured in the presence of protein; thus, the production of new double-stranded replicative form RNA is terminated. **(b)** Coat protein dimers proceed to bind contiguously along the RNA molecule. **(c)** When a sufficient number of stem-loops and coat protein subunits have been formed, the complexes coalesce into a fluid but disordered aggregate. **(d)** The complexes of RNA and coat protein sort themselves out through cooperative intramolecular and intermolecular interactions, and the pentameric nucleus of a virion forms. **(e)** Additional RNA–protein subcomplexes add to the nucleus until the entire RNA–protein mass condenses into the icosahedral virion, with protein on the outside and RNA on the inside.

systematic manner would be essential. Although there could be many variations, there is only one general arrangement, illustrated in Figure 2d. Connections would consume an additional 80 nucleotides. The course of the entire RNA chain built according to this scheme is illustrated in Figure 3a and its encapsidated form is shown in Figure 3b,c.

A model for assembly

Assembly [29–31] is ultimately constrained by the structure of the protein–nucleic acid complex or virion. The protein–protein and protein–RNA interactions known from X-ray crystallography represent the hooks and eyes that draw and fasten the virion together. It was implicit in

the RNA model that encapsidation proceeds before nucleic acid synthesis is complete, that is, as daughter RNA strands emerge from the replication complex. Thus, the RNA structural elements, helical stem-loops, would appear in a temporal and spatially linear sequence to provide, in an ordered manner, binding sites on the RNA for the capsid protein.

If the RNA inclines ultimately toward physiological conformations that cannot be encapsidated, then some agent must intervene and stabilize the sequence of stem-loop structures. The only plausible candidate is the capsid protein. STMV makes no other proteins and there is no evidence that either the host cell or the helper virus

provide assistance. RNA and protein are co-conspirators in a cooperative process. As illustrated in Figure 4, the RNA specifies, by sequential presentation of metastable substructures, the order of binding of protein dimers and their disposition along the nucleic acid. The protein subunits, through associations with one another, direct the overall organization of the RNA, guiding the formation of its tertiary structure and the shaping of an icosahedral capsid around it. Although the RNA does not, itself, specify icosahedral interactions, it is compliant in the assembly process. Protein subunits, even dimers, do not aggregate to form capsids in the absence of RNA. Only complex formation between a protein dimer and an RNA stem-loop can structurally complete a capsid unit and activate it, through the creation of appropriate new interfaces, for assembly into the virion.

Cooperativity, which is the essence of self-assembly, implies that one event progressively increases the probability of subsequent events, so that those events become inevitable. The mechanism of cooperativity in macromolecular systems is the creation of new surfaces or the exposition of new chemical groups as a consequence of some interaction. The favorable participation of these novel features in later events in turn increases the likelihood of those events recurring. Thus, assembly propagates in an ordered manner.

In STMV, there are three classes of interaction (excluding water): RNA–RNA, protein–RNA and protein–protein. In assembly, the first to occur involves only local base pairing, itself a cooperative process, which produces the stem-loop elements that ultimately program assembly. Protein–RNA interactions initiate as capsid protein dimers bind to RNA stem-loop elements. Finally, capsid protein dimers, associated with RNA stem-loops, cooperatively interact with one another and with disparate RNA elements [7,8] to form the virion.

Interactions may not appear in a rigorously specified order, but conjoin in many different patterns. It matters little, as all paths lead to the same product, the icosahedral virion. None, once the process begins, is rigorously dependent on the specific sequence of others. Such a process, which leads to the same end by a variety of paths, is inherently favored over strict adherence to a single path.

Conclusions

The central role of RNA in directing the assembly of STMV is remarkable. The RNA is a double code that executes its genetic function of specifying the amino acid sequence of the coat protein, while the formation of unique secondary and tertiary structural elements, such as the replicase recognition site and the tRNA-like structure, instructs the physiological functions of replication and translation. In addition, its sequence also codes the formation of secondary structural elements along its length, the stem-loops, that dictate the order, rate and pattern of assembly of the virion. Furthermore, the RNA provides a means by which protein subunits are structurally activated and their inherent potential to cooperatively associate and self-assemble is realized.

STMV illustrates another principle of self-assembly in biological systems. The components, RNA and protein, need not be rigorously specified or conform precisely to a fixed, periodic motif. There is slack in the system. Stems, loops and connecting strands can all be of variable sizes and lengths; the opportunities for diversity are numerous, but the variations, as long as they are not too radical, can still be accommodated. This is important in terms of natural mutation and evolution, and for explaining the occurrence of numerous strains of encapsidated RNA in the wild [19,20,22]. Just as the architecture of the virion is not 'overdetermined' in an engineering sense, neither is the assembly process 'over-determined.'

The physiologically active forms of STMV RNA represent more thermodynamically preferred conformations, whereas the linear array of helical stem-loop elements is a metastable, kinetic intermediate that arises during synthesis. In the absence of protein, the latter is destined to transform into the former. If protein is present, capsid subunits bind stem-loop substructures, both stabilizing them and directing the entire RNA into a condensed, encapsidated form. Thus, the virion does not incorporate the energetically most favored states of the individual components, but utilizes less favored, metastable intermediates of one or more of them.

Acknowledgements

References

1. Mirkov TE, Mathews DM, DuPlessis DH, Dodds JA: **Nucleotide sequence and translation of STMV-RNA**. *Virology* 1989, **170**:139-146.

2. Zucker M, Stiegler P: **Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information**. *Nucleic Acids Res* 1981, **9**:133-148.

3. Devereux J, Haeberli P, Smithies O: **A comprehensive set of sequence analysis programs for the VAX**. *Nucleic Acids Res* 1984, **12**:387-395.

4. Maramorosch K (Ed): *Viroids and Satellites: Molecular Parasites at the Frontier of Life*. Boca Raton, FL: CRC Press Inc; 1991.

5. Maramorosch K, McKelvey JJ (Eds): *Subviral Pathogens of Plants and Animals: Viroids and Prions*. Orlando, FL: Academic Press Inc; 1985.

6. Pritsch C, Mayo X: **Satellites of plant viruses**. In *Plant Viruses, vol I, Structure and Replication*. Edited by Mandahar CL. Boca Raton, FL: CRC Press; 1989:289-321.

7. Larson SB, Koszelak S, Day J, Greenwood A, Dodds JA, McPherson A: **Double helical RNA in satellite tobacco mosaic virus**. *Nature* 1993, **361**:179-182.

8. Larson SB, Koszelak S, Day J, Greenwood A, Dodds JA, McPherson A: **Three-dimensional structure of satellite tobacco mosaic virus at 2.9 Å resolution**. *J Mol Biol* 1993, **231**:375-391.

9. Larson SB, Day J, Greenwood A, McPherson A: **Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution**. *J Mol Biol* 1998, **277**:37-59.

10. Rossmann MG: **The structure of southern bean mosaic virus.** In *Biological Macromolecules and Assemblies, vol 1, Virus Structures.* Edited by Jurnak FA, McPherson A. New York: John Wiley; 1984:62-64.

11. Chen Z, Stauffacher C, Li Y, Schmidt T, Bomn W, Kamer G, Shanks M, Lomonossoff GP, Johnson JE: **Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution.** *Science* 1989, **245**:154-159.

12. Leontis NB, Santa Lucia JS (Eds): *Molecular Modeling of Nucleic Acids.* Washington DC: American Chemical Society; 1997.

13. Joshi S, Haenni A-L: **Plant RNA viruses: strategies of expression and regulation of viral genes.** *FEBS Lett* 1984, **177**:163-174.

14. Dodds JA: **Structure and function of the genome of satellite tobacco mosaic virus.** *Can J Plant Pathol* 1991, **13**:192-195.

15. Felden B, Florentz C, McPherson A, Giege R: **A histidine accepting tRNA-like fold at the 3′-end of satellite tobacco mosaic virus RNA.** *Nucleic Acids Res* 1994 **22**:2882-2885.

16. Primrose SB, Dimmock NJ (Eds): *Introduction to Modern Virology.* Oxford: Blackwell Science Publishing; 1980.

17. Witz J, Strazielle C: **Viral RNAs.** In *Subunits in Biological Systems. Part B.* Edited by Fasman GD, Timasheff SN. New York: Marcel Dekker; 1973:207-252.

18. Kaper JM: *The Chemical Basis of Virus Structure, Dissociation and Reassembly,* vol 35. Amsterdam: North Holland; 1975.

19. Fresco JR, Alberts BM, Doty P: **Some molecular details of the secondary structure of ribonucleic acid.** *Nature* 1960, **188**:98-101.

20. Crane HR: **Principles and problems of biological growth.** *The Scientific Monthly* 1950, **70**:376-389.

21. Kurath G, Rey ME, Dodds JA: **Analysis of genetic heterogeneity within the type strain of STMV reveals several variants and a strong bias for G to A substitution mutations.** *Virology* 1992, **189**:233-244.

22. Kurath G, Heick JA, Dodds JA: **RNase protection analyses show high genetic diversity among field isolates of satellite tobacco mosaic virus.** *Virology* 1993, **194**:414-418.

23. Domingo E, Holland JJ: **High error rates, population equilibrium and evolution of RNA replication systems.** In *RNA Genetics,* vol 3. Edited by Holland J, Domingo E, Ahlquist P. Boca Raton, FL: CRC Press; 1988:3-35.

24. Kurath G, Dodds JA: **Mutation analyses of molecularly cloned satellite tobacco mosaic virus during serial passage in plants: evidence for hotspots of genetic change.** *RNA* 1995, **1**:491-500.

25. Spirin AS: **Some problems concerning the macromolecular structure of ribonucleic acids.** *Prog Nucleic Acid Res* 1963, **1**:301-345.

26. Studier FW: **Conformational changes of single stranded DNA.** *J Mol Biol* 1969, **41**:189-197.

27. Tinoco I, Uhlenbeck OC, Levine MD: **Estimation of secondary structure in ribonucleic acids.** *Nature* 1971, **230**:362-367.

28. Johnson JE, Rueckert RR: **Packaging and release of the viral genome.** In *Structural Biology of Viruses.* Edited by Chiu W, Burnett RM, Garcea R. Oxford: Oxford University Press; 1997:269-287.

29. Symons RH, Haseloff J, Visvader JE, Keese P, Murphy PJ, Gill DS, Gordon KHJ, Bruening G: **On the mechanism of replication of viroids, virusoids, and satellite RNAs.** In *Subviral Pathogens of Plants and Animals: Viroids and Prions.* Edited by Maramorosch K, McKelvey JJ. Orlando, FL: Academic Press Inc; 1985:235-265.

30. Knolle P, Hohn T: **Morphogenesis of RNA Phages.** In *RNA Phages.* Edited by Zinder N. New York: Cold Spring Harbor Laboratory Press; 1975:147-201.

31. Casjens S: **Nucleic acid packaging by viruses.** In *Virus Structure and Assembly.* Edited by Casjens S. Boston: Jones and Bartlett; 1985:76-147.