Published in final edited form as:
Biochemistry. 2013 May 7; 52(18): 3182–3190. doi:10.1021/bi4001535.

Long-Range Architecture in a Viral RNA Genome

Eva J. Archer[1,3], **Mark A. Simpson**[1,3], **Nicholas J. Watts**[1,3], **Rory O'Kane**[1,3], **Bangchen Wang**[1,3], **Dorothy A. Erie**[1], **Alex McPherson**[2], and **Kevin M. Weeks**[1,*]

[1]Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290

[2]Department of Molecular Biology and Biochemistry, University of California, Irvine CA 92697

[3]UNC Undergraduate Transcriptome Project

Abstract

We have developed a model for the secondary structure of the 1058-nucleotide plus-strand RNA genome of the icosahedral satellite tobacco mosaic virus (STMV) using nucleotide-resolution SHAPE chemical probing of the viral RNA isolated from virions and within the virion, perturbation of interactions distant in the primary sequence, and atomic force microscopy. These data are consistent with long-range base pairing interactions and a three-domain genome architecture. The compact domains of the STMV RNA have dimensions of 10 to 45 nm. Each of the three domains corresponds to a specific functional component of the virus: The central domain corresponds to the coding sequence of the single (capsid) protein encoded by the virus, whereas the 5′ and 3′ untranslated domains span signals essential for translation and replication, respectively. This three-domain architecture is compatible with interactions between the capsid protein and short RNA helices previously visualized by crystallography. STMV is among the simplest of the icosahedral viruses but, nonetheless, has an RNA genome with a complex higher-order structure that likely reflects high information content and an evolutionary relationship between RNA domain structure and essential replicative functions.

Icosahedral plus-strand RNA viruses are diverse and infect organisms from all branches of life.[1] The RNA genomes in these plus-strand viruses encode information at two levels: in their primary sequences, which direct synthesis of viral proteins, and in higher-order structures, which govern RNA packaging and form complex regulatory signals.[2–4] For several icosahedral viruses, it has been possible to visualize portions of the RNA structure in crystallographic studies.[2] Global secondary structure models for the genomes of several icosahedral viruses have been proposed based on interpretation of crystallographic information, computational secondary structure prediction, and chemical probing experiments. The resulting models have emphasized both simple local stem-loop structures[2,5–10] and more complex structures that feature long-range base pairing between nucleotides distant in the primary sequence.[11–16] The well-defined partial helices seen in some crystal structures[2] and compact RNA structures visualized in imaging studies[14,17,18] point to high levels of organization, but there have been few direct evaluations of long-range base pairing and its contribution to higher-order structure in RNA viruses.

Satellite tobacco mosaic virus (STMV) represents the "hydrogen atom" for plus-strand icosahedral viruses. STMV forms a well-defined icosahedral capsid, contains an RNA genome of 1,058 nucleotides, and encodes the protein that forms 30 protein dimers that eventually coalesce to form the capsid.[6,19] As visualized crystallographically, each capsid dimer binds a short RNA helix (Fig. 1A). Averaged over all symmetry positions, each helix

[*]correspondence, weeks@unc.edu, 919-962-7486.

spans 5–7 base pairs with well-defined density and weaker density extending out to ~9 base pairs. Roughly 45% of the RNA forms base-paired elements that are ultimately visualized in the context of 60-fold crystallographic symmetry.[20,21] The terminal nucleotides in each helix are less well defined, with higher crystallographic B-factors, than the nucleotides in the centers of the helices. The connectivities between helices are not visualized by crystallography. The internal diameter of the capsid is 10 nm. The RNA occupies 75% of the available internal space in the capsid; most of the unused space appears to be in the center.[6,21]

Until recently, secondary structural modeling has not been reliable for RNAs of the lengths of the genomes of icosahedral viruses. Methods for modeling of RNA secondary structures have recently become more robust due to the improved ability to incorporate experimental information into structure prediction algorithms.[22–26] Here, we used SHAPE chemistry[27,28] to probe authentic STMV RNA genomes at single-nucleotide resolution and used this reactivity information to develop an experimentally-constrained secondary structure model. The results contrasted with current models positing simple arrayed stem-loop structures (shown schematically in Fig. 1B). Instead, our experimentally-directed structure model supported formation of extensive long-range base pairing interactions (shown schematically in Fig. 1C). We performed two tests of the proposed long-range interactions: First, we selectively disrupted putative base pairings and measured induced changes in SHAPE profiles, and, second, we directly visualized the RNA with atomic force microscopy (AFM). Data from these experiments were consistent with a complex three-domain architecture for the STMV RNA genome that contains a set of internal base-paired elements sufficient to account for the RNA helices visualized by crystallography. All experiments in this project were designed, implemented, and interpreted by undergraduate students in the Undergraduate Transcriptome Project at the University of North Carolina.

Methods

STMV Virion and RNA Purification

STMV virions were purified from leaves of infected tobacco plants[21] and dialyzed against 25% ammonium sulfate, which induces the virions to crystallize. This solution (~2 mL, at ~1.5 mg virion/mL) was dialyzed overnight against 2 L of 50 mM HEPES (pH 8.0), 200 mM NaCl, and 5 mM $MgCl_2$. These virions were frozen at −20 °C and were thawed immediately before use. For purification of *ex virio* RNA, dialyzed STMV particles were treated with proteinase K (1 mg/mL) and 1% (w/vol) sodium dodecyl sulfate for 1 h at 37 °C. Purified genomic RNA was then isolated by three extractions with phenol (equilibrated with virion dialysis buffer) and five with chloroform. Virion and viral RNA preparation protocols avoided use of denaturants, high temperature, or metal ion chelation that would disrupt STMV RNA structure.

SHAPE and RNA Structure Modeling

For *ex virio* analyses, STMV RNA was incubated in folding buffer [50 mM HEPES (pH 8.0), 200 mM potassium acetate (pH 8.0), and 3 mM $MgCl_2$] for 20 min and modified by a 5-min treatment with one-tenth volume 50 mM 1-methyl-7-nitroisatoic anhydride (1M7) in DMSO. For *in virio* experiments, one-tenth volume 20 mM 1M7 was incubated with dialyzed virions prior to capsid removal with proteinase K and SDS treatment and phenol and chloroform extractions. Locations of 2′-*O*-adducts were detected using reverse transcription with fluorescently labeled primers. cDNA products were quantified by capillary electrophoresis. Each DNA primer was 21 nucleotides in length; primers were designed to bind the RNA at positions 285-264, 514-493, 622-601, 793-772, 854-833, and 1058-1037. The 1508-1037 primer contained five locked nucleic acid (LNA) residues at positions 1058, 1055, 1052, 1049, and 1046. Data were corrected for signal decay, quantified, and normalized by the box-plot approach[22] using *ShapeFinder*.[29] Two to four replicates were obtained for each primer read, and the average of these was used to generate a continuous profile across the full-length STMV RNA. Secondary structure models were generated by inputting single-nucleotide reactivities into the *RNAstructure* program as pseudo-free energy change terms[22] using recently optimized parameters[26] of $m = 1.9$ and $b = -0.7$. No pseudoknots were detected when data for the *ex virio* and *in virio* states were analyzed using a new algorithm able to detect pseudoknots[26].

LNA-Induced Structure Perturbation

LNAs (Exiqon) were nine nucleotides long and complementary to positions 180-188 or 536-544; every position except the 3′-most contained an LNA nucleotide. LNA oligomers were designed to hybridize to single sites within the RNA based on hybridization strengths estimated using the *OligoWalk* routine in *RNAstructure*.[30] STMV RNA (1 pmol) was incubated with a 1.5-fold molar excess of LNA in folding buffer (10 μL final volume) for 20 min at 37 °C. The LNA-bound RNA and no-LNA controls were subjected to a standard SHAPE experiment, and the RNA was recovered by precipitation with ethanol. The two SHAPE experiments were processed independently. Final SHAPE traces for the no-LNA experiments were scaled to those for the plus-LNA experiments over 251-nt windows, calculated every 10 nts and summed across the full RNA sequence. An LNA-induced perturbation factor at each nucleotide was calculated as: Perturbation factor = [(plus-LNA reactivity − no-LNA reactivity) / (1 + no-LNA reactivity)]. Data were then smoothed over a window of 5 nucleotides. LNA-induced perturbations were taken to be significant at regions where multiple consecutive peaks had absolute perturbation factors ≥ 0.15.

Atomic Force Microscopy

STMV RNA was diluted to 2 nM in sterile 50 mM ammonium acetate (pH 8.0). Aliquots of 10 μl of this solution were deposited onto freshly peeled mica and dried in desiccator for 10 minutes or under a gentle stream of N_2 gas. Imaging was performed with a Nanoscope III AFM (Veeco) with 512 × 512 pixel resolution and a scan rate of 2 Hz in tapping mode. AFM tips were from Olympus (resonance 280–325.5 kHz and spring constant 34.3–54.2 N/m) or Nanosensors (resonance 146–236 kHz and spring constant 21–98 N/m). Images were flattened using Nanoscope 5.12r5 software, and volume measurements were made using ImageSXM [S.D. Barrett (2012) http://www.ImageSXM.org.uk]. Individual molecules with estimated volumes of 320 ± 100 nm^3 were selected for analysis. This range was selected to allow for uncertainties in volume measurements created by the likely compressibility of mica-immobilized RNA molecules and convolution by the dimensions of the imaging tip. Molecules were classified based on their overall structure into one of seven categories (Fig. 5). The length of each branch or major feature was measured. The most compact states that we visualized in this work were similar to the spherical states seen in prior AFM studies of STMV; however, we saw many fewer extended and fewer aggregated structures.[17,31] The differences likely reflect the fact that the RNA was not precipitated with ethanol or heated (to ≥65 °C) as was done previously. Instead, low ionic strength buffer was used to inhibit aggregation and favor formation of partially spread out structures.

RNA Dimension Calculations

Estimates for lengths of features in the STMV RNA secondary structure models assumed a rise of 2.5-Å rise per base pair and a 26-Å length span for each nested helix. Symmetrical mismatches were counted as an equivalent number of canonical base pairs. For intermediate-sized single-stranded loop regions (for example, positions 320-323 and 553-554 in Fig. 3A), the strand with the smaller number of residues was used to estimate the effective length of the segment. These length estimates apply to a (theoretical) state in which the base paired RNA is fully extended.

Results

Strategy for Examining STMV RNA Secondary Structure

We interrogated the authentic STMV RNA genome in two states, in each case starting with virions obtained from tobacco leaves (Fig. 1D). An initial dialysis step functioned to remove plant-derived polymeric material that otherwise caused viral particles and genomic RNA to aggregate. First, we analyzed the protein-free *ex virio* state, in which the genomic RNA was extracted from the virion and subjected to SHAPE. Second, we examined the STMV RNA genome structure by SHAPE inside intact virions – the *in virio* state. STMV virions and STMV RNA were maintained in buffers containing monovalent and divalent ions to stabilize native RNA structures. No denaturing or precipitation steps were used prior to structure probing. SHAPE structure probing was performed using the 1M7 reagent.[25,32] Data for >90% of STMV nucleotides were obtained using a series of six overlapping primer extension reactions (Fig. 2); products were detected by capillary electrophoresis and analyzed as outlined previously.[29,33] Data could not be collected on ~10 and ~40 nucleotides at the 5′ and 3′ ends, respectively, and a small number of internal positions.

Ex Virio Secondary Structure

SHAPE data were initially used to direct modeling of the *ex virio* state by incorporating the chemical probing data as pseudo-free energy change terms into a nearest-neighbor thermodynamic model. SHAPE-directed secondary structure modeling has proven highly accurate in experiments closely analogous to those performed here. For example, the 16S and 23S ribosomal RNAs from *E. coli*, at 1542 and 2904 nucleotides, are both longer than the STMV genome. When total RNA is gently extracted from *E. coli* cells using methods similar to those used to extract STMV RNA in this work, SHAPE-directed folding recovers >90% of the accepted base pairs in the protein-free, native-like 16S and 23S rRNAs.[22,23,26]

SHAPE-directed modeling of the STMV genome suggests that the RNA forms a three-domain structure (Fig. 3A). The central domain is a large T-shaped structure with three long helical branches. Flanking the T-shaped domain are 5′ and 3′ domains. The base pairs in the central domain are very well defined by the SHAPE data. Structures in the 5′ and 3′ domains are likely to be generally accurate, but the SHAPE data are consistent with multiple structures with local variations. In the STMV RNA structure model, 62% of all nucleotides are base paired (328 base pairs total), with most helices between five and 10 base pairs in length. Most single-stranded elements linking these helices are short; the *ex virio* structure model contains no single-stranded regions longer than 19 nucleotides (Fig. 3A). The SHAPE data were essential for defining the structure; only 40% of base pairs in the SHAPE-directed structure are identical to those calculated using a similar minimum free energy algorithm[34] but omitting the experimental data.

The SHAPE-directed *ex virio* secondary structure is quite different from models proposed for the genome structures of icosahedral viruses including STMV.[6,8,10] In general, previous models have proposed that the 30 short helices visualized by crystallography are arrayed as a series of short stem-loop elements connected by short linkers (Fig. 1B). In contrast, the SHAPE-directed model includes long-range base pairing interactions. The 30-nm arm features interactions between nucleotides 260 residues apart in primary sequence, and the base of the T-domain is formed from base pairs ~460 nucleotides distant in the sequence (Fig. 3A). It is possible to obtain a model of the STMV RNA that contains only short stem-loop structures by prohibiting base pair formation by nucleotides greater than 50 residues apart (Fig. 3B); however, this linked stem-loop structure has a net thermodynamic stability much lower than that of the three-domain model.

SHAPE-directed modeling makes use of two energy change terms: The first nearest-neighbor term, $\Delta G°_{NN}$, reflects contributions from each base pair stack, calibrated from extensive experimentation.[35,36] The second SHAPE pseudo-free energy change term, $\Delta G°_{SHAPE}$, is added at each base pair stack and provides a base pairing bonus or penalty for nucleotides with low and high SHAPE reactivity, respectively.[22,23,26] This $\Delta G°_{SHAPE}$ term provides an analytical measure of agreement of any two structures, given an experimental SHAPE dataset. The much lower and more favorable values for the three-domain model compared to those of the model constrained to form only short-range pairs (Table 1) suggest that the three-domain model is more consistent with in-solution structure probing information.

The overall structures for the two models are very different in terms of both connectivity and linear dimensions. Using standard estimates of 2.5 Å rise per base pair and a 26 Å length across each nested helix, we approximated the lengths of the major branches of the SHAPE-supported three-domain model and of the model with only short-range base pairing (Fig. 3, dimensions are emphasized with gray lines). Individual features of the three-domain model are expected to span 10–30 nm and have overall dimensions of 40–45 nm. In contrast, when fully extended, the linked stem-loop model would span ~127 nm (Fig. 3). We pursued two approaches to further distinguish between the three-domain and linked stem-loop models: measurement of reactivity changes after perturbation of specific helices by binding of locked nucleic acid (LNA) oligonucleotides and direct visualization by atomic force microscopy.

Long-Range Interactions Detected by Site-Selective LNA Binding

LNA oligonucleotides are able to bind to and locally disrupt the structure of an RNA molecule[37], and SHAPE readily detects long-range RNA structure perturbations.[33,38] We incubated the STMV RNA under our standard conditions and added a 1.5-fold excess of 9-nt long LNAs targeted to positions 180-188 or 536-544 (Fig. 4A, open boxes). These segments are base paired through long-range interactions in the three-domain model. We first confirmed site-specific binding by each LNA by its ability to inhibit reverse transcriptase-mediated primer extension. SHAPE experiments were then performed on the LNA-bound STMV RNA, LNAs were removed by adding an excess of oligonucleotide complementary to the LNA, and primer extension was used to detect the sites of SHAPE adducts in the RNA. Data were compared with that from otherwise identical experiments performed on LNA-free RNA. We calculated a structure perturbation factor, the difference in SHAPE reactivities normalized by the SHAPE reactivity of the free RNA (see Methods), to identify sites that underwent large changes in structure upon LNA binding. We focused on positive perturbation values of 3 or more standard deviations greater than the baseline changes, equal to positive perturbations of ≥0.15.

Binding by the LNA complementary to positions 180-188 induced two large disruptions to the RNA structure, one immediately adjacent to the LNA binding site (positions 173-179) and the second spanning positions 625-635 (Figs. 4A, 4B; emphasized in blue). This latter region, located ~450 nucleotides away from the LNA binding site, is the site predicted to form the other half of the long-range interaction that defines the central T-shaped domain in the three-domain model. Binding by the second LNA, complementary to positions 536-544, also caused large perturbations to the RNA structure. These RNA structure changes occurred immediately adjacent to the LNA binding site (positions 525-535 and 545-553) and at positions 356-369, consistent with the proposed structure of a long, base-paired “arm” in the three-domain STMV structure model (Fig. 4A, 4C; emphasized in green). Binding of the

536-544 LNA also induced low-level perturbations throughout the RNA (Fig. 4C), suggestive of additional long-range structural communication.

Visualization by Atomic Force Microscopy

To visualize the *ex virio* STMV RNA by AFM, two experimental procedures were used to produce extended RNA structures and to reduce RNA aggregation. First, extensive dialysis (Fig. 1D) removed plant-based contaminants. Second, the RNA was diluted into a low ionic strength, volatile [50 mM ammonium acetate (pH 8.0)] buffer to destabilize low affinity and non-specific interactions and to reduce accumulation of salt on the mica substrate. RNAs were deposited onto freshly cleaved mica, dried under nitrogen gas, and visualized in air.

Volumes of individual species measured after image processing fell in a broad peak centered at 320 nm^3 (Fig. 5A), in good agreement with the calculated volume for an RNA of this size.[39] Molecules with lower volumes were likely genome fragments, and those with higher volumes were presumably aggregates. Molecules within 100 nm^3 of the expected volume (263 molecules total) were sorted into seven categories based on visualized structural features (Fig. 5B). The most compact molecules were approximately spherical and similar in appearance to compact forms of the STMV RNA visualized by AFM previously.[17,31] The remaining molecules had less compact structures.

The multiple forms reflect different orientations of deposition and the degree to which the RNA spread out on the mica substrate. For each category of RNA (Fig. 5B), we measured the end-to-end distances of each major feature (Fig. 5C). The lengths of the most common features were approximately 10, 20, 30, 45, and 55 nm. These distances corresponded closely with the estimates of feature dimensions in the predicted *ex virio* structure (distances shown in Figs. 3 and 5C). Very few molecules (~6%) had features with lengths greater than 60 nm. None of the forms observed had the 127-nm length expected of a fully extended structure with a series of short stem loops connected by single-stranded regions. In sum, as visualized by AFM, the STMV RNA populates a family of conformations whose features are consistent with a compact overall structure.

In Virio STMV RNA Genome Structure

Finally, we examined the structure of the unperturbed STMV genome RNA inside virions. We performed SHAPE on intact virion particles by inverting the modification and extraction steps in our experimental approach (Fig. 1D). The overall pattern of SHAPE reactivity for RNA in the intact viral particle was similar to that for the protein-free RNA (Fig. 2, *in virio* panel), and the resulting SHAPE-directed structure predicted for the *in virio* STMV RNA was similar to that of the *ex virio* structure. Most long-range pairings, the three-domain architecture, and an overall T-shape for the central domain were present in both models (compare Fig. 3 and 6). Intriguingly, in regions where the SHAPE reactivities did differ, the *in virio* RNA was generally more reactive towards the 1M7 reagent than was the *ex virio* RNA (Fig. 2, difference panel); this suggests that capsid binding weakens or strains some intramolecular RNA interactions or that removal of the coat protein allows the RNA to form additional structure.

The largest differences between *in virio* and *ex virio* RNA occurred at positions 460-465, 490-510, and 579-590 (Fig. 6, labels *a*, *b*, *c*). In each case, the *in virio* RNA was highly reactive, whereas the *ex virio* was unreactive. Differences at the latter two sites resulted in differences in the secondary structure models: At these sites, the RNA is predicted to form base pairs *ex virio* but is predicted to be single stranded *in virio*. In the right-hand branch of the central domain, the *in virio* structure has additional stem loops and more local pairings than the *ex virio* structure. Both the 5′ and 3′ domains are also less structured, on average, *in virio* than *ex virio*. In the 5′ domain, four stem-loop elements occur in both models but with different inter-helix connectivities. Of the nine stem-loop elements in the 3′ domain in the *in virio* structure, eight are present in the *ex virio* model. In general, for both *ex virio* and *in virio* structures, the 5′ and 3′ domains are less well defined than is the central T-shaped domain. This difference may reflect sampling of different low free-energy structures.

Discussion

Structural motifs within icosahedral viral genomic RNAs play important roles in multiple distinct stages of viral replication.[3,4,40,41] Most current working models for retrovirus,[42] rhinovirus,[7] STMV,[6,8,10] and satellite tobacco necrosis virus (STNV)[9] RNA genomes emphasize short single-stranded and stem-loop motifs without long-range base pairing. Recent exploratory studies using chemical probing experiments lead to two very different models for the structure of STMV RNA.[8,16] Here, we report multiple lines of evidence that support a model in which the STMV RNA genome folds into three structural domains stabilized by long-range base pairing (shown schematically in Fig. 1C). These structures are likely to influence genome compaction and packaging.

In the SHAPE-directed RNA secondary structure model, approximately 60% of the nucleotides are involved in base pairs (Fig. 3A), consistent with the high level of structure visualized crystallographically.[20,21] The SHAPE structure probing data are more consistent with the three-domain model than models with only short-range stem-loops (Fig. 3, Table 1). We did not find support for the pseudoknots previously proposed to form at the 3′ end of the STMV genome,[44,45] suggesting that a non-pseudoknotted structure predominates in the STMV RNA isolated from virions. The structure of the central T-shaped domain is essentially identical to that suggested by recent SHAPE experiments that probed an *in vitro* transcript corresponding to the STMV RNA.[16] This agreement emphasizes the robustness of SHAPE-directed secondary structure modeling. The median correlation between the *ex virio* SHAPE data (Fig. 2) and the data obtained using *in vitro* transcripts was high in most regions ($R \geq 0.8$). In regions of lower correlation, the *in vitro* transcript was generally more highly structured (lower SHAPE reactivities) than the authentic *ex virio* STMV RNA, which was, in turn, more highly structured than the RNA inside virions (Fig. 2). This trend suggests that virion assembly and protein-RNA interactions destabilize some RNA structures.

The long-range domain organization we propose is consistent both with detection of interactions between nucleotides far apart in the primary sequence as measured by LNA-induced structure disruption (Fig. 4) and with AFM imaging at low ionic strength (Fig. 5). The dimensions estimated based on calculation of helix lengths and measured after deposition on a flat surface by AFM are consistent with formation of compact domains (Figs. 3A and 5C). AFM imaging shown here (Fig. 5B) and performed previously[17,31] shows that the RNA maintains higher-order structure even after the capsid is removed. A reversible structural transition occurs when the STMV RNA is heated from 4 to 65 °C with a discontinuity around 55 °C.[31] This transition may correspond to a rearrangement of the three-domain STMV architecture. The domains in our proposed model (Fig. 3 & Fig. 6, emphasized in color) are consistent with the branched structures visualized by AFM, and the most compact structures observed by AFM would fit into and largely fill the 10-nm internal diameter of the STMV capsid.

Crystallographic experiments with STMV indicate that short RNA helices are bound at the dimer interfaces of the STMV viral capsid protein.[20,21] Given the icosahedral symmetry of the virus, this suggested that the RNA likely forms roughly 30 helices that occupy similar sites in the virion (Fig. 1A). The helices predicted by SHAPE-directed modeling of both the *ex virio* and *in virio* RNAs can be arranged to match the 30 edges of the capsid (Fig. 6, see numbered helices). The stem-loop structures in the 5′ and 3′ domains are connected by single-stranded RNA elements and can be placed at the helix interaction sites formed by the capsid dimer, similar to prior models.[6,10] Unique to our model, helices in the extended structure of the central T-domain are 6 to 11 base pairs in length and connected by bulges and loops whose constituent nucleotides are reactive by SHAPE. If the sequence of helices traces a short path through the binding sites, then the angles between the helices would be acute (roughly 60 degrees; see Fig. 1A and Fig. 6, *inset*). Compilations of helix junctions from the crystallographic databases show clearly that internal loops of approximately 2 nucleotides in each strand are sufficient to allow for sharp, ~60° bends in an RNA.[46] The STMV secondary structure model accommodates the bulge structures required for the tight bends needed to thread short helices through the binding sites on the capsid (Fig. 6, *inset*). Other specific arrangements of extended helices can be fit within the geometric constraints of the capsid. Critically, the proposed STMV RNA genome model – with extensive long-range base pairing interactions – is fully compatible with the icosahedral virus structure of STMV.

A final intriguing observation is that the three-domain architecture of the viral genome corresponds closely with the evolutionary origins of each functional component of the virus. The central T-shaped domain almost exactly spans the coding sequence of the capsid protein (Figs. 3 and 6, start and stop codons are boxed). STMV is a satellite virus that requires the tobacco mosaic virus (TMV) for replication. STMV is thought to have arisen in plant cells co-infected by multiple viruses.[19] During mixed replication, a 3′-UTR replicase recognition element of ~350 nucleotides (from the TMV helper virus) became genetically linked to an expressed icosahedron-forming coat protein. The 5′-UTR, which includes signals for translation initiation, was likely derived from a second virus. Strikingly, each functional region of the virus – the 5′ and 3′ untranslated regions important for translation and genome replication, respectively, and the capsid open reading frame – is encoded within its own modular domain (Fig. 6). The sequestration of regulatory and coding sequences into distinct domains may reflect an evolutionary and structural relationship between RNA and protein structure and may prove to be an organizational feature of many viral RNAs. Although STMV is among the simplest of the icosahedral viruses, this work indicates that the RNA genome forms a complex and stable higher-order structure with high information content encoded in all levels of its organization, principles likely to apply broadly to viral RNA genomes.

Acknowledgments

This work, including the UNC Undergraduate Transcriptome Project, was supported by grants from the NSF (MCB-0919666 and MCB-1121024 to K.M.W.), by a UNC Summer Undergraduate Research Fellowship (SURF) to E.J.A, and by grants from the NIH (GM079480 to D.A.E., and GM080412 to A.M.). We are indebted to many members of the Weeks and Erie laboratories for mentorship and discussions.

References

1. Koonin EV, Dolja VV. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences, Crit. Rev Biochem Mol Biol. 1993; 28:375–430.

2. Schneemann A. The structural and functional role of RNA in icosahedral virus assembly, Annu. Rev Microbiol. 2006; 60:51–67.

3. Rao AL. Genome packaging by spherical plant RNA viruses, Annu. Rev Phytopathol. 2006; 44:61–87.

4. Simon AE, Gehrke L. RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. Biochim Biophys Acta. 2009; 1789:571–583. [PubMed: 19501200]

5. Hellendoorn K, Mat AW, Gultyaev AP, Pleij CW. Secondary structure model of the coat protein gene of turnip yellow mosaic virus RNA: long, C-rich, single-stranded regions. Virology. 1996; 224:43–54. [PubMed: 8862398]
6. Larson SB, McPherson A. Satellite tobacco mosaic virus RNA: structure and implications for assembly, Curr. Opin Struct Biol. 2001; 11:59–65.
7. Palmenberg AC, Spiro D, Kuzmickas R, Wang S, Djikeng A, Rathe JA, Fraser-Liggett CM, Liggett SB. Sequencing and analyses of all known human rhinovirus genomes reveal structure and evolution. Science. 2009; 324:55–59. [PubMed: 19213880]
8. 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. Biophysical J. 2011; 101:167–175.
9. Bunka DH, Lane SW, Lane CL, Dykeman EC, Ford RJ, Barker AM, Twarock R, Phillips SE, Stockley PG. Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: implications for the assembly of a T=1 capsid. J Mol Biol. 2011; 413:51–65. [PubMed: 21839093]
10. Zeng Y, Larson SB, Heitsch CE, McPherson A, Harvey SC. A model for the structure of satellite tobacco mosaic virus. J Struct Biol. 2012:110–116. [PubMed: 22750417]
11. Rodriguez-Alvarado G, Roossinck MJ. Structural analysis of a necrogenic strain of cucumber mosaic cucumovirus satellite RNA in planta. Virology. 1997; 236:155–166. [PubMed: 9299628]
12. Simmonds P, Tuplin A, Evans DJ. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. RNA. 2004; 10:1337–1351. [PubMed: 15273323]
13. Badorrek CS, Weeks KM. Architecture of a gamma retroviral genomic RNA dimer. Biochemistry. 2006; 45:12664–12672. [PubMed: 17042483]
14. Davis M, Sagan SM, Pezacki JP, Evans DJ, Simmonds P. Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses. J Virol. 2008; 82:11824–11836. [PubMed: 18799591]
15. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009; 460:711–716. [PubMed: 19661910]
16. Athavale SS, Gossett JJ, Bowman JC, Hud NV, Williams LD, Harvey SC. In vitro secondary structure of the genomic RNA of satellite tobacco mosaic virus. PLoS One. 2013; 8:e54384. [PubMed: 23349871]
17. Kuznetsov YG, Daijogo S, Zhou J, Semler BL, McPherson A. Atomic force microscopy analysis of icosahedral virus RNA. J Mol Biol. 2005; 347:41–52. [PubMed: 15733916]
18. Sagan SM, Nasheri N, Luebbert C, Pezacki JP. The efficacy of siRNAs against hepatitis C virus is strongly influenced by structure and target site accessibility, Chem. Biol. 2010; 17:515–527.
19. Dodds JA. Satellite tobacco mosaic virus, Annu. Rev Phytopathol. 1998; 36:295–310.
20. 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. [PubMed: 8421525]
21. 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. [PubMed: 9514737]
22. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination, Proc. Natl Acad Sci USA. 2009; 106:97–102.
23. Low JT, Weeks KM. SHAPE-directed RNA secondary structure prediction. Methods. 2010; 52:150–158. [PubMed: 20554050]
24. Gherghe C, Lombo T, Leonard CW, Datta SA, Bess JW, Gorelick RJ, Rein A, Weeks KM. Definition of a high-affinity Gag recognition structure mediating packaging of a retroviral RNA genome, Proc. Natl Acad Sci USA. 2010; 107:19248–19253.
25. Leonard CW, Hajdin CE, Karabiber F, Mathews DH, Favorov OV, Dokholyan NV, Weeks KM. Principles for understanding the accuracy of SHAPE-directed RNA structure modeling. Biochemistry. 2013; 52:588–595. [PubMed: 23316814]

26. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots, Proc. Natl Acad Sci USA. 2013; 110:5498–5503.

27. Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc. 2005; 127:4223–4231. [PubMed: 15783204]

28. Weeks KM, Mauger DM. Exploring RNA structural codes with SHAPE chemistry, Acc. Chem Res. 2011; 44:1280–1291.

29. Vasa SM, Guex N, Wilkinson KA, Weeks KM, Giddings MC. ShapeFinder: a software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. RNA. 2008; 14:1979–1990. [PubMed: 18772246]

30. Lu ZJ, Mathews DH. OligoWalk: an online siRNA design tool utilizing hybridization thermodynamics. Nucleic Acids Res. 2008; 36:W104–108. [PubMed: 18490376]

31. Kuznetsov YG, Dowell JJ, Gavira JA, Ng JD, McPherson A. Biophysical and atomic force microscopy characterization of the RNA from satellite tobacco mosaic virus, Nucl. Acids Res. 2010; 38:8284–8294.

32. Mortimer SA, Weeks KM. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. J Am Chem Soc. 2007; 129:4144–4145. [PubMed: 17367143]

33. Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, Giddings MC, Weeks KM. High-Throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. PLoS Biology. 2008; 6:e96. [PubMed: 18447581]

34. Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics. 2010; 11:129. [PubMed: 20230624]

35. Mathews DH, Turner DH. Prediction of RNA secondary structure by free energy minimization, Curr. Opin Struct Biol. 2006; 16:270–278.

36. Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res. 2010; 38:D280–282. [PubMed: 19880381]

37. Kauppinen S, Vester B, Wengel J. Locked nucleic acid: high-affinity targeting of complementary RNA for RNomics. Handb of Exp Pharmacol. 2006:405–422.

38. Duncan CDS, Weeks KM. SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA. Biochemistry. 2008; 47:8504–8513. [PubMed: 18642882]

39. Voss NR, Gerstein M. Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. J Mol Biol. 2005; 346:477–492. [PubMed: 15670598]

40. Dreher TW, Miller WA. Translational control in positive strand RNA plant viruses. Virology. 2006; 344:185–197. [PubMed: 16364749]

41. Liu Y, Wimmer E, Paul AV. Cis-acting RNA elements in human and animal plus-strand RNA viruses. Biochim Biophys Acta. 2009; 1789:495–517. [PubMed: 19781674]

42. D'Souza V, Summers MF. How retroviruses select their genomes, Nature Rev. Microbiol. 2005; 3:643–655.

43. Gherghe C, Leonard CW, Gorelick RJ, Weeks KM. Secondary structure of the mature ex virio Moloney murine leukemia virus genomic RNA dimerization domain. J Virol. 2010; 84:898–906. [PubMed: 19889760]

44. Gultyaev AP, van Batenburg E, Pleij CW. Similarities between the secondary structure of satellite tobacco mosaic virus and tobamovirus RNAs. J Gen Virol. 1994; 75 (Pt 10):2851–2856. [PubMed: 7931178]

45. 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; 22:2882–2886. [PubMed: 8065897]

46. Bindewald E, Hayes R, Yingling YG, Kasprzak W, Shapiro BA. RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. Nucleic Acids Res. 2008; 36:D392–397. [PubMed: 17947325]

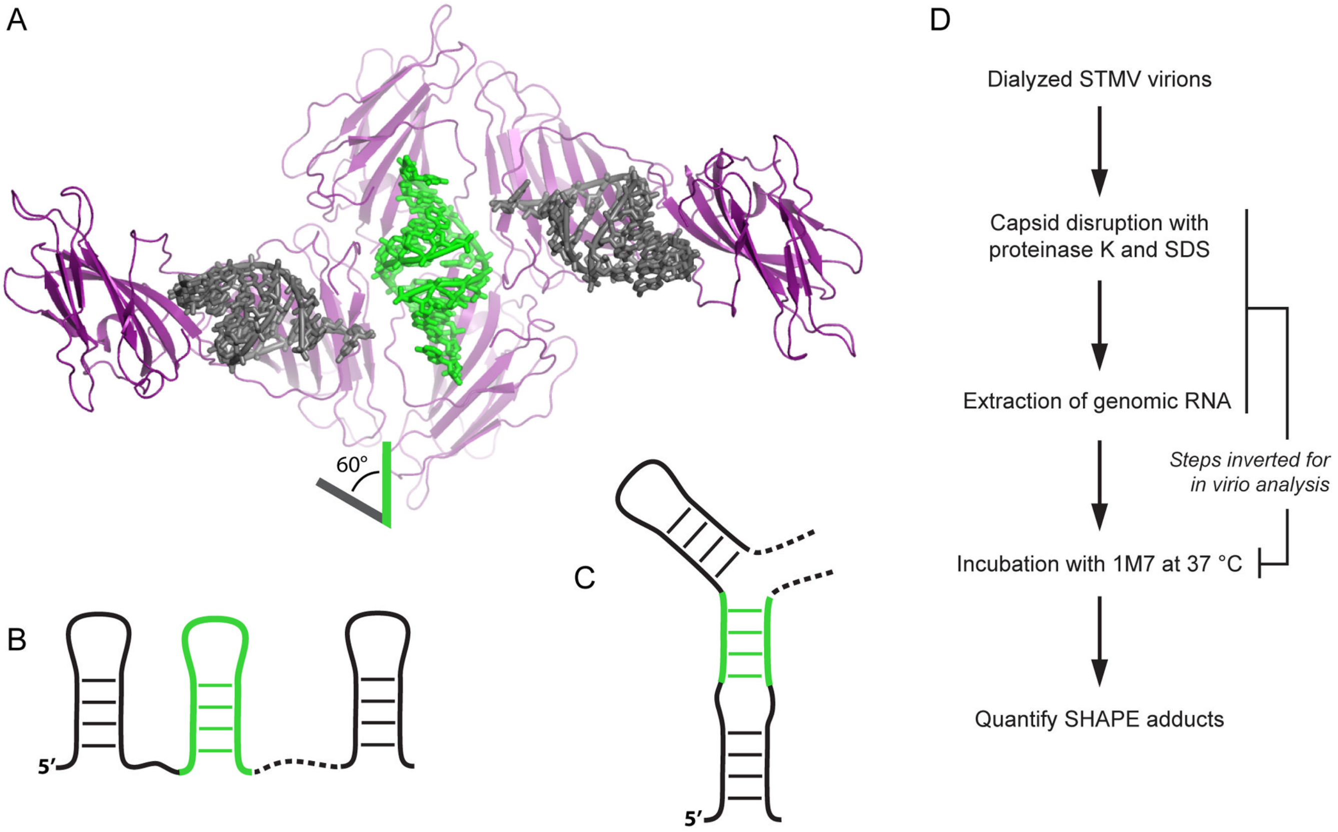
A
60°
B
5′
C
5′
D
Dialyzed STMV virions
Capsid disruption with proteinase K and SDS
Extraction of genomic RNA
Steps inverted for in virio analysis
Incubation with 1M7 at 37 °C
Quantify SHAPE adducts

Figure 1.

Models for the organization of RNA helices in the STMV capsid and analysis of native RNA genome structure by SHAPE. (A) Crystallographic model of helices and capsid interactions in STMV (1a34).[21] Three capsid dimers and their respective bound RNA helices are shown. (B, C) Schematic interpretations of the helices visualized by crystallography in terms of (B) linked stem-loop and (C) long-range base pairing models. (D) Approach for analyzing STMV genome structure under native-like conditions by SHAPE.

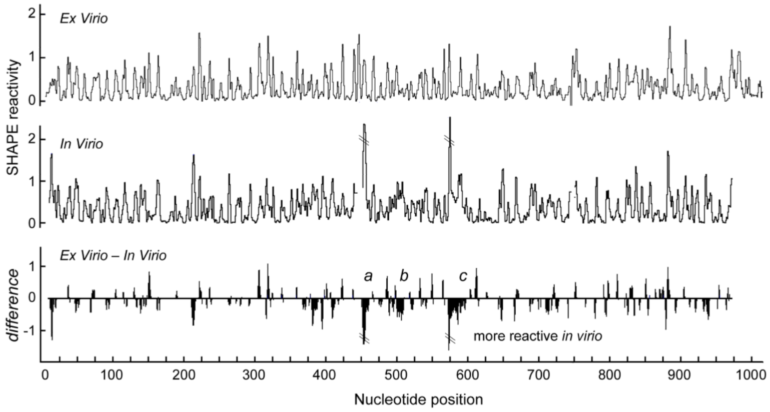

Figure 2.

SHAPE reactivity profiles for *ex virio* (top) and *in virio* (middle) STMV RNAs. Difference plot of SHAPE reactivities (bottom). Positive and negative values indicate protection from versus enhanced SHAPE reactivity *in virio*.

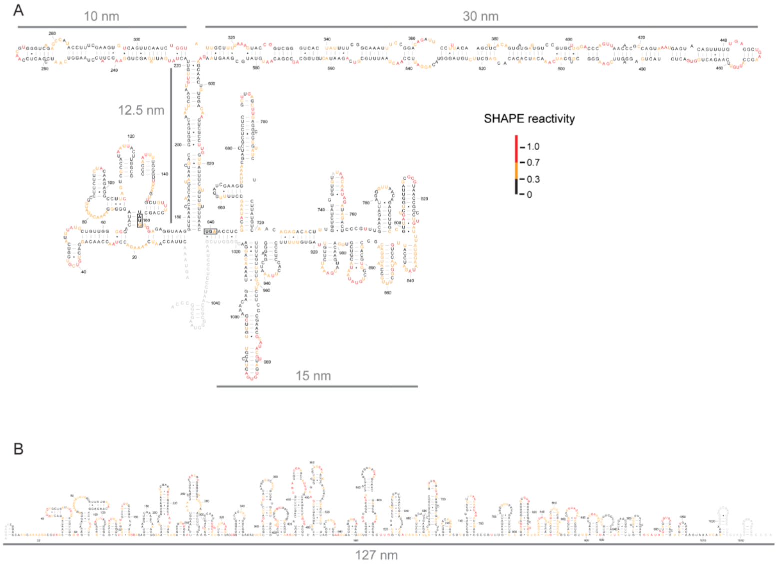

Figure 3.

Secondary structure models for the STMV RNA *ex virio*. (A) SHAPE-directed model. Maximum allowed base pairing distance was 600 nucleotides.[22] The start and stop codons for the capsid protein are boxed. (B) Linked stem-loop model, created using SHAPE data and parameters designed to force formation of short stem-loop motifs by restricting the maximum base pairing distance to ≤50 nucleotides. Nucleotides are colored by SHAPE reactivity (see legend); gray indicates no data were obtained. Calculated lengths of major structural features in each structure are shown (in nanometers).

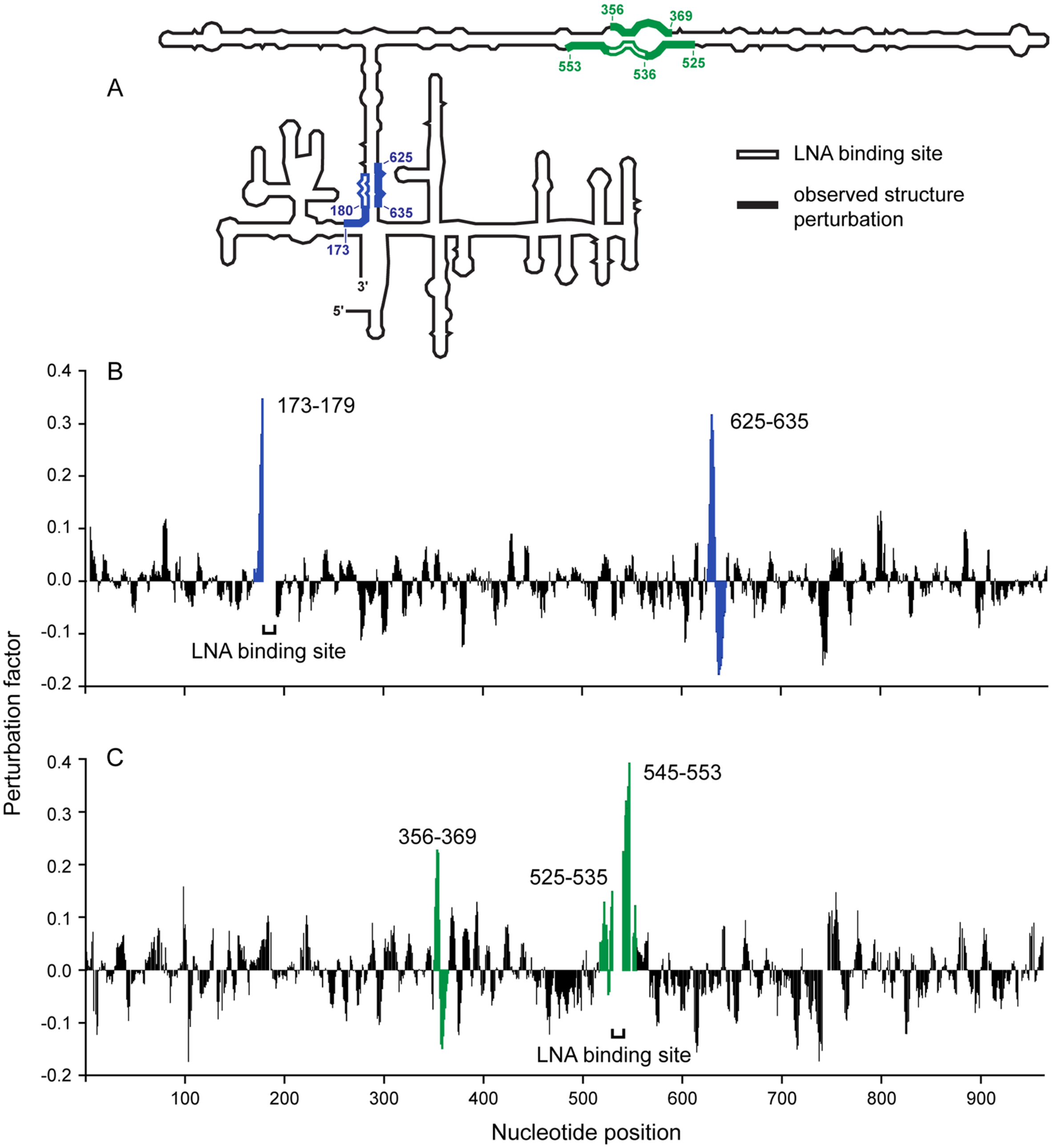

Figure 4.

Analysis of long-range interactions in the STMV RNA genome by LNA-mediated structure disruption. (A) Schematic image showing sites of LNA binding (open boxes) and the resulting SHAPE-detected perturbation (heavy lines). Quantification of LNA-induced structure perturbations for (B) an LNA bound at positions 180-188 and (C) an LNA bound at positions 536-544. The largest observed changes for each LNA (≥ 0.15) are indicated by color.

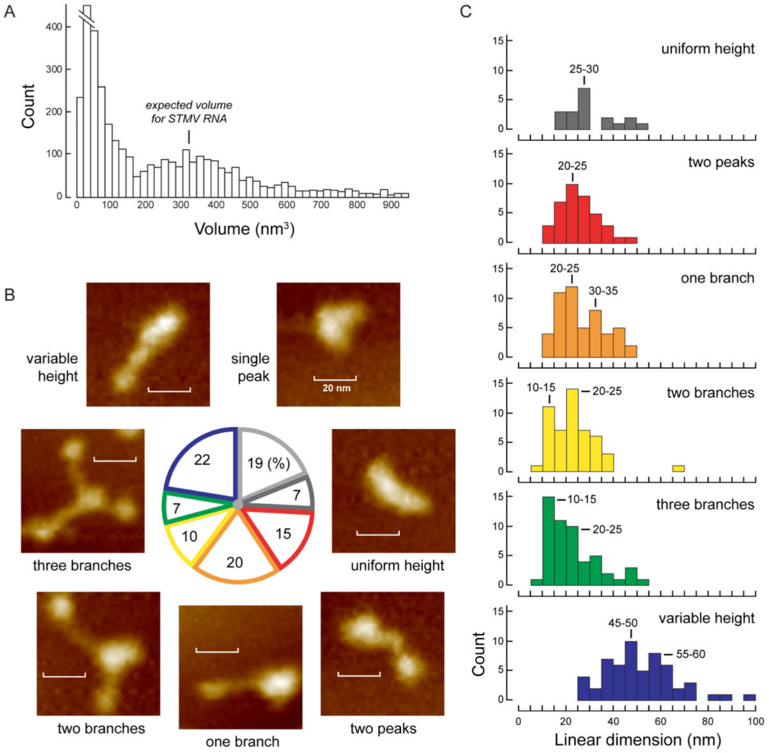

Figure 5.

AFM visualization of STMV RNA structure. (A) Volume distribution for all species in AFM images. The peak at 320 nm^3 is consistent with the calculated molecular volume of a 1058-nt RNA. (B) Classification of single RNA molecules by structural features. Central chart shows the fraction of RNAs in each category. Observed species suggest a general unfolding from most condensed to extended conformations of three branches. (C) Lengths of observed features based on branch length and peak-to-peak distance. Feature lengths corresponding to peaks in each histogram are labeled explicitly. A single length was measured for the uniform height, two peaks, one branch, and variable height molecules; two and three lengths, respectively, were measured for the two and three branches molecules. No molecules had a length greater than 100 nm.

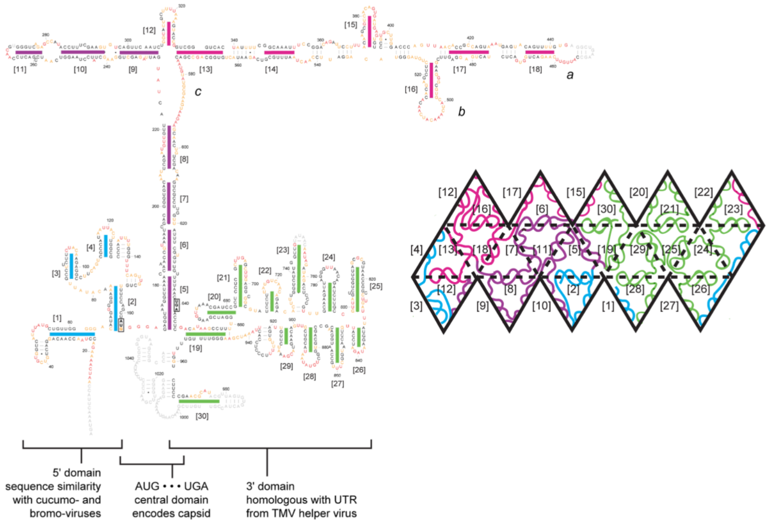

Figure 6.

Model for the secondary structure and domain architecture of the STMV RNA genome derived from SHAPE measurements performed inside intact virions. Thirty representative helices (numbered in brackets) that constitute plausible sites of interaction with capsid protein dimers[21] are emphasized with colored lines. Three regions with the largest differences relative to the *ex virio* structure are identified as *a*, *b*, and *c*. The start and stop codons for the capsid protein open reading frame are boxed. Nucleotides are colored by SHAPE reactivity using the scheme shown in Fig. 3. (*inset*) Schematic map of STMV RNA helices and connectivity superimposed onto icosahedral geometry. Each edge corresponds to an RNA-capsid interaction site.

Table 1

Nearest-neighbor ($\Delta G°_{NN}$) and SHAPE pseudo-free energy ($\Delta G°_{SHAPE}$) changes for the three-domain and linked stem-loop models for the STMV RNA genome.

Model	base-pair cutoff	$\Delta G°_{NN}$	$\Delta G°_{SHAPE}$	$\Delta G°_{Total}$
Three-domain	600	−296	−366	−662
Linked stem-loop	50	−253	−293	−546
difference		43	73	116

Free energy changes in kcal/mol; lower (more negative) energies are more favorable or are more consistent with the experimental SHAPE probing information. $\Delta G°_{NN}$ is the standard Turner nearest-neighbor free energy change[36] and provides an estimate for the relative stability of each structure. $\Delta G°_{SHAPE}$ is not a physical energy but represents a measure of the relative agreement of each structure with the experimental in-solution SHAPE probing information. $\Delta G°_{Total}$ is the sum of $\Delta G°_{NN}$ and $\Delta G°_{SHAPE}$.