Rewriting nature’s assembly manual for a ssRNA virus

Nikesh Patel, Emma Wroblewski, German Leonov, Simon E. V. Phillips, Roman Tuma, Reidun Twarock, and Peter G. Stockley

"Satellite tobacco necrosis virus (STNV) is one of the smallest viruses known. Its genome encodes only its coat protein (CP) subunit, relying on the polymerase of its helper virus TNV for replication. The genome has been shown to contain a cryptic set of dispersed assembly signals in the form of stem-loops that each present a minimal CP-binding motif AXXA in the loops. The genomic fragment encompassing nucleotides 1–127 is predicted to contain five such packaging signals (PSs). We have used mutagenesis to determine the critical assembly features in this region. These include the CP-binding motif, the relative placement of PS stem-loops, their number, and their folding propensity. CP binding has an electrostatic contribution, but assembly nucleation is dominated by the recognition of the folded PSs in the RNA fragment. Mutation to remove all AXXA motifs in PSs throughout the genome yields an RNA that is unable to assemble efficiently. In contrast, when a synthetic 127-nt fragment encompassing improved PSs is swapped onto the RNA genome, a CP-binding sequence with subsequences that are continuous nucleotides, e.g., 5’-AXXAXXXAXXAX-3’, delivers an improved assembly substrate. These data confirm details of the PS-mediated assembly mechanism for STNV and identify an efficient approach for production of stable virus-like particles encapsidating nonnative RNAs or other cargoes.

Significance

Viruses composed of a shell of coat proteins enclosing ssRNA genomes are among the simplest biological entities. Their lifecycles include a range of processes, such as specific genome encapsidation and efficient capsid self-assembly. Until recently, these were not linked, but we have shown that many viruses in this class encode multiple, degenerate RNA sequence/structure motifs that bind cognate coat proteins collectively. This simultaneously ensures specific genome packaging and efficient virion assembly via an RNA-encoded instruction manual. Here we extract essential features of this manual in a viral RNA genome, creating a synthetic sequence with an assembly substrate superior to the natural equivalent. Such synthetic viral RNA assembly substrates should lead to dramatic improvements in VLP technology.

Author contributions: N.P. and P.G.S. designed research; N.P., E.W., and G.L. performed research; G.L., R. Tuma, and R. Twarock contributed new reagents/analytic tools; N.P., G.L., R. Twarock, and P.G.S. analyzed data; and N.P., E.W., S.E.V.P., R. Tuma, R. Twarock, and P.G.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence should be addressed. Email: p.g.stockley@leeds.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706951114/-/DCSupplemental.
AXXA loop acts as a CP recognition motif and that all five of the PS sites act cooperatively to condense the RNA fragment and assemble a complete T = 1 CP shell around it (12) (Fig. 1C). The loop of the PS3 site is the most important for generating these effects, which require multiple PSs, consistent with the relative spacing to neighboring PSs being vital for accurate control of assembly. STNV is therefore an ideal test case for the design of a synthetic assembly substrate. Here, we analyze in depth the contribution of different molecular features to the cooperative assembly, arriving at an improved, synthetic RNA that surpasses the native sequence in assembly efficiency, paving the way for improved VLP production.

Results

Sequence-Specific Recognition of Individual PS Sites. There are multiple consequences of sequence-specific RNA-CP recognition in the STNV system (Fig. 1). Titration of CP into oligonucleotides encompassing only PS3 (or B3) initially results in formation of a trimeric capsomer (R_h ~ 5 nm), followed by formation of T = 1 VLPs (R_h ~ 11.3 nm) as the CP concentration is raised gradually. R_h distribution plots of the smFCS data at the end of the titration suggest that the VLPs formed are homogeneous, while EM images and RNase challenge assays suggest that they are composed of complete protein shells. A similar titration with a PS3/B3 variant having a loop sequence of UUUU showed that CP binds such SLs, but the complex formed is unable to assemble to VLPs (12). The natural 127-mer, encompassing PS1–PS5, shows more complex behavior. Addition of low CP concentrations triggers a collapse in its R_h by about 20–30%, mimicking the behavior seen for the full-length genome (11). Subsequent CP additions result in cooperative conversion to T = 1 VLPs with the same properties as those formed around PS3 alone. PS variants within this fragment confirm that AXXA is a CP recognition motif. Its presence is only absolutely required in PS3. However, the variants no longer assemble with wild-type cooperativity (12). STNV-1 CP alone does not aggregate below 15 μM under these conditions, and therefore everything in the titrations shown here is a consequence of RNA-CP interaction.

To identify the critical features of PS3 recognition, we produced a series of SLs encompassing variant loop sequences but retaining the PS3 stem sequence (SI Appendix, Fig. S1 and Table S1). The variants have altered nucleotides in the “inner” two positions (CC, AA, GG, GU, and UG) compared with the wild-type CA of PS3. “Outer” variants (AUUA, AUUG, GUUA, GUUG, GUUU, UUUG, UUUA, and AUUU), in which both inner nucleotides were altered to uridines, were also tested. Our expectation was that there would be no base specificity at the middle positions, while the adenines would be preferred at the first and last positions of the tetraloop. We examined their abilities to support assembly of both the T = 1 shell and the trimeric capsomer. Capsid reassembly assays (SI Appendix) were carried out at a molar ratio of 1:3 RNA:CP with a final CP concentration of 6 μM. Note, reassembly remains sensitive to the loop sequence at these concentrations (17). The results were assayed by velocity sedimentation analysis and in EM images, and yields were quantitated by quasi-elastic light scattering (QELS) chromatography. The inner nucleotide variants form T = 1 capsids with roughly similar efficiency as PS3, confirming that their identities are not part of the CP recognition motif (Fig. 2 A and B and SI Appendix, Fig. S1). The outer nucleotide variants showed differing behavior, with only the AUUU, UUUA, and AUUA variants having a peak in a similar position to PS3, confirming that the outer adenines are part of the CP recognition motif.

To examine the relative importance of the loop sequence for CP affinity, we adapted the smFCS assay (Fig. 1C). Labeled B3 was titrated with CP to form the trimer, as judged by the R_h value, and then a 100-fold molar excess of each outer sequence variant was added to see if this would displace the bound dye-label. Variants that do not bind with a similar affinity to B3 fail to displace the labeled RNA, whereas B3 and other variants displace the labeled species restoring the R_h to that of CP-free RNA (Fig. 2C). The results (Fig. 2D) show the percentage R_h change following this challenge, revealing a wide variation between

~1.2-kb positive-sense ssRNA genome only encodes its CP gene. Previously, we identified the highest-affinity STNV PS using RNA SELEX against the cognate CP (31). One aptamer, B3 (Fig. 1B), forms an SL with an ACAA sequence in the loop. This matches a previously identified PS, B3 (Fig. 1). Single-molecule fluorescence correlation spectroscopy (smFCS) reassembly assays can be carried out at low nanomolar concentrations. These conditions reveal the effects of PS-mediated assembly that are lost at higher concentration (13, 17). We showed previously by this method that the

Fig. 1. The STNV system. (A) Ribbon diagram of the STNV T = 1 capsid (green) (Left, PDB 354G) viewed along a fivefold axis with a trimeric capsomer highlighted (magenta/pink) and (Right) a CP monomer (magenta, PDB 4BCU). Side-chains mutated here are shown and labeled. The disordered N-terminal amino acid sequence is shown as a dashed line, next to the sequence of the first 25 amino acids. (B) Sequence and putative secondary structure of the 127-nt 5’ STNV-1 genomic fragment showing the locations of the PS SLs, named 5’ to 3’ as PS1 to PS5, respectively. Each contains the CP recognition motif, AXXA, in their loops (white circles, black outline). The B3 aptamer is shown similarly above. Nucleotides are color-coded as indicated, here and throughout. (C) Example smFCS assays. Hydrodynamic radius (R_h) values for CP-free, fluorescently labeled RNAs (black line for PS1–PS5 and red line for B3) are determined before and during STNV CP titration at fixed time points (vertical dashed lines). The R_h values were allowed to equilibrate after each step. The PS1–PS5 R_h initially collapses by up to 30% until the CP concentration reaches a threshold, triggering cooperative assembly to T = 1 VLPs (R_h ~ 11 nm). At the end of each titration, the complexes formed are challenged by addition of RNase A. Largely unchanged R_h values were assumed to indicate that the RNA is in a closed VLP.
loop sequence variants. All those with guanine substitutions and the AUUU variant fail to displace B3. The superior performance of the UUUU variant suggests that the 3’ A is the most important for CP recognition. Alternatively, the A-U base pair at the top of the adjacent stem may break and present an AUUU variant of the B3 motif that is still recognized by the CP. Either way, AXAX outperforms all variants, suggesting that SLs carrying tetratrop motifs of AXAX encompass the best CP recognition motif for assembly into VLPs.

Roles of Electrostatics and PS Cooperativity in VLP Assembly. The results above demonstrate that sequence specificity of RNA-CP interaction is the major determinant of assembly. Previous work with other plant viruses having similarly positively charged N-terminal tails has, in contrast, suggested that the major assembly driving force is electrostatic neutralization. STNV-1 CP is typical of many viruses having many basic amino acids in its N-terminal tail giving it a net +8 charge, including the N terminus; cf. cowpea chlorotic mosaic virus (3), which has a charge of +10. Three of these sidechains, R8, R14, and K17 (Fig. 1), are close to the RNA duplex seen in the crystal structure of the B3 VLP (17). To examine the effects of these positive charges on assembly we produced mutations with A or D in place of K or R. Since R14 and K17 are adjacent in three dimensions, their variants were made as the double mutants, i.e., R14A/K17A and R14D/K17D. All the mutant CPs express normally (SI Appendix, Fig. S2 and Table S2), but only the ones at R8 form VLPs equivalent to those seen with wild type (SI Appendix), consistent with the basic sidechains contributing positively to assembly. All the variant proteins were examined for their abilities to bind RNA oligos encompassing either a single PS (B3) or the 127-mer fragment (Fig. 3 and SI Appendix, Fig. S3). Neither double mutant bound either RNA under these conditions. R8A assemblies around B3 but requires a much higher (>10-fold) CP concentration to do so, consistent with it having a lowered affinity for the RNA. By 1 μM CP it forms T = 1 shells that are resistant to RNase challenge. The R8D variant fails to form any stable higher-order species with either RNA (SI Appendix, Fig. S3B). In contrast, the R8A variant binds to the 127-mer very similarly to wild-type CP, including undergoing an initial collapse in R8D, implying that reduced intrinsic RNA affinity can be compensated by cooperative PS binding. Unfavorable electrostatic interaction presumably explains the lack of assembly when R8D is titrated against the 127-mer. If we assume that the mutations do not significantly alter the unliganded CP conformation, these effects probe the role(s) of electrostatic interactions during assembly. They imply that charge neutralization is not an absolute requirement for assembly on longer natural RNA fragments. This is consistent with the PS-mediated, but not purely electrostatic assembly mechanism for this virus.

Given that multiple RNA PSs can overcome lower RNA-CP affinity, as expected for a process in which PSs act collectively, we examined how many PSs are required to generate cooperative assembly. Given the importance of PS3 and the effects seen for fragments containing five PSs, three subfragments of the 127-mer each containing PS3 were tested (Fig. 5B and SI Appendix, Figs. S4 and S5). These are PS1–PS3, PS2–PS4, and PS3–PS5. Each could bind CP at PS3 but differs in the numbers of flanking PS sites, from two 5’ or 3’ of PS3 to just one on each flank. Only the fragment with PS3 centrally located assembles RNase-resistant T = 1 shells, although it does not show a collapse, and the overall yield is lower than for the 127-mer. The other fragments form nonspecific aggregates that eventually spontaneously dissociate.

The interpretation of these results is nontrivial. The effects are clearly not purely electrostatic in nature since the PS2–PS4 fragment (66 nt) which assembles is shorter than PS1–3 (76 nt) and 1 nt shorter than PS3–PS5. To understand the specificity of these reactions we need to consider the folding propensity of each of the PS-encoding sequences. The secondary structure of the 127-mer shown (Fig. 4) was arrived at by constraining its folds to capture the maximum number of SLs with AXAX loop motifs present. In this fragment only PS1 and PS3 are predicted to have a favorable folding free energy (Mfold; ref. 32) in isolation. This is consistent with our previous assays, in which replacement of the CP recognition motifs within each PS with UUUU, and variations in their relative spacing with respect to PS3, resulted in markedly different assembly behavior (12). In solution these RNA molecules will exist as an ensemble of different conformations. Interaction with the STNV CP displaces this equilibrium, preferentially selecting a single or a few assembly-competent conformations in which the PSs are present. The assembly efficiency seen may therefore be related to the relative populations of such conformers in the ensemble and thus to the free-energy costs of imposing this conformation. Assessing the extent of a conformational ensemble is difficult. A sense of the likelihood of alternate structures can be obtained based on the first 100 secondary structures returned by the Sfold algorithm (33).
In principle, the minor conformer could promote assembly, but the spacing between PS1 and PS3 is too large to facilitate the cooperative effects of multiple PSs. Similar analysis of PS2–PS4 suggests that the dominant secondary structure does not contain any of the PS folds expected for the 127-mer. However, its predicted secondary structure contains two alternative SLs that are almost always present, one of which presents an AXXA sequence (SI Appendix, Fig. S5). Their relative spacing (4 nt) is short enough to see a cooperative effect. The PS3–PS5 fragment forms two SLs within 10–12 nt of one another, one presenting an AXXA motif as PS5. This would suggest an assembly-competent structure. However, in the ensemble of possible structures, this SL is only present in 6% of the potential folds (SI Appendix, Table S3), which may account for its assembly behavior (Fig. 3B).

The conformational scrambling behavior described above for the fragments encompassing three PSs probably reflects events in vivo, where it is known that sequences within the 127-mer participate in formation of a translational enhancer with sequences in the 3′ UTR (34). That complex cannot be present in the wild-type 127-mer. The natural viral sequences connecting these SLs were then replaced with strings of As and Gs until only one fold was most likely (Fig. 4 and SI Appendix, Fig. S6). The relative separations of the base-paired stems were kept identical to those in the wild-type 127-mer. As a result of these changes, PSs 1, 2, 4, and 5 have been stabilized compared with the wild-type 127-mer, with all SLs having favorable folding propensity.

To assess the importance of the folding propensity of the dominant PS3 site we also created the following synthetic PS1–PS5 cassettes: (i) unstable PS1–PS5, cassette 1 (C1), in which the folding free energy of PS3, the central PS, is positive (0.3 vs. −2.6 kcal/mol), i.e., a scenario in which PS3 is unlikely to fold spontaneously; (ii) stable PS1–PS5, cassette 2 (C2), in which the folding free energy of the central PS is more negative (−3.5 vs. −2.6 kcal/mol for the 127-mer), i.e., where PS3 is more stable; (iii) all PS3, cassette 3 (C3), in which all five PSs mimic PS3, with stems of all PSs extended to the same length (7 bp) and all CP recognition motifs identical to that in wild-type PS3; and (iv) synthetic, stabilized PS1–PS5, cassette 4 (C4), containing the artificial PSs 1, 2, 4, and 5 from stable PS1–PS5 and the artificial extended SL for PS3 from the all-PS3 construct. The latter is hyperstabilized with respect to PS3 in both the wild-type fragment and C2 (−7.6 vs. −2.6 or −3.5 kcal/mol, respectively).

To compare the behaviors of these cassettes we examined their predicted secondary structures. SI Appendix. Table S3 lists the frequency of occurrence of each PS in the ensemble of the first 100 secondary structures returned by the Sfold algorithm (33), together with their relative spacings. In addition, we compared their circular dichroism (CD) spectra. CD provides a physical signal (35), the molar ellipticity at 260 nm, that is proportional to the percentage of base-paired residues and/or tertiary structure. The measurements were made in a buffer containing calcium ions since these are required in the reassembly buffer, there being several Ca2+–binding sites within the STNV capsid (36). Titration of the test RNAs using 2 mM calcium, the concentration in reassembly buffers, results in mild increases (9–17%) in the 260-nm ellipticity, as expected (SI Appendix, Fig. S7A). The only exception is C1, which does not respond to the presence of the cation. The molar ellipticity values of all test RNAs in this buffer decline as expected with temperature (SI Appendix, Fig. S7B). CD ellipticities at 260 nm of all the RNAs differ, illustrating the complexity of comparing RNA conformational ensembles. The C1 is much less structured throughout the temperature range. Perhaps surprisingly given the apparent secondary structures, the wild-type 127-mer has the highest amount of structure at the lower temperatures. At the lowest temperature tested all the RNAs except C1 have roughly similar ellipticity values, implying that they had reached similar levels of denaturation.

All these cassettes, with the exception of C1, trigger assembly of VLPs at lower CP concentrations than the wild-type 127-mer, i.e., under these conditions it is a better assembly substrate. Remarkably, C3 also assembles more efficiently than wild type, even though it encompasses PSs that are longer than those found in the 127-mer, suggesting that there is some leeway in the PS secondary structure context in which the recognition motif is presented. This is a little surprising given the critical dependency on PS spacing around PS3 observed previously (12). The efficiency of assembly and the folding propensity of C3 notwithstanding, C4 is by far the best assembly substrate, assembling to VLPs most efficiently (i.e., it...
assembles more rapidly following the 100-nM CP titration point) (Fig. 4B).

These results suggest that it is possible to abstract the critical assembly features from a viral genomic RNA fragment. Given the alterations in the stem lengths and loop sizes in the synthetic fragments it would also appear that there is considerable scope for engineering templates with improved PS folding propensity.

Transfer of Critical Assembly Features to Genomic-Scale RNAs. As a test of whether these experiments have successfully identified essential assembly features, we examined how inclusion of this improved RNA “cassette” alters the assembly efficiency of the STNV-1 genome. As a control we created a genome lacking PSs by altering all AXXA motifs within stable SLs to UXXU (ΔAXXA) (SI Appendix, Fig. S9). This has only a modest effect on the total number of SLs that can form in the modified genome. We then created a series of chimeric genomes fusing C1, C4, and a 127-mer PS1–PS5U, in which all the loop motifs are substituted by Us, which we have shown previously are unable to support assembly in isolation (12), onto the wild-type genomic fragment from 128 to 1239 nt. In addition, we fused C4 to the equivalent ΔAXXA 3’ fragment (SI Appendix, Table S4). The resultant smFCS assembly curves, R0 distributions, and EM images are shown in Fig. 5 and SI Appendix, Fig. S9. The ΔAXXA genome aggregates, the RNA remains RNase-sensitive, and it is clearly a very poor assembly substrate, as expected for an RNA lacking PSs. C4, which is the best assembly substrate in isolation, only partially reverses these properties, creating an RNA that collapses but remains RNase-sensitive, forming only a few misshapen VLPs. This result confirms that genome assembly relies on PSs outside the 127-mer. Indeed, PS1–PS5U 127-mer lacking CP-binding motifs fused to a wild-type 3’ fragment has improved assembly properties, although it fails to collapse and appears to form aggregates of T = 1 capsids that resolve into discrete particles and malformed structures on RNase treatment. Clearly, well-regulated assembly requires both sets of PSs to be functional, and this is confirmed with the C1 chimera, which collapses but struggles to form T = 1 capsids, with the RNA remaining accessible to nuclease. The proof of these ideas is seen with C4 fused to a wild-type 3’ fragment. It collapses rapidly to T = 1, forming nuclease-resistant T = 1 shells in high yield. Thus, an artificial sequence encompassing improved PS characteristics is able to regulate the assembly pathway of a fragment that is over 10 times its own size.

Confirmation of this interpretation of the results was obtained by directly comparing assembly efficiency of wild-type genome versus the C4–wild-type chimera, each carrying a different dye, under conditions where there was only enough CP to package one of these RNAs fully. QELS and EM images of the products following elution from a gel filtration column identifies two types of VLPs (PLs). Fluorescence emission (SI Appendix, Table S5) for each VLP suggest that the chimera constitutes up to 70% of the resultant VLPs (SI Appendix, Fig. S9), i.e., the C4 cassette chimera outcompetes the wild-type genome for CP binding.

Discussion
We have shown that the dual codes inherent in RNA PS-mediated virus assembly, i.e., that genomic RNAs simultaneously encode a genetic message as well as instructions for efficient capsid assembly, are separable. An important question is why do the codes not separate during the course of viral evolution? RNA replication in ssRNA viruses occurs via error-prone processes that lead to creation of a quasi-species of genome variants. There are now three examples of viruses using RNA PS-mediated virus assembly where we have structural information that partially answers this question. In bacteriophage MS2 (12), human parvovirus-1 (19) and STNV (12), at least one of the PS sites in the viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.

The focus here is the assembly code liberated from the wild-type viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.

The focus here is the assembly code liberated from the wild-type viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.

The focus here is the assembly code liberated from the wild-type viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.

The focus here is the assembly code liberated from the wild-type viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.

The focus here is the assembly code liberated from the wild-type viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.

The focus here is the assembly code liberated from the wild-type viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.

The focus here is the assembly code liberated from the wild-type viral RNA sequence. Indeed, by sequentially investigating each aspect of the STNV assembly sequence in its natural context we have been able to reproduce its effects in triggering in vitro assembly of STNV CPs using a synthetic nonviral RNA cassette. Additional refinements allowed us to produce sequences that are either less or more efficient than the wild-type STNV 127-mer, and we demonstrated that these effects can be transferred to genome-length RNAs. These results confirm the nature of PS-mediated assembly for STNV. Assembly in vitro initiates within the 127-mer by CP recognition of the PS3 stem-loop. Higher-order CP binding is dependent on the correct positioning and folding of the neighboring PSs (PS2 and PS4), each presenting a consensus CP recognition motif in the loop. The 127-mer potentially encompasses five PSs that make the initial binding cooperative with respect to protein concentration, leading to a collapse in the hydrodynamic radius of the RNA, a necessary precursor to encapsidation. Thereafter additional PSs 3’ to the 127-mer ensure accurate completion of the viral lifecycle.
viral capsid. Electrostatic interactions contribute to these protein-RNA contacts but are not the major driving force, which instead is a high-affinity sequence-specific interaction of the stem and loop regions of the PSs with the inner surface of the protein capsid.

Previously, Wilson and colleagues (37, 38) showed they could drive assembly of nonvirion RNAs into rods of tobacco mosaic virus (TMV) CP by creating RNA chimeras encompassing the TMV assembly initiation site. This was successful, with the length of the protein-coated rods formed being determined by the length of the RNA being packaged, as expected from the known assembly mechanism. This approach was less successful when applied to spherical sRNA viruses, where the highest-affinity MS2 PS has positive effects on in vitro encapsidation of short RNAs, but is less important on longer ones (39). Note, all these experiments were done at micromolar concentrations, where the effects of PS-mediated assembly are obscured by the tendency of the CP-CP to form spontaneously (40). The results described above suggest an efficient route for encapsidation of bespoke, nonviral RNAs in shells of viral CPs. In vitro assembly may be possible for a large number of CP-RNA combinations, but it differs from in vivo assembly where, in many viruses, there is good evidence suggesting that only nascent genomic transcripts emerging from the viral polymerase complex are packaged into progeny virions. In such reactions, the RNA is very likely to fold kinetically, avoiding some of the issues with RNA conformational ensembles in the in vitro reactions such as those described here.

Viruses and VLPs are finding increasing potential in medical applications as gene therapy or drug-delivery vectors (28–30), as well as acting as nonreplicating synthetic vaccines. Viral protein shells are also of interest for nanotechnology applications. The results described here offer an important insight into ways to create such structures with high efficiency and potentially carrying non-viral RNAs with advantageous properties. This will be essential for the production of designer synthetic viros.

Materials and Methods

Wild-type and mutant STNV CP was prepared by dissociation of recombinant VLPs produced in Escherichia coli. RNAs used in assembly assays were either transcribed as described previously (12), or gene blocks were purchased (IDT) and cloned into a PACYC184 plasmid for subsequent transcription (SI Appendix). smFCS assays and data analysis were performed as described previously, with any variations described in SI Appendix (11).

ACKNOWLEDGMENTS. We thank John Proks, Peter Prevelige, University of Alabama, Birmingham, and Ian Robinson, University College London, for stimulating our analysis of “synthetic” packaging sequences. This work was supported by grants from the UK Biotechnology and Biological Sciences Research Council (BB/00667/X1 and BB/L02095/1). R. Twarock acknowledges funding via a Royal Society Leverhulme Trust Senior Research Fellowship (LT130088), P.G.S., and R. Twarock thank The Wellcome Trust for financial support for virus work (Joint Investigator Awards 011045 and 01146).

1. Stockley PG, et al. (2016) Bacteriophage MS2 genomic RNA encodes an assembly instruction manual for its capsid. Bacteriophage 6:1157666.
2. Prevelige PE, Jr (2016) Follow the yellow brick road: A paradigm shift in virus assembly. J Mol Biol 428:416–418.
3. Garmann RF, et al. (2014) Role of electrostatics in the assembly pathway of a single-stranded RNA virus. J Virol 88:10472–10479.
4. Rudnick J, Brunsma R (2005) Tosaheaded packing of RNA viral genomes. Phys Rev Lett 94:038101.
5. van der Schoot P, Bruinsma R (2005) Electrostatics and the assembly of an RNA virus. Biophys J 89:1663–1668.
6. Belyi VA, Muthukumar M (2006) Electrostatic origin of the genome packaging in viruses. Proc Natl Acad Sci USA 103:17174–17178.
7. Balint R, Cohen SS (1985) The effects of dicyclohexylammonium on polyamine biosynthesis and incorporation into turnip yellow mosaic virus in Chinese cabbage protoplasts infected in vitro. Virology 144:194–203.
8. Brunsma RF (2006) Physics of RNA and viral assembly. Eur Phys J E Soft Matter 19:303–310.
9. Bercen C, Dreesens LL, Liu KN, Knobler CM, Gelbart WM (2017) The role of viral RNA secondary structure in the self-assembly of viral capsids. Biophys J 113:339–347.
10. Liu F, Tabor JS (2012) The prevention of turnip crinkle virus definition by a specific packaging signal and RNA size. J Virol 76:1428–1435.
11. Borodavka A, Tuma R, Stockley PG (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. Proc Natl Acad Sci USA 109:15769–15774.
12. Patel N, et al. (2015) Revealing the density of encoded functions in a viral RNA. Proc Natl Acad Sci USA 112:2227–2232.
13. Dykeman EC, Stockley PG, Twarock R (2014) Solving a Levinthal’s paradox for virus assembly identifies a unique antiviral strategy. Proc Natl Acad Sci USA 111:5361–5366.
14. Basnak G, et al. (2010) Viral genomic single-stranded RNA directs the pathway toward the T=3 capsid. J Mol Biol 395:924–936.
15. Rolfsen Ø, Toropova K, Ranson NA, Stockley PG (2010) Mutually-induced conformational switching of RNA and coat protein undergrad efficient assembly of a viral capsid. J Biol Chem 285:309–314.
16. Rolfsen Ø, et al. (2010) Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. J Mol Biol 408:431–448.
17. Ford RJ, et al. (2013) Satellite tobacco necrosis virus (STNV) virus like particle in complex with the B3 aptamer. J Mol Biol 425:1050–1064.
18. Steward H, et al. (2016) Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. Sci Rep 6:22952.
19. Shackle S, et al. (2017) Genomic RNA folding mediates assembly of human parvovirus. Nucleic Acids Res 45:2076–2077.
20. Patel N, et al. (2017) HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. Nat Microbiol 2:17098.
21. Dykeman EC, et al. (2011) Simple rules for efficient assembly predict the layout of a packaged viral RNA. J Mol Biol 408:399–407.