Supplementary Information

Supplementary Tables

Loop motif	Sequence	ΔG of folding (kcal mol^{-1})
ACAA	ACAUGCAACAAUGCACAC	-2.6
AUUU	ACAUGCAAUUUUGCACAC	-2.6
UUUA	ACAUGCAUUUAUGCACAC	-2.6
GUUU	ACAUGCAGUUUUGCACAC	-2.1
UUUG	ACAUGCAUUUGUGCACAC	-2.9
AUUG	ACAUGCAAUUGUGCACAC	-2.6
GUUA	ACAUGCAGUUAUGCACAC	-3.4
GUUG	ACAUGCAGUUGUGCACAC	-2.5
AUUA	ACAUGCAAUUAUGCACAC	-2.6

Supplementary Table 1: B3 sequence variants. Loop motif (left), full sequence (middle) and the Mfold folding free energy values (right) of the B3 sequence variants.

CP variant	$A_{260/280}$
Wild Type	1.73
R8A	1.68
R8D	1.25
R8D’	0.91
R14K17A	0.82
R14K17D	0.85

Supplementary Table 2: $A_{260/280}$ ratios of the purified charge change STNV VLPs

	% of Correctly presented -A.X.X.A- motifs					% of 'correct' spacings				Maximum spacing variance
Cassette	**PS1**	**PS2**	**PS3**	**PS4**	**PS5**	**1-2**	**2-3**	**3-4**	**4-5**	
WT PS1-5	72	0	8	0	10	8	0	0	0	+7
C1	98	97	0	73	27	73	0	0	67	+7
C2	100	87	99	68	31	5	72	55	66	+7
C3	100	99	100	100	100	92	90	100	90	+4
C4	100	87	100	75	73	9	69	76	72	+7
PS1-3	78	2	9	*N/A*	*N/A*	0	0	*N/A*	*N/A*	*N/A*
PS2-4	*N/A*	0	2	0	*N/A*	*N/A*	0	0	*N/A*	+3
PS3-5	*N/A*	*N/A*	34	4	6	*N/A*	*N/A*	0	0	*N/A*

Supplementary Table 3: Analysis of suboptimal structures of RNA assembly cassettes. An ensemble of 100 secondary structure folds was created for each cassette using the Sfold algorithm (1). These folds were then assessed by the following criteria: The presence of an -A.X.X.A- loop in PSs1-5 was verified and shown as a percentage of "correctly presented motifs" (green = 60+, orange = 40+, red 0-39 throughout the table). The nucleotide spacing between each SL was measured, compared to the expected value (SI Appendix, Sup Fig 6) and also displayed as a percentage of "correct spacings". Where these spacings differed, the maximum nucleotide difference is given.

Genome organisation		***Rh Change with CP***	***RNase resistant***	***EM appearance***
5' (nts 1-127)	***3' (nts 128-1239)***			
WT	WT	Collapse with recovery to *T*=1	✓	Complete *T*=1 shells
Unstable	WT	Collapse with incomplete recovery to *T*=1	X	Few, mostly complete *T*=1 shells
Synthetic Stabilised	WT	Immediate collapse to *T*=1	✓	Complete *T*=1 shells
ΔAXXA throughout	ΔAXXA throughout	Increases, aggregates	X	Incomplete *T*=1 shells
PS1-5U	WT	Increases. Aggregates until addition of RNase A, upon which degradation occurs to the size of a *T*=1 particle	X	Mostly incomplete *T*=1 shells
Synthetic Stabilised	ΔAXXA throughout	Immediate collapse to *T*=1	X	'Squashed' *T*=1 shells

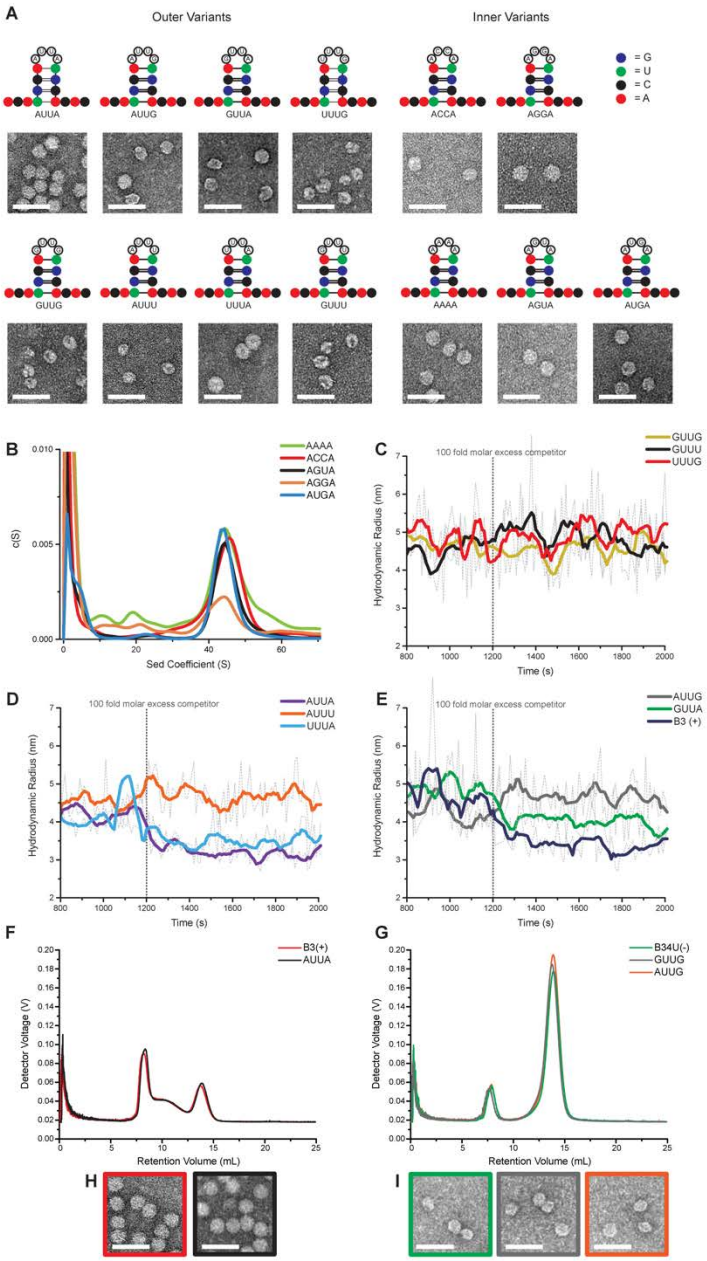
Supplementary Table 4: Summary of results with variant genomes.

Competition	Relative Yield of main peak / %	Relative Yield of minor peak / %	Main Peak		Minor Peak	
			RFU ʎ510 nm	RFU ʎ617 nm	RFU ʎ510 nm	RFU ʎ617 nm
STNV-1 (AF488) vs C4-WT (AF594)	72	28	8	205	68	0
STNV-1 (AF594) vs C4-WT (AF488)	66	34	268	0	0	12

Supplementary Table 5: Yields of VLPs from a head-to-head competition assay with differing genomes.

Reassembly reaction were set up with STNV-1 (AF488-labelled, 1 nM), C4-WT (AF594-labelled, 1 nM) and 60 nM STNV-1 CP. To avoid bias caused by the dyes, the oppositely labelled RNAs were also examined. The reassembly products were analysed by QELS using a 658 nm laser following elution from a gel filtration column. Two peaks eluted from each experiment that are both complete VLPs, as judged by EM images and the fact that the RNAs are RNase resistant. The relative yields of each VLP were calculated by integrating the area under the main peak (9.5 ml, SI Appendix, Sup Fig 9) and minor peak (11.5 ml, SI Appendix, Sup Fig 9) using the peak analyser function in Origin Pro 9. The relative fluorescence emission units (RFU) at the expected wavelength maxima (510 nm and 617 nm) were measured to determine the identity of the genome packaged. The major peak contains VLPs, which always contain C4-WT and the VLPs within the minor peak appears to always contain the STNV-1 genome.

Supplementary Figures

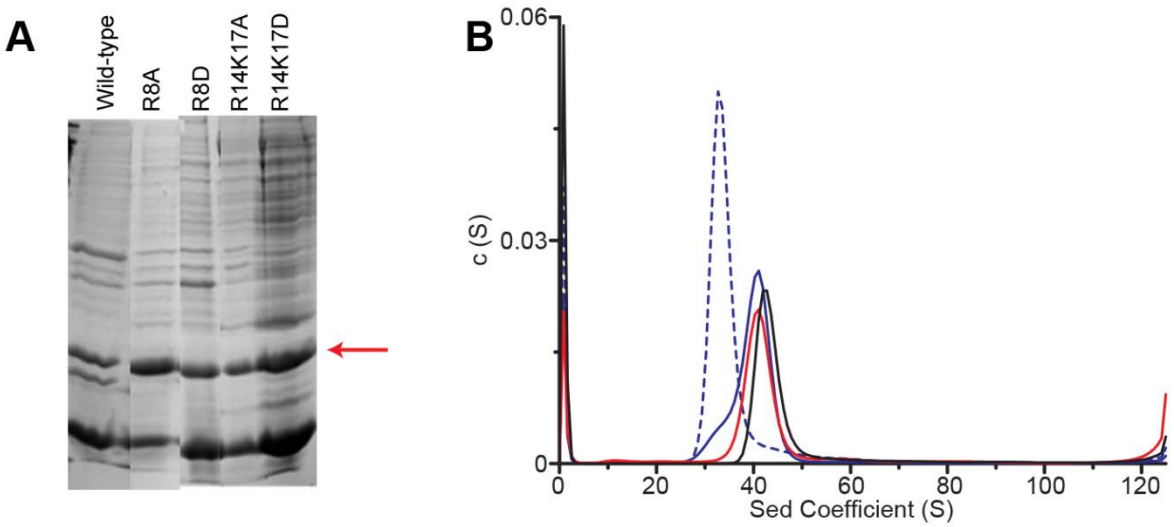


Supplementary Figure 1.

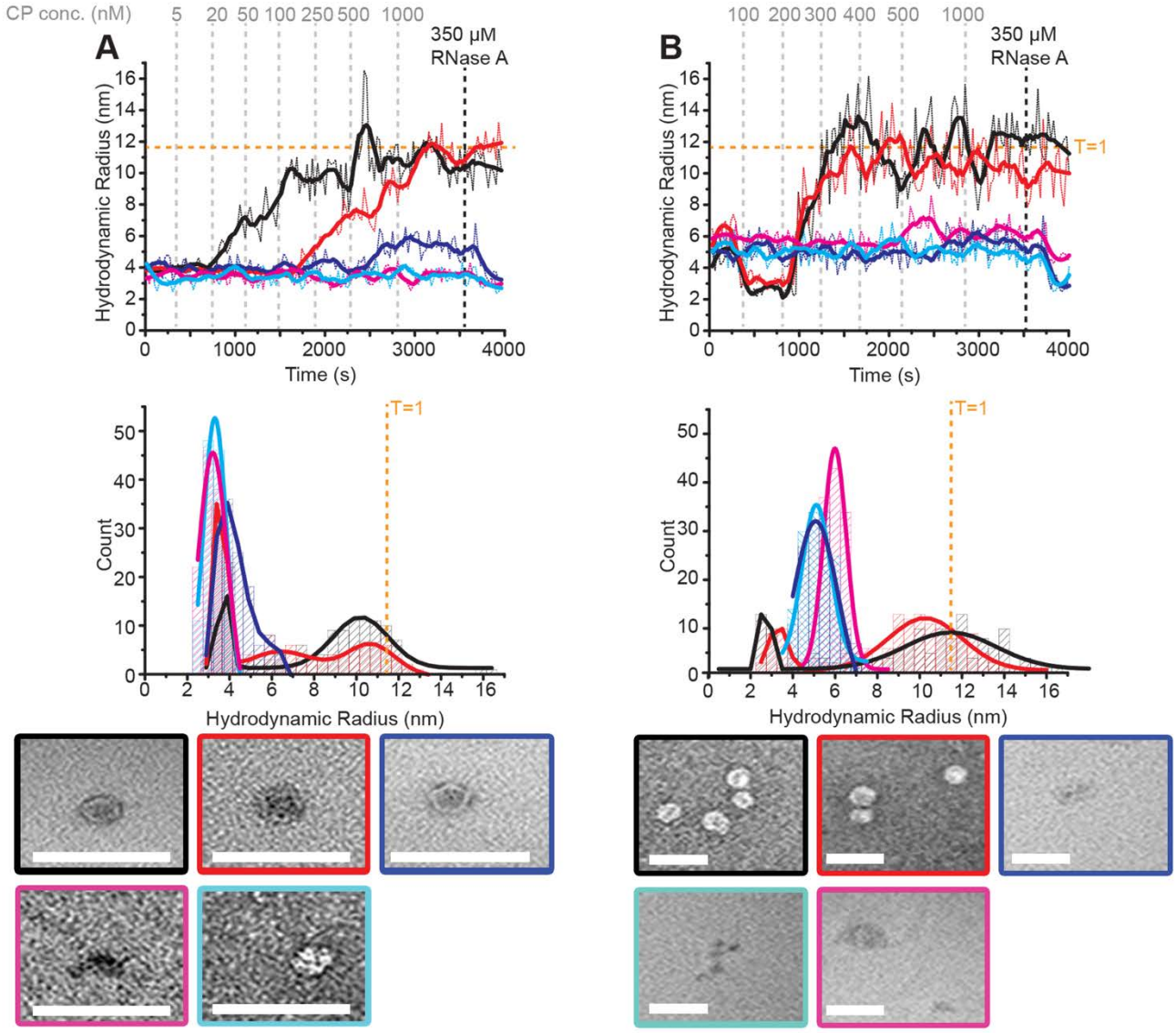
(A-C): Assembly behaviour of B3 variants. (A) MFold structures of all B3 variants with loop sequences given under each. EM images of B3 variant assemblies at 6 µM CP concentration are displayed below corresponding structure. Scale bar, here and throughout = 50 nm. (B) Comparison of STNV CP reassembly using B3 inner variant RNAs via svAUC (Methods). Reassembly reactions were carried out at a 1:3 RNA:CP ratio with final CP concentration of 6 µM monomer.

(C-E):B3 Variant-capsomer displacement assays. (C) 15 nM STNV CP and 5 nM Alexa Fluor 488 labelled B3 were used to form a capsomer of ~5 nm R_h. This was equilibrated for 1200 s, at which point 100-fold molar excess of a B3 variant GUUG (gold), GUUU (black) or UUUG (red) was added and the resulting R_h change tracked for 800 s. The data were then used to compute the percentage change in the R_h (Fig 2C). (D) As in (C) with B3 variant AUUA (purple), AUUU (orange) or UUUA (cyan). (E) As in (C) with B3 variant AUUG (grey), GUUA (green) or unlabelled B3 (navy blue).

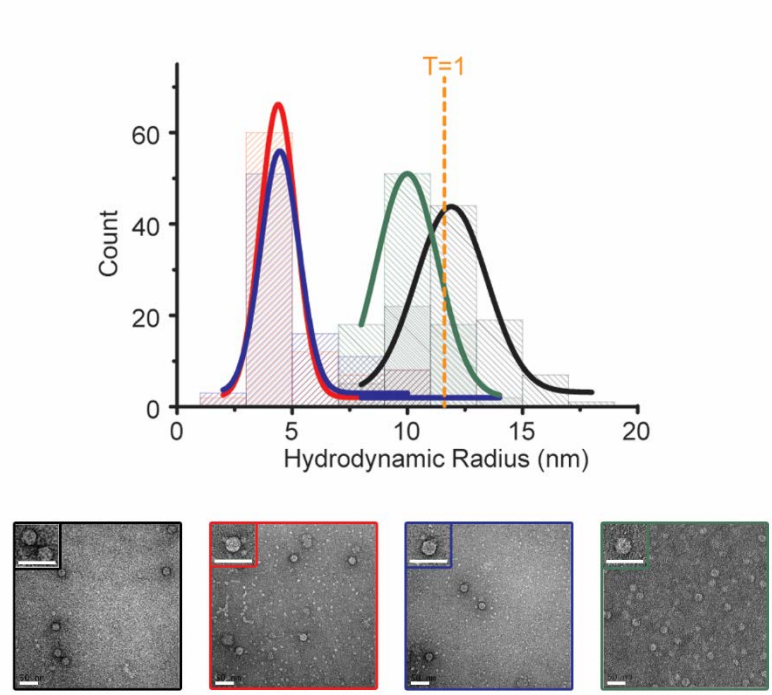
(F-I) B3 Variant DLS. (F) Ensemble reassembly of 'good' B3 variant RNAs, analysed by dynamic light scattering (DLS), B3 (red), AUUA (black). Areas under the curves calculated for the assembled VLP peaks (7.0-9.6 mL) B3 = 0.139 and AUUA = 0.137. (G) Ensemble reassembly of 'bad' B3 variant RNAs, analysed by DLS, B34U (green), GUUG (grey) and AUUG (orange). Areas under the curves for the VLP peak (7.0-9.6 mL) B34U = 0.085, GUUG = 0.059 and AUUG = 0.088. (H) & (I) EM images of the B3 variant assemblies shown in (G) and (H) at 6 µM CP, coloured outlines match the traces above.



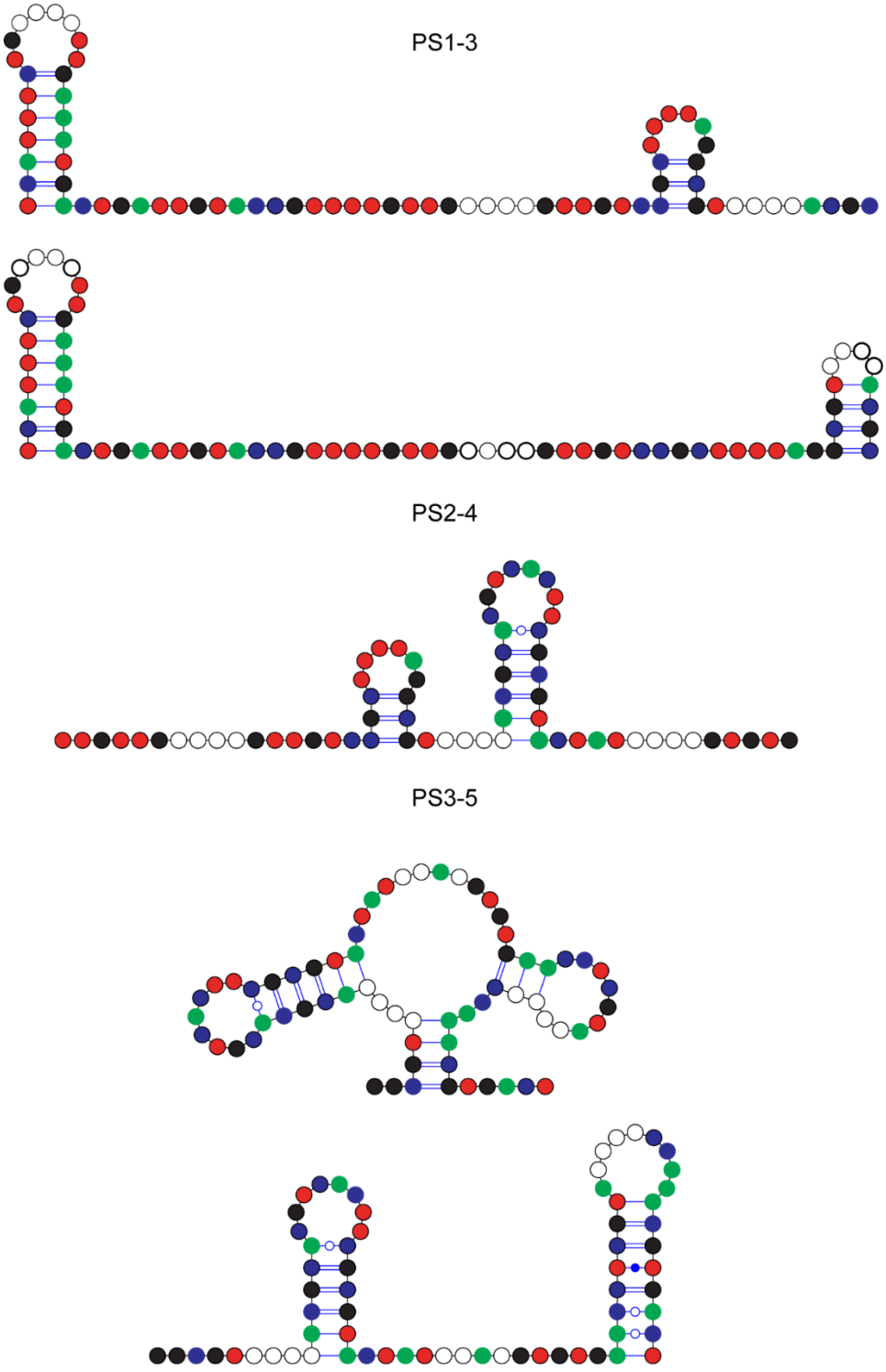
Supplementary Figure 2: Characterisation of the STNV CP charge-change mutants: (A) SDS PAGE of charge-change mutants expressed in *E.coli,* red arrow indicates the STNV CP band. **(B)** svAUC of charge-change mutants purified from *E.coli*. Only mutants that produce enough VLPs (wild-type (black), R8A (red) and R8D (blue and dashed blue [termed R8D']]) were analysed. Further detail on the purification and characterisation of charge-change mutants can be found in the Methods section.



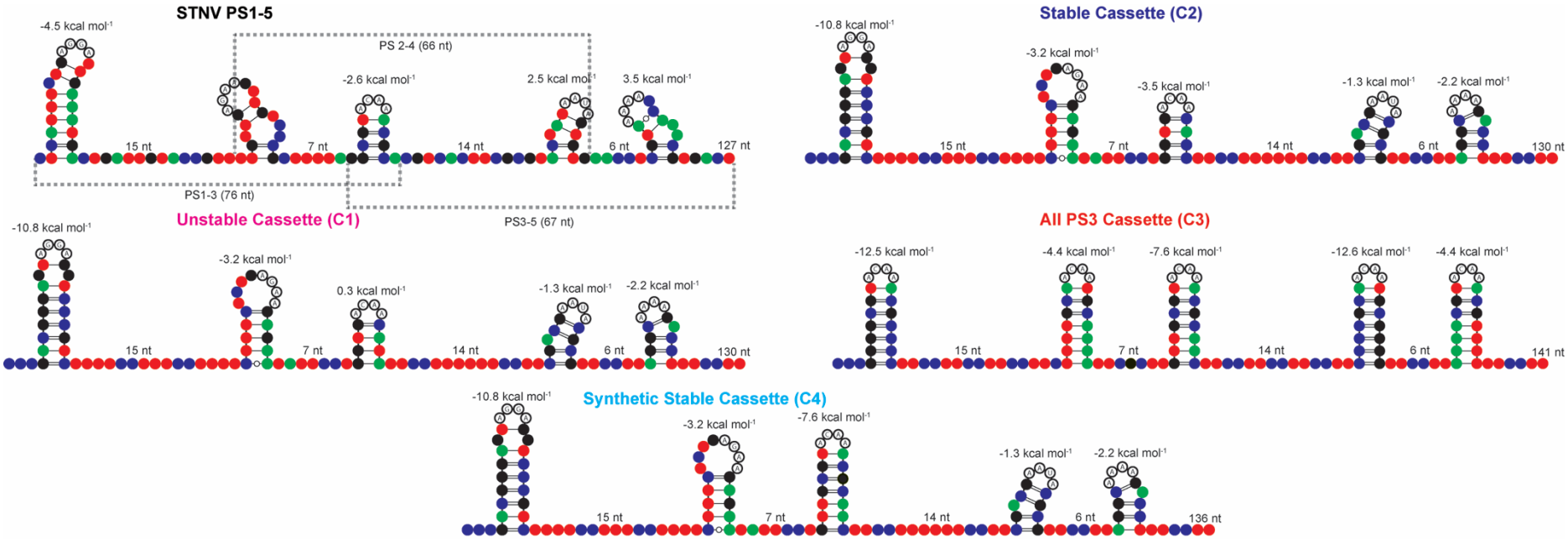
Supplementary Figure 3: Assembly behaviour of STNV CP charge-change mutants with single or multiple packaging signals. (A) Top: Wild-type (black), R8A (red), R8D (blue), R14K17A (cyan) or R14K17D (magenta) was titrated into 1 nM B3 and the resulting hydrodynamic radius monitored as described in Fig 3. Middle: R_h values from the upper chart shown as a distribution, with 0.5 nm bin sizes, were fitted with Gaussian peaks to highlight the heterogeneity present in solution throughout the reaction. The R_h of a T=1 particle is indicated by the orange dashed line. Bottom: EM images of the above reactions, with the same colour coding. **(B)** As in (A) but with CP titrated into 10 nM PS1-5.



Supplementary Figure 4: Assembly behaviour of three PS fragments. Top: R_h values corresponding to 3000-4000 secs of the reactions between STNV CP and PS 1-5 (black), PS 1-3 (red), PS 3-5 (blue) and PS 2-4 (green) from Fig 3B, plotted as a distribution, with 2 nm bin sizes and fitted with Gaussian peaks. Bottom: EM images of the reactions from Fig 3B.

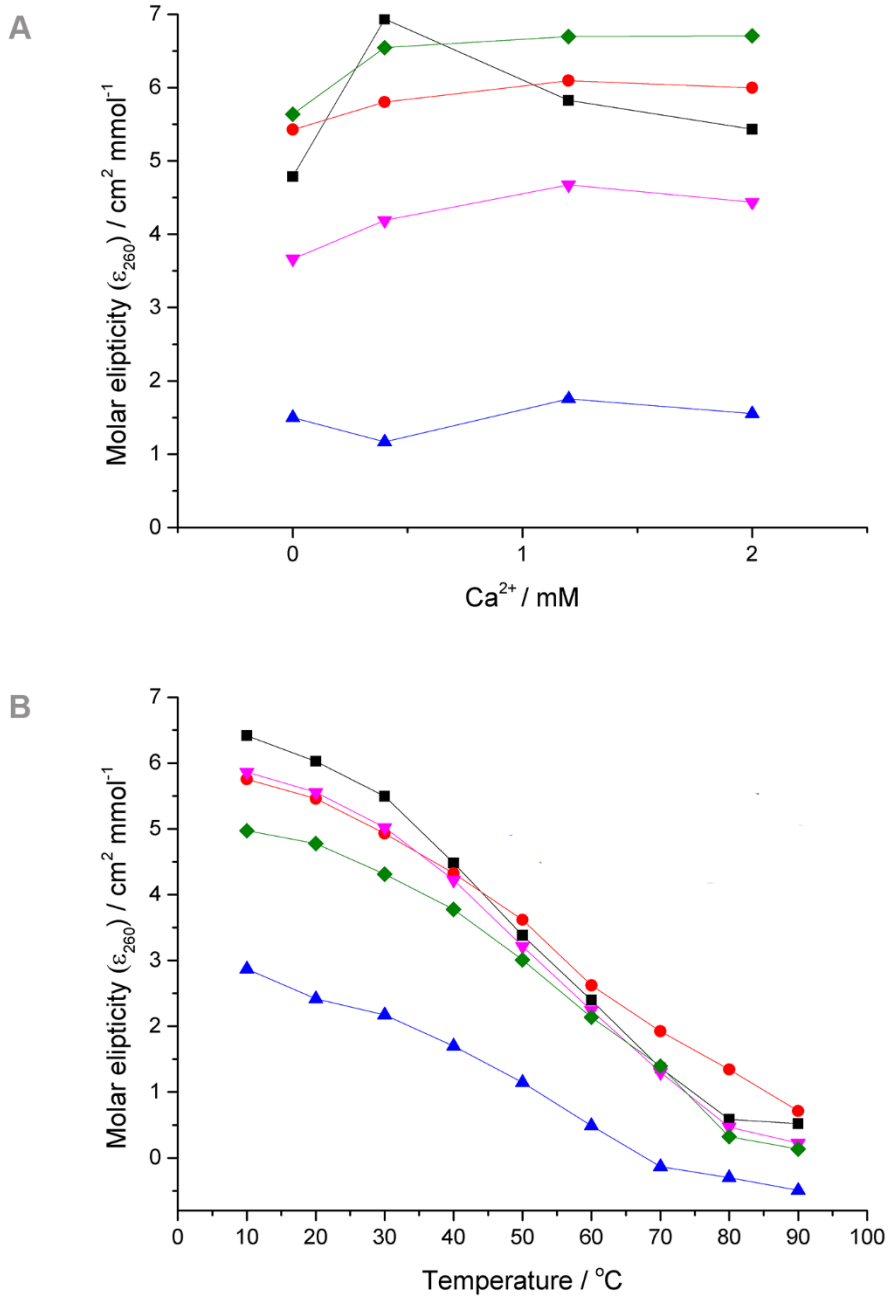


Supplementary Figure 5: Stable structures of the three PS fragments. Most stable structures of PS1-3, PS2-4 and PS3-5 as predicted by Mfold. Colour coding is as in Fig 1 and throughout. The two main PS1-3 folds predominantly present two -A.X.X.A- SLs, however these appear to be too distant for cooperative assembly. PS2-4 presents two SLs, one with an -A.X.X.A- motif, and another one 4 nts away. The two main folds of PS3-5 also present two SLs, but whilst these SLs are close (10-12 nt), only one presents an -A.X.X.A- motif, and that only 6% of the time.

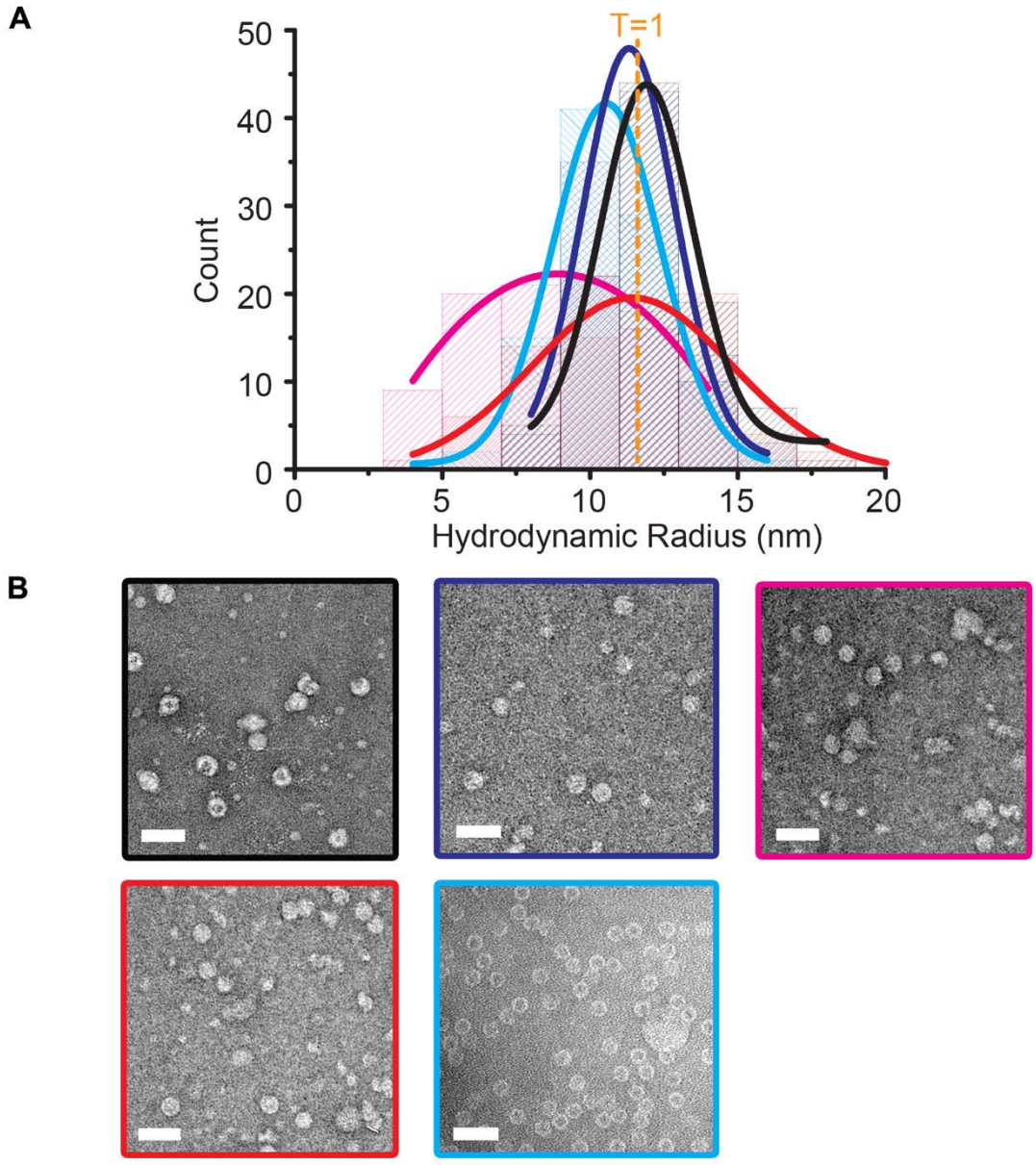


	Free Energy of Folding (kcal mol^{-1})				
Oligonucleotide	**PS1**	**PS2**	**PS3**	**PS4**	**PS5**
PS1-5	**-4.5**	**N/A**	**-0.3**	**2.5**	**3.5**
C1	**-10.8**	**-3.2**	**0.3**	**-1.3**	**-2.2**
C2	**-10.8**	**-3.2**	**-3.5**	**-1.3**	**-2.2**
C3	**-12.5**	**-4.4**	**-7.6**	**-12.6**	**-4.4**
C4	**-10.8**	**-3.2**	**-7.6**	**-1.3**	**-2.2**

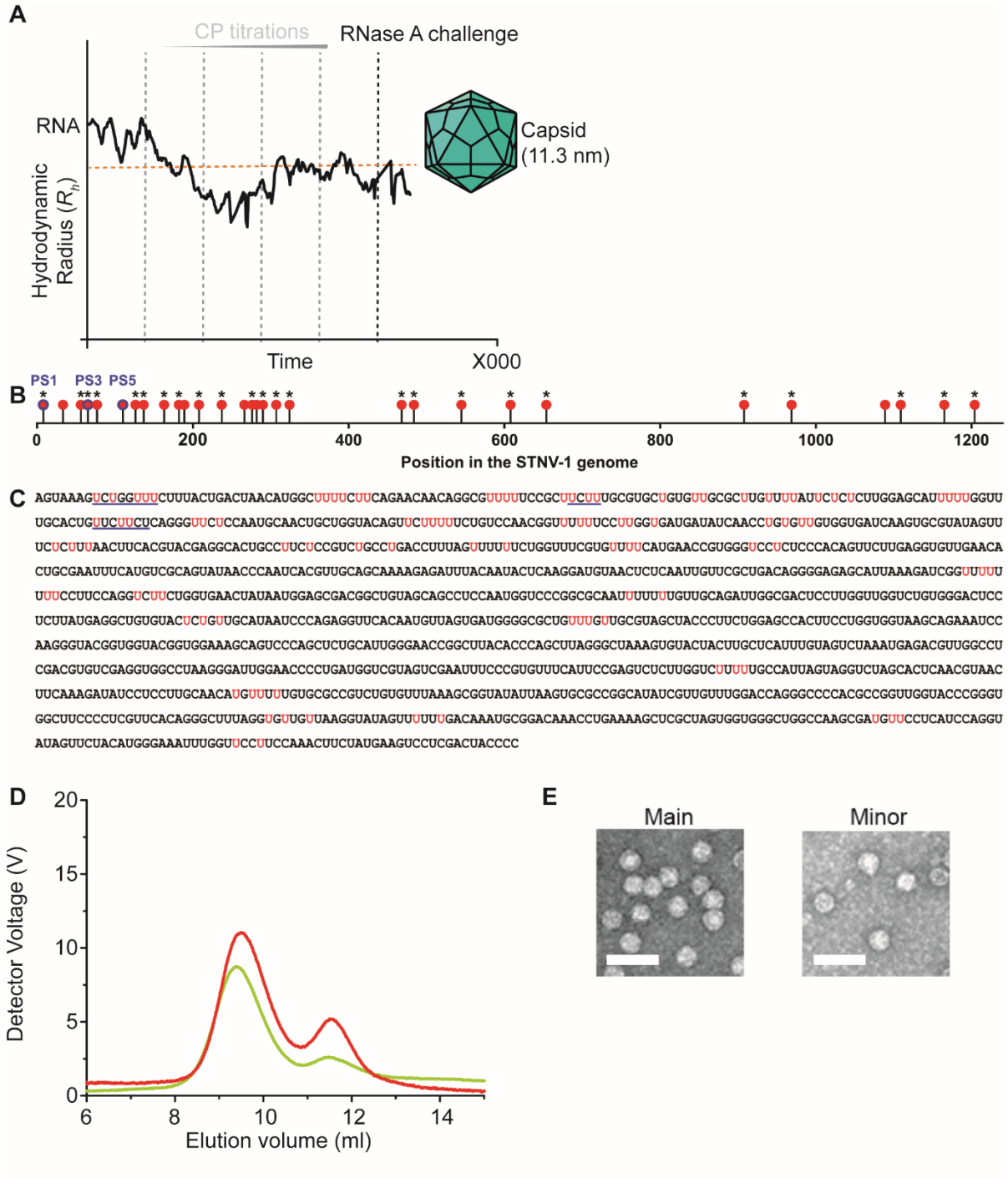
Supplementary Figure 6: 127-mer cassette structures. Top: sequences, putative secondary structures and folding free energies as determined by Mfold of the STNV wild-type PS 1-5 fragment and its synthetic counterparts. Bottom: table comparing the relative folding free energies of each SL as determined by Mfold. Oligos are colour-coded as in Fig 4.



Supplementary Figure 7: Interrogation of the structure and stability of the 127-mer cassettes. (A) Molar ellipticity of RNA oligonucleotides WT PS1-5 (black squares), C1 (blue triangles), C2 (red circles), C3 (magenta triangles) and C4 (green diamonds) at 260 nm with increasing Ca^{2+} concentration. **(B)** Molar ellipticity at 260 nm during thermal melts of the RNA oligonucleotides, colour-coded as in (A).



Supplementary Figure 8: Assembly behaviour of the synthetic 127 mer cassettes. (A) The R_h values between 3000–4000 secs of reactions containing STNV CP and PS 1-5 (black), unstable cassette (C1, magenta), stable cassette (C2, blue), cassette PS3 (C3, red) or synthetic, stable cassette (C4, cyan) from Fig 4A, plotted as in SI Appendix, Sup Fig 4. **(B)** EM images of the above reactions, colour coded as in Fig 4A.



Supplementary Figure 9: Assembly behaviour of the genomic chimeras. (A) smFCS reassembly using STNV-1 genomic RNA. R_h values were measured before and during STNV CP titration (vertical dashed lines), allowing the R_h values to equilibrate after each step. The STNV-1 genome initially collapses by up to ~30%, and then the R_h recovers to a value corresponding to a *T*=1 VLP (R_h ~11 nm). **(B)** Positions of SLs containing the AXX(X)A motif in the WT STNV-1 genome that have been mutated (red), including previously known PSs (blue). SL structures that have not been disrupted as a result of the introduced mutations have been marked with an asterisk (*). **(C)** The genomic sequence of STNV-1 mutant. The 102 single

adenine to uracil mutations have been highlighted in red. The apical loop sequence of PS1, PS3 and PS5 have been underlined in blue. **(D)** QELS data of a head-to-head competition between wild-type STNV-1 labelled with Alexa Fluor 488 and a C4-WT genomic chimera labelled with Alexa Fluor 594 (red trace), and the same genomes labelled with the opposite dyes (green), were reassembled with only enough STNV CP to form capsids around 50% of the input RNAs. These reactions were carried out in a buffer containing 50 mM HEPES (pH 7.5), 150 mM NaCl, 2 mM $CaCl_2$ and 1mM DTT. The samples were separated on a TSKgel G6000PWxl SEC column (Tosoh). Peak areas are listed in SI Appendix, Sup Table 5. The fluorescence intensities of each eluted peak were measured at each of the emission maxima for each dye to determine the amount of each genome in each peak. This revealed that the two peaks are due to the differing internal RNA content of each VLP. The C4-WT chimera elutes first whichever dye it is labelled with, and is always the larger of the two peaks (SI Appendix, Sup Table 5). **(E)** Representative EM images of the VLPs in each peak. The genomes in all four VLP products were resistant to RNase A treatment, confirming that the VLPs were complete shells. Since EM images suggest that the VLPs formed with both RNAs are superficially identical this behaviour suggests there is a somewhat altered internal arrangement of the two different RNAs in the VLPs.

Supplementary Materials & Methods

Purification of recombinant STNV CPs

Recombinant STNV VLPs were purified from E. coli (2). STNV charge-change mutant plasmids were created using primers designed using Agilent, and a Quikchange site directed mutagenesis kit (Agilent). CP monomers were purified by disassembly in 50 mM Tris (pH 8.5), 10 mM EDTA, in the presence of Complete Protease Inhibitor Cocktail (Roche, United Kingdom). STNV CP was separated from the mRNA by sequential Q-Sepharose, and SP-Sepharose columns (GE Healthcare, Sweden). STNV CP was washed with 20 column volumes of 50 mM HEPES (pH 7.5) and 25 mM NaCl to remove residual EDTA, and subsequently eluted using a 0.025 - 2 M NaCl gradient in buffer. CP elutes at 0.8 M NaCl. STNV CP was analysed by SDS-PAGE and its concentration determined by UV absorbance. Fractions with an A_{260}:A_{280} ratio of 0.6 or lower were used in assembly assays.

Characterisation of recombinant STNV CPs

Mutant CPs plasmids were designed and produced as stated above. Overexpression in *E.coli* resulted in comparable amounts of protein produced in the charge change mutants cf. wild-type STNV-1 CP (SI Appendix, Sup Fig 2). All STNV CP variants were purified over a 15-45% (w/v) sucrose gradient, isolating the STNV VLPs formed around *E.coli* mRNA. In the case of R8D, two bands of STNV VLP were purified, with one band, termed R8D' migrating slower through the sucrose gradient. The quantities of STNV VLPs produced in *E.coli* corresponded to the degree of the charge change. Wild-type produced the largest amount of VLPs, whilst R8A, R8D, R14K17A and finally R14K17D produced decreasing amounts of VLPs. The RNA content across the STNV variants appeared to differ, the VLPs appearing emptier as estimated by $A_{260/280}$ ratio, as more charges became less positive (SI Appendix, Sup Table 2). Analysis by svAUC was carried out on the wild-type and R8 variants, however, the double mutant variants did not yield enough VLP for svAUC analysis. Charge change VLPs R8A and R8D sedimented at 42 S, identical to wild-type while R8D' sedimented at 32 S.

Preparation of RNA Oligonucleotides

dsDNA transcripts encoding the RNA oligonucleotides used in this study were produced using primers and the KAPA2G system (KAPA biosystems) following the manufacturer's protocol. Transcriptions were carried out using a Highscribe T7 High yield RNA synthesis kit (NEB). Products were run on a denaturing RNA gel. The Alexa Fluor 488 labelled B3 oligonucleotide used throughout was synthesised and HPLC purified by DNA Technology A/S (Denmark). Other RNA oligonucleotides requiring a 5' fluorophore were labelled with an amino GMP during transcription and cross linked to an Alexa Fluor 488 SDP ester (Invitrogen) prior to gel purification as described previously (3).

Genomic chimeras were created by purchasing Gene blocks of the different genome constructs with a 5′ T7 promoter (IDT DNA technology), possessing BamHI and HindIII cleavage sites at either end to create sticky ends after restriction digestion. PACYC184 plasmid was also digested and dephosphorylated using Antarctic phosphatase (NEB). These gene blocks were then ligated into the PACYC184 plasmid using T4 DNA ligase (NEB). Transcription was carried out as above after linearization using HindIII.

RNA was annealed prior to each experiment by heating to 80°C for 90 s and cooling slowly to 4°C in a buffer containing 50 mM NaCl, 10 mM HEPES and 1 mM DTT at pH 7. Genomes were only heated to 65°C.

Sedimentation Velocity Analytical Ultracentrifugation (svAUC) of STNV CP Reassembly in the Presence of B3 Variants

Reassembly reactions were carried out in the presence and absence of B3 variants in a 1:3 RNA:CP ratio at a final CP concentration of 6µM, by dialysis into a buffer containing 50 mM HEPES (pH 7.5) and 2 mM Ca^{2+}. All samples were analysed by EM and AUC. For AUC, 0.32 mL of each sample was placed in a 1.2 cm path length 2-sector meniscus matching epon centrepiece cell constructed with sapphire windows. The samples were centrifuged at 15,000 rpm in an Optima XL-1 analytical ultracentrifuge (Beckman Instruments, Inc., Palo Alto, California 94304) at 20°C in an An50-Ti rotor. Changes in absorbance at 260 nm were detected by absorbance optics with 100 scans taken in approximately 11 hrs 30 min. Buffer densities and viscosities were calculated using the program Sednterp version 1.09 (4). Data analysed using Sedfit, version 12.1b 2010, using a continuous distribution c(s) Lamm equation model (5).

smFCS Data Collection and Analysis

FCS measurements were performed on a custom-built smFCS facility. Excitation laser (Sapphire CW blue laser, 488 nm, Coherent, USA) power was set to 65 µW. The focus position was adjusted to 20 µm from the cover slip inner surface (maintained by piezoelectric feedback loop, Piezosystems Jena, Germany). Immersion oil (refractive index 1.515, type DF, Cargille Laboratories, USA) was used with immersion oil objective (63 x magnification, numerical aperture 1.4). The photon count was recorded and analysed by an ALVL5000 multiple tau digital correlator (ALV-GmbH) in single channel mode. FCS data was analysed using non-linear, least-squares fitting with a single component diffusion model autocorrelation function corrected for the triplet state in Matlab. Diffusion time was used in the calculation of apparent hydrodynamic radius (R_h) and plotted as a function of assembly time. R_h calculations were based on the measured diffusion time for Alexa Fluor 488 dye with the estimated R_h of the dye (= ~0.7 nm in assembly buffer).

smFCS Assembly and Displacement assays

Initial measurements of Alexa Fluor 488 labelled RNA oligonucleotides were taken for at least 10 runs of 30 secs (5 min). Purified STNV CP was titrated into labelled RNA, as indicated by a black dashed vertical line. Sample additions, CP and RNase, take ~1 minute and the FCS trace and the derived R_h values often become very noisy just before and after these changes as a consequence of the way the data are averaged to produce the bold smoothed curve. Usually the large bumps or dips observed resolve rapidly. Each titration was measured for a minimum of 10 30 secs runs. In assembly assays this was repeated until full capsid assembly had occurred. At this point RNase A was added to probe RNA protection (indicated by a dashed vertical blue line). In displacement assays, once the sample had formed a capsomer structure (R_h=~5 nm) the sample was monitored for a further 120 runs of 10 secs (20 min to ensure stability). At this point unlabelled B3 short/B3 variant competitor was added in 100-fold molar excess and measured for 120 runs of 10 secs.

CD analysis

Transcribed oligonucleotides were diluted to 1.5 µM in 300 µl, in a buffer containing 10 mM MES, 50 mM NaCl and 1 mM DTT at pH 6. Measurements were performed on a Jasco J715 spectropolarimeter, from 200 to 350 nm, with a bandwidth of 2 nm. Each Ca^{2+} and STNV titration was inverted 5 times and allowed to reach equilibria for 2 min prior to the next measurements. Thermal denaturations were performed using a Peltier temperature control from 10- 95 °C in 5 °C steps, and an end scan was performed at 10 °C to check for cleavage. Each measurement was performed in triplicate and averaged. Data was converted to molar ellipticity using the equation: $\Delta\varepsilon$ (cm^2 mM^{-1}) = θ/(32980 C(mM) L(cm) N(no. of nt)).

ΔAXXA genomic mutant design

The STNV-1 (GenID: NC_001557) genome was scanned using a 60 nucleotide probe from the 5' end of the genome using the Sfold algorithm (1), that returned 1000 sampled RNA structure folds for each probe, from which all stem-loops (SLs) were extracted. SLs that were present in >10% of the samples and contained the AXX(X)A motif in an apical loop of up to 10 nucleotides in length had all adenines mutated to uracils. The new mutant genome was generated and probed again, repeating the process described above until no more SLs with the cognate motif could be detected.

Light scattering assay of reassembly with genomic RNA variants

The STNV-1 and C4 - Wild-type (C4-WT) genomes were designed and transcribed as described in methods, however an aliquot of each was labelled with Alexa Fluor 488, and a second aliquot was labelled with Alexa Fluor 594. Reassemblies were performed with 1 nM each of these oppositely labelled genomes (total genome concentration 2 nM) in a 96 well plate as in the smFCS assays, with CP titrated in until a final concentration of 60 nM STNV CP was reached, only enough CP to package 50% of the genomic RNA present. This 96 well plate was concentrated through a 100 kDa centricon (millipore) at 10k xg for 5 min and run on a TSKgel G6000PWxl SEC column (Tosoh) with an AKTA pure system (GE Healthcare) connected to a DAWN HELEOS and Optilab TrEX for QELS and refractive index measurements. The column flow-rate was 0.4 ml min^{-1} for 50 min. Peaks were fractionated, $A_{260/280}$ ratios measured, fluorescence intensities at ʎ510 and 617 nm measured with a nanodrop 3300 (Thermofisher) to identify the genome packaged and EM images obtained (SI Appendix, Sup Fig 9).

Two peaks eluted, one immediately after the other (~9 min and ~11 min respectively), which we term the main and minor peaks. These peaks were concentrated through a centricon as described above and the fluorescence intensity measured at ʎ510 nm with a blue LED, to check for Alexa Fluor 488 labelled RNA and ʎ617 nm with a white LED to check for Alexa Fluor 594 labelled RNA (SI Appendix, Sup table 5). The main peak always contained the C4-WT cassette VLPs while the minor peak contained wild type STNV-1 VLPs. The average $A_{260/280}$ ratios of these peaks were 1.65 and 1.62 respectively, showing no discernible differences in amounts of nucleic acid packaged assuming these were complete VLPs. RNase A treatment confirmed this, as R_h values pre and post RNase treatment remained constant.

Data Access

Original data files for smFCS, DLS and AUC experiments are available upon request.

References

1. Ding Y, Chan CY, Lawrence CE (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res* 32(Web Server):W135–W141.

2. Lane SW, et al. (2011) Construction and crystal structure of recombinant STNV capsids. *J Mol Biol* 413(1):41–50.

3. Patel N, et al. (2015) Revealing the density of encoded functions in a viral RNA. *Proc Natl Acad Sci USA* 112(7):2227–2232.

4. Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, eds Harding SE, Rowe AJ, Horton JC (Cambridge, Royal Society of Chemistry), pp 90–125.

5. Schuck P, Zhao H, Brautigam C, Ghirlando R (2015) Analytical Ultracentrifugation Basics. *Basic Principles of Analytical Ultracentrifugation* (CRC Press), pp 1–22.