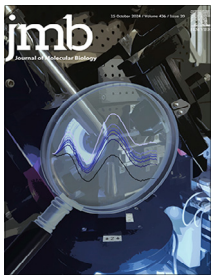

Visualizing Viral RNA Packaging Signals in Action

Emma Wroblewski [1,†], Nikesh Patel [1,†], Abid Javed [1,†], Carlos P. Mata [1,†,‡], Rebecca Chandler-Bostock [1], Lekshmi B. G. [2,3], Sabine M. Ulamec [1,||], Sam Clark [2,3,§], Simon E. V. Phillips [1,¶], Neil A. Ranson [1], Reidun Twarock [2,3] and Peter G. Stockley [1]

*1 - **Astbury Centre for Structural Molecular Biology,** University of Leeds, Leeds LS2 9JT, United Kingdom*
*2 - **York Centre for Complex Systems Analysis,** University of York, YO10 5DD, United Kingdom*
*3 - **Departments of Mathematics and Biology,** University of York, YO10 5DD, United Kingdom*

Correspondence to Nikesh Patel, Reidun Twarock and Peter G. Stockley:Corresponding authors at: York Centre for Complex Systems Analysis, University of York, YO10 5DD, United Kingdom (R. Twarock). *fbsnpat@leeds.ac.uk (N. Patel), reidun.twarock@york.ac.uk (R. Twarock), p.g.stockley@leeds.ac.uk (P.G. Stockley)*
https://doi.org/10.1016/j.jmb.2024.168765
Edited by M.F. Summers

Abstract

Here we confirm, using genome-scale RNA fragments in assembly competition assays, that multiple subsites (Packaging Signals, PSs) across the 5′ two-thirds of the gRNA of Satellite Tobacco Necrosis Virus-1 make sequence-specific contacts to the viral CPs helping to nucleate formation of its $T = 1$ virus-like particle (VLP). These contacts explain why natural virions only package their positive-sense genomes. Asymmetric cryo-EM reconstructions of these VLPs suggest that interactions occur between amino acid residues in the N-terminal ends of the CP subunits and the gRNA PS loop sequences. The base-paired stems of PSs also act non-sequence-specifically by electrostatically promoting the assembly of CP trimers. Importantly, alterations in PS-CP affinity result in an asymmetric distribution of bound PSs inside VLPs, with fuller occupation of the higher affinity 5′ PS RNAs around one vertex, decreasing to an RNA-free opposite vertex within the VLP shell. This distribution suggests that gRNA folding regulates cytoplasmic genome extrusion so that the weakly bound 3′ end of the gRNA, containing the RNA polymerase binding site, extrudes first. This probably occurs after cation-loss induced swelling of the CP-shell, weakening contacts between CP subunits. These data reveal for the first time in any virus how differential PS folding propensity and CP affinities support the multiple roles genomes play in virion assembly and infection. The high degree of conservation between the CP fold of STNV-1 and those of the CPs of many other viruses suggests that these aspects of genome function will be widely shared.

Introduction

Viruses can have devastating consequences for the health of the organisms they infect[1,2] despite often having only few molecular components. For example, all three known strain variants of Satellite Tobacco Necrosis Virus (STNV-1; 2 and C) encapsidate only the genetic information required to encode their coat protein (CP) subunits on a positive-sense, single-stranded (ss), genomic (g) RNA.[3] Tobacco Necrosis Virus is an essential co-infecting helper virus, which encodes the RNA-dependent RNA polymerase needed for replication of both genomes.[4] Each of the satellite viruses assembles an icosahedral $T = 1$ capsid[4–6] comprising 60 copies of their cognate CP subunits that

encapsidates their individual gRNAs. For STNV-1, this is a 1239 nt long gRNA (Figure 1). Its 195 amino acid long CP subunit consists of a shell-forming, eight-stranded, jelly-roll -barrel globular domain, and an extended amino-terminal helix ordered between residues 9–22.[5–7] These structural features are widely shared amongst many viral CPs[7–13] STNV-1 enables the principles of virion assembly, and the role(s) of its gRNA & CP in this process, to be studied in a simple setting.

We previously demonstrated that virion assembly in this model system relies on the RNA Packaging Signal (PS)-mediated mechanism.[14,15] *In vitro* studies of STNV-1 capsid assembly by single-molecule fluorescence correlation spectroscopy and analytical ultracentrifugation[16,17] as well as X-ray crystallographic structures of the resultant VLPs[18] revealed that the PSs consist of stem-loops (SL) in which the loop sequence is either 5´-A.X.X.A-3´ or a close variant. It is this feature which is recognised

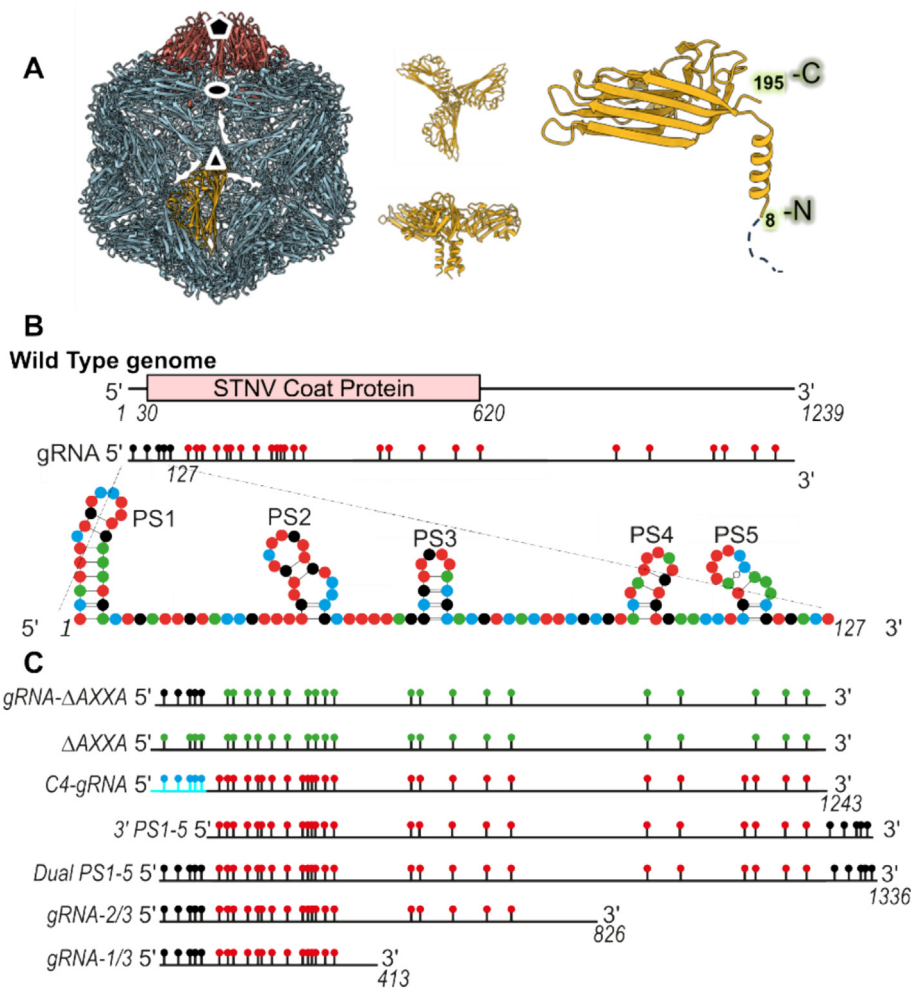

Figure 1. Sequence-specific assembly of STNV-1. (A) The STNV-1 virion structure. Left view shows the $T = 1$ capsid (PDB 4BCU) seen along a 3-fold axis. CPs are shown as ribbons, with a monomer (yellow) and a pentamer (red) highlighted. The middle image shows a CP trimer viewed from above (top) and the side (bottom). The rightmost image shows a single CP chain viewed from the side, with its ordered N and C termini labelled. **(B)** Top line shows the map of the STNV-1 genome, with the position of the CP ORF indicated. Middle line shows the assumed secondary structure of genomic RNA (gRNA; see Methods – PS folding analysis), showing the locations of putative PS sites (black stems topped by coloured circles, with those circles coloured black representing the proposed point of assembly nucleation, PSs1-5[16,17]; Bottom line shows the presumed secondary structure and sequence of the wild-type PSs 1–5 within the 5´ 127 nt (nucleotides are coloured as follows red = A; blue = G; green = U; black = C). **(C)** Schematics of the gRNA variants used here in assembly competition assays (sequences in Sup Figure 1). PS sites are shown as in (B), those lacking CP recognition sites or in which the folding propensity has been improved are coloured (AXXA) green or (C4-gRNA) blue, respectively[16,17]. Below are schematics of the gRNAs with transposed (3´ PS1-5) or duplicated (Dual PS1-5) PSs1-5, and the truncated gRNA variants (2/3-gRNA & 1/3-gRNA) used to create VLPs.

sequence-specifically by CP.[16,17] RNA PS-mediated assembly implies that viral genomes contribute to virion assembly regulation, in contrast to a purely electrostatic interactions between itself and the subunits of its CP shell.[15] Evidence for similar assembly regulation has been observed in a number of virions.[14,16] Its hallmarks in terms of sequence-specific genome CP shell contacts are numerous,[14,16,19,20] and these can be seen directly in asymmetric cryo-EM reconstructions,[21–24] or detected using X-ray footprinting experiments. However, the roles of the PS distribution along the gRNA linear sequence, and in the fully assembled virion during assembly and genome release, have not yet been characterised in any viral system. STNV-1 gRNA potentially encodes a series of at least 30 PS sites (Figure 1), each of which could fold to form a SL topped with a loop sequence encompassing 5´-A.X.X.X-3´, where X can be any nucleotide.[16–18,25,26] Assembly *in vitro* appears to initiate towards the 5´-end of the gRNA, where there are 5 PS sites located within the 5´ 127 nts of the gRNA.[16,17] The roles of the other putative PSs in the genome have not been determined before.[16,17]

Here we report the results of competition assays of sub-genomic fragments containing different numbers of PSs with structural data from asymmetric cryo-EM maps to dissect how different subgroups of STNV-1 gRNA PSs regulate assembly of the $T = 1$ shell of the virion efficiently around only the positive-sense gRNA, whilst also ensuring that the virion remains dynamic to enable subsequent intracellular release and replication of its genome. In particular, we characterise the numbers, and locations in the packaged structure, of the PSs involved in nucleation[27], providing unprecedented mechanistic details of nucleation in virion assembly. The similarity between the STNV-1 CP architecture, both with the other satellite viruses, and many other ssRNA virions,[9–13] suggests that similar assembly mechanisms are likely to be common in Nature.

Results

The roles of putative PSs across the STNV-1 genome

Assembly initiation *in vitro* occurs at PSs1-5 within the 5´ 127 nts of the natural gRNA[16,17]. The potential assembly roles of sets of PSs, and/or the negatively charged backbone of the STNV-1 gRNA, for assembly initiation and capsid formation were dissected here by creating a series of gRNA variants (Figure 1; Sup Figure 1). These variants probe the assembly roles of PS sequences 3´ to those contained within the 5´ 127 nts studied previously.[16,17] The outcomes of *in vitro* assembly with these new variants allow us to determine whether the relative positions of PSs, and their numbers in the gRNA, are important for assembly. Genome variants (3´ PS1-5 & Dual PS1-5) probe whether assembly initiation is directional and/or unique by moving the 127 nts, encompassing the PSs 1–5 sequence to the 3´-end, or by duplicating it at both ends of the gRNA. We also studied three gRNA variants in which PSs 1–5 are; (a) stabilised by replacing A.U & wobble base pairs with G.C base pairs, and otherwise ensuring that each PS is fully base-paired (yielding C4 PSs1-5[17]) or (b) with all the potential PSs across the gRNA missing their preferred loop sequences (i.e., by replacing A.X.X.A with U.U.U.U, = A.X.X.A); or (c) by ablating only potential PS sites 3´ to PSs 1–5 (gRNA-A.X.X.A). The impact of gRNA length on assembly was also probed using 3´-truncated versions of the wild-type genome (i.e. the 5´ 1/3 or 2/3 gRNAs). These variants eliminate 11 or 8 potential PSs, respectively. They also test whether purely electrostatic interactions between positively-charged CPs and negatively-charged gRNA fragments act as a driving force for assembly.

To determine the relative assembly efficiencies of these gRNA variants, *in vitro* pairwise assembly competition assays were used (see *Methods*). Note, all RNAs including those without PS sites, will trigger capsid-like reassembly *in vitro* in the presence of sufficient quantities of STNV-1 CP,[18] potentially complicating analysis. However, gRNA variants trigger assembly *in vitro* in the presence of stoichiometric amounts of CP, i.e. when there are 60 CP monomers/gRNA.[17] To distinguish the variant gRNAs within each competition experiment, we differentially-labelled them at their 5´ ends with distinct fluorophores. Reassembly reactions were carried out using robotic titration to ensure reproducibility.[28,29] Titration to excess CP stoichiometry of both aliquots of the differentially-labelled gRNA transcripts to be assessed in competition[30] was used to show that differential end-labelling does not influence assembly (not shown).

For each competition experiment, two aliquots of differentially-labelled gRNAs were allowed to reassemble *in vitro* in the presence of sufficient CP to assemble fully just a single aliquot of the gRNA. This process generates two types of virus-like particles (VLPs). Those that appear fully assembled in negatively-stained transmission electron micrographs (TEMs), have R_h values expected for a $T = 1$ capsid by light-scattering, and migrate like wild-type VLPs on sucrose density gradients or gel filtration columns (see “Peak 1” in Figure 2). The second type of VLP is incompletely assembled and migrates more slowly upon gel filtration (See “Peak 2” in Figure 2). Fluorescence emission values at 520 and 617 nm, the emission maxima of the dyes used, were determined for “Peak 1”; the native-like, VLP peak from such non-competitive assembly reactions (Table 1). The ratio of the differing fluorescence intensities of the gRNA fragments in this peak was then used to normalise the molar ratios of each

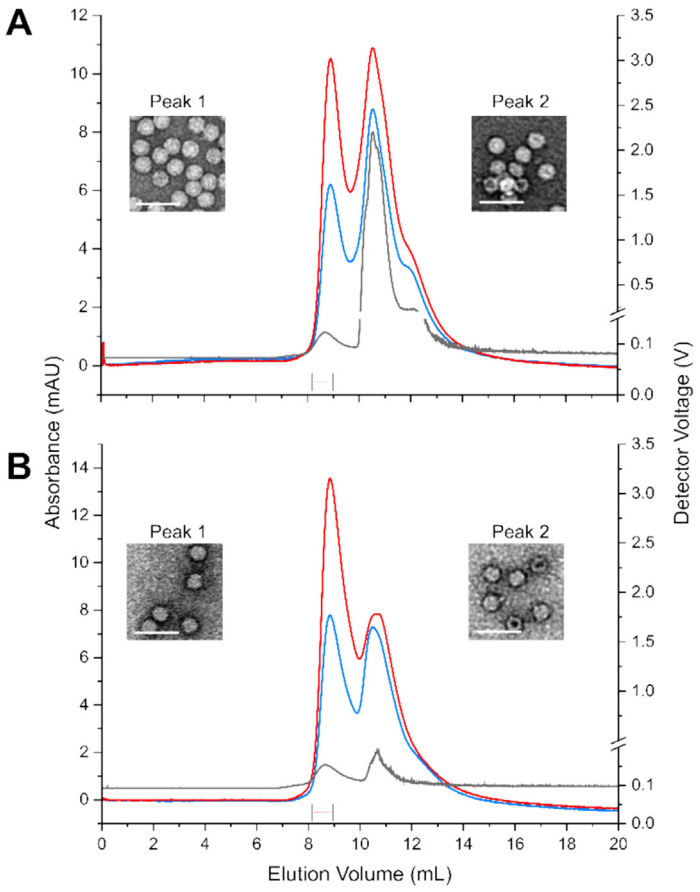

Figure 2. Gel-filtration elution profiles for *in vitro* reassembly experiments. Exemplar gel filtration profiles of *in vitro* reassembly products applied to a TSKgel G6000PWxl SEC column (Tosoh) connected to a SEC-MALLS system, showing the absorbance traces at 260 nm (blue), 280 nm (red) and light-scattering at 658 nm (grey) for: **(A)** gRNA-AF488 and gRNA-AF594 assembled in the presence of excess STNV CP and: **(B)** a competitive reassembly reaction between gRNA-AF488 and gRNA-AF594. Insets: Negative stain TEM images of material from Peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000. (Samples were pooled and analysed from the main VLP peak as indicated below.) gRNA-AF488 and gRNA-AF594. Inset: Negative stain TEM images of material from peaks 1 and 2. Scale bar = 100 nm, magnification = 30,000. **(**VLPs were pooled and analysed from the centre of the main VLP peak as marked.

Table 1 **Fluorescently-labelled RNA present in Peak 1 from the competitive reassembly experiments.** Left**,** reassembly conditions; competitive reassembly in the presence of 120 nM CP, with the 2 RNAs used in the reassembly experiments highlighted, centre, the fluorescent emission from the centre of the major VLP peak at $_{510}$, $_{617}$ normalised using the quantum yield of each fluorophore, and the ratio of $_{617/510}$, giving the amount of red (RNA2) to green fluorescently labelled RNA (RNA1), and right, the winner of the competition experiment with the fold difference in labelled RNA packaged in brackets

Assembly Competitions		$_{510}$ nm	$_{617}$ nm	$_{617/510}$ nm	Best Assembly Substrate (Fold differences between RNAs 1 and 2)
RNA1	RNA2				
gRNA	gRNA	2	2	1	N/A
gRNA	C4-gRNA	0.4	8.8	22	C4-gRNA (22)
gRNA	AXXA	8.4	2.4	0.3	gRNA (3.5)
gRNA-AXXA	AXXA	8.9	4.4	0.5	gRNA-AXXA (2)
gRNA	2/3 gRNA	5.5	6	1.1	2/3 gRNA (1.1)
gRNA	1/3 gRNA	1.9	0.9	0.5	gRNA (2.1)
gRNA	3´ PS1-5	3.1	1.8	0.6	gRNA (1.7)
gRNA	Dual PS1-5	3.8	1.5	0.4	gRNA (2.5)
3´ PS1-5	Dual PS1-5	0.1	0.2	2	Dual PS1-5 (2)

gRNA fragment (Table 1) in VLPs assembled under competitive conditions. The percentage of each fluorophore in the resultant "Peak 1" was then used to estimate the relative assembly efficiency of each genome variant under identical conditions (Table 1; Figure 1). Under non-competitive conditions, i.e. one in which a wild-type sequence RNA fragment labelled with one dye does not compete with an identical RNA labelled with the other dye, the native-like VLPs contained equimolar ratios of each dye, as expected for RNAs with similar labelling efficiencies.

The role(s) of the various STNV-1 gRNA regions/sequences for assembly were then compared (Suppl. Table 1). The largest assembly difference occurs when the natural gRNA sequence competes against the C4-gRNA variant, i.e. with the gRNA mutated to contain artificially stabilised base pairs within PSs 1–5. This variant has increased assembly kinetics[17] and is >20 fold more likely to be found in "Peak 1" compared to an unmodified wild-type gRNA. This outcome is consistent with assembly initiation occurring at PSs 1–5. However, when the wild-type 127-mer sequence is translocated to the 3'-end of a gRNA, or is duplicated at both ends, the wild-type gRNA sequence is the favoured assembly substrate (by at least 1.7-fold, see Table 1). This outcome implies that assembly initiation at PSs1-5 is normally followed by directional assembly 5′ to 3′ across the remaining STNV-1 gRNA. We assume that two assembly initiation complexes form on the gRNA variant carrying two copies of the PSs 1–5 sequence. Failure to convert such gRNA molecules carrying two assembly initiation complexes into VLPs efficiently presumably accounts for the nuclease sensitivity of that construct. The inferior performance of the gRNA containing a single copy of the 127 nt-long assembly initiation cassette at the 3′ end highlights the importance of the entire PS distribution. The PSs proximal to the assembly-initiation cassette at the 5′ end are more numerous and, on average, of lower folded free energies than those towards the 3′ end, with a gradual decrease in numbers. Our result suggests that this asymmetry in the PS distribution is important for assembly, contributing to assembly efficiency over and above the nucleation cassette itself.

The competition assay is also sensitive to alterations in assembly regulation by the other gRNA variants, e.g. when the template lacking the recognition features of PSs (A.X.X.A gRNA) competes with unmodified gRNA (Table 1). These RNAs are of the same length, yet the latter, i.e. the natural genome, is preferentially packaged by a factor of >3, confirming that sequence-specific recognition by the CP occurs.[16,17,23–25,31] These results are consistent with previous *in vitro* assembly studies.[16,17,25,29,32–38] Note, when the amount of CP required to form a complete shell around the natural gRNA molecule is added as a single aliquot, i.e. without the titration steps, regulation via PS-mediated interactions is dramatically reduced (4 fold) (Table 1). At the highest concentration the CP presumably binds to the gRNA at many differing PS sites, partially mimicking the outcome with two copies of the PSs 1–5 sequence at opposite ends of a genome.

Truncation of the native gRNA, removing either the putative 3′ PSs 25–30 (gRNA-2/3, a fragment 826 nts long), or PSs 20–30 (STNV-1/3, a fragment 413 nts long) (Figure 1) yields molecules shorter than the natural gRNA that have different properties (Table 1). The 2/3 gRNA fragment is a slightly better assembly substrate than its full-length competitor, under these conditions. The outcomes suggest that PSs 25–30, and/or the additional 413 negative charges of the deleted region, add very little to assembly efficiency. This is consistent also with the low frequency of occurrence of all but one of these PSs in global folds of the gRNA secondary structure (not shown), i.e. their inability to fold spontaneously in the absence of CP. Indeed, this applies to all putative PSs 3′ to nucleotide 550, i.e. beyond the 3′ end of the CP ORF. It may also be due to the reduced size of that gRNA fragment.[39] This region of the gRNA is known to encompass a translational enhancer, presumably for the CP ORF, and that might explain its continued presence[40] despite its negative impact on assembly efficiency. The shorter fragment lacks PSs 20–30, which encompass several of the morpresue frequently occurring PSs in the global fold, and 826 negative charges relative to the wild-type gRNA. Its reduced performance in the assembly competition assays suggest that one, or both, of these factors contribute to assembly efficiency.

Structure determination of STNV-1 VLPs

The competition data are consistent with STNV-1 CP subunits assembling in response to sequence-specific interactions with RNA sequences/structures within the 5′ 60 % of the gRNA. In order to visualise the inferred CP-gRNA contacts created in such particles, we prepared two VLPs (VLP PSs 1–5 or VLP gRNA-2/3, containing 127 or 826 nts of the gRNA) and determined their structures using cryogenic electron microscopy (cryo-EM) (Figure 3). These VLPs were formed by titrating a 1.5 M excess of CP, i.e. 90 CPs/RNA, into a solution containing 2 nM of either gRNA fragment (see Methods). These truncated gRNA fragments encompass all the putative PSs that contribute positively to VLP assembly *in vitro* (Table 1). Both assembly reactions produced VLPs in high yield. Surprisingly, both had identical $A_{260/280}$ ratios (=1.31) post-treatment with RNase, implying that multiple copies (7) of the PSs1-5 fragment were internalised in the same capsid,

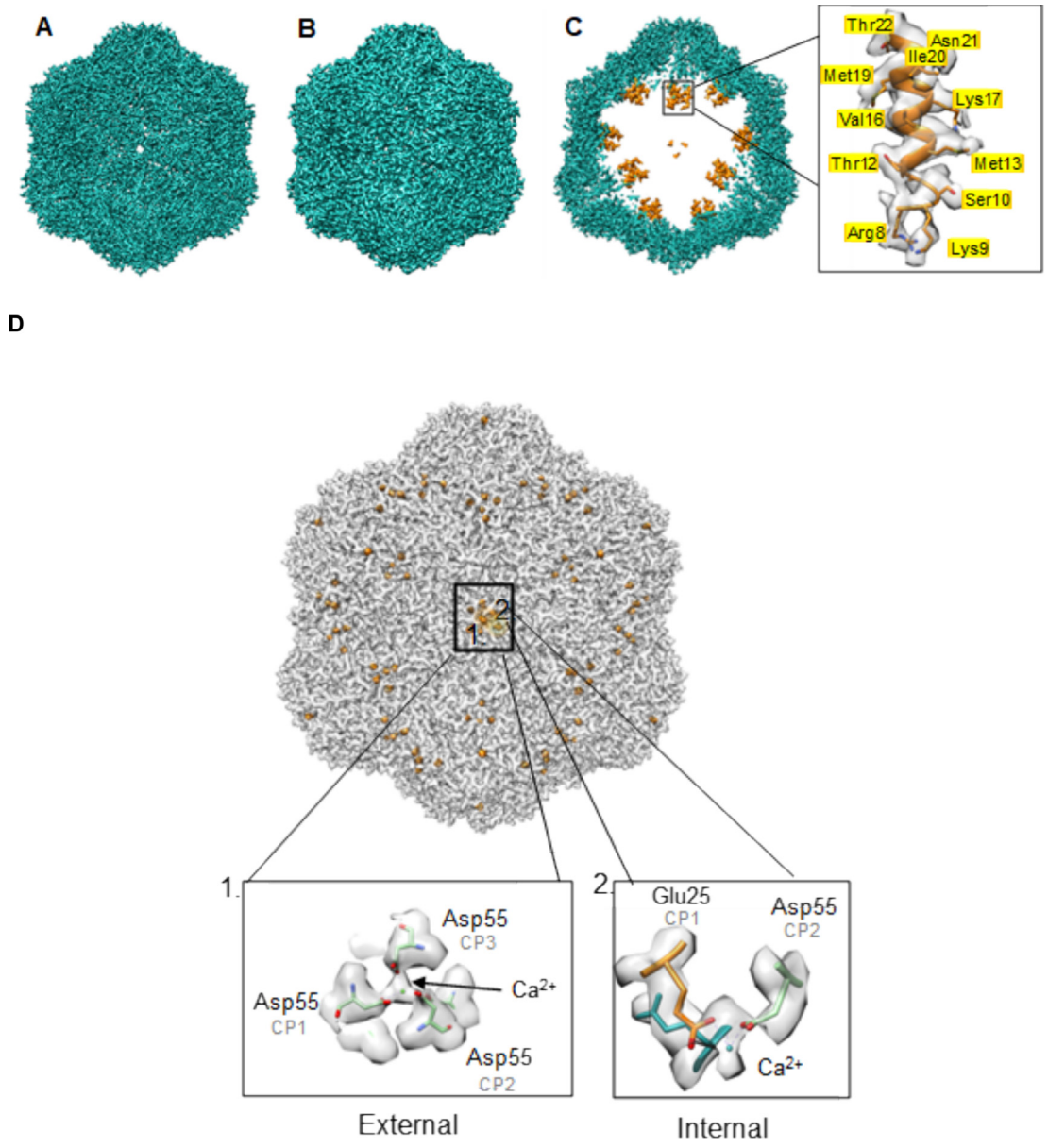

Figure 3. Icosahedrally-averaged structures of the STNV-1 VLPs assembled around PSs1-5 or 2/3-gRNA – the roles of chelated metal ions. Cryo-EM reconstructions of STNV-1 VLPs containing: (**A**) PS1-5 gRNA (EMD-18465); (**B**) 2/3-gRNA (EMD-18469) with the CPs coloured teal; **(C)** A cross-section of the capsid shown in (**A**) with the ordered N-terminal helical segments (orange) pointing towards the capsid interior. Boxed, inset: the ordered N-terminal helix of the CP (PDB ID: 4BCU, orange) rigid-body fitted into the map density. (D) Top: Map of the STNV-1 VLP CP (grey, transparent) encompassing PS1-5 RNA. (Note, this structure is essentially identical to the VLP encompassing 2/3-gRNA.) There is clear density for the presence of multiple Ca^{2+} ions in the VLP (shown in orange). These ions are bound in two calcium ion binding pockets located at the 3-fold (electron densities and model fits shown as inset #1) and at the 5-fold axes (shown as inset #2). These sites are labelled as “external” & “internal” to indicate their relative positions with respect to the VLP surface.

compared to the single copy of the 2/3 gRNA fragment in its VLP.

Cryo-EM reconstructions of purified VLPs (Figures 2 & 3) were calculated either by assuming icosahedral symmetry, or by expanding that symmetry to produce asymmetric electron density maps (Figure 4). As expected,[4,5] the icosahedral analyses yielded structures for both VLPs at higher resolution, with the CP subunits in both VLPs organised as $T = 1$ shells. The fold of the globular domain of each CP subunit is essentially identical in both reconstructions, and it also matches the fold for this region seen in previous crystal structures (Figure 3).[26,41,4,5] The N-terminal region of each

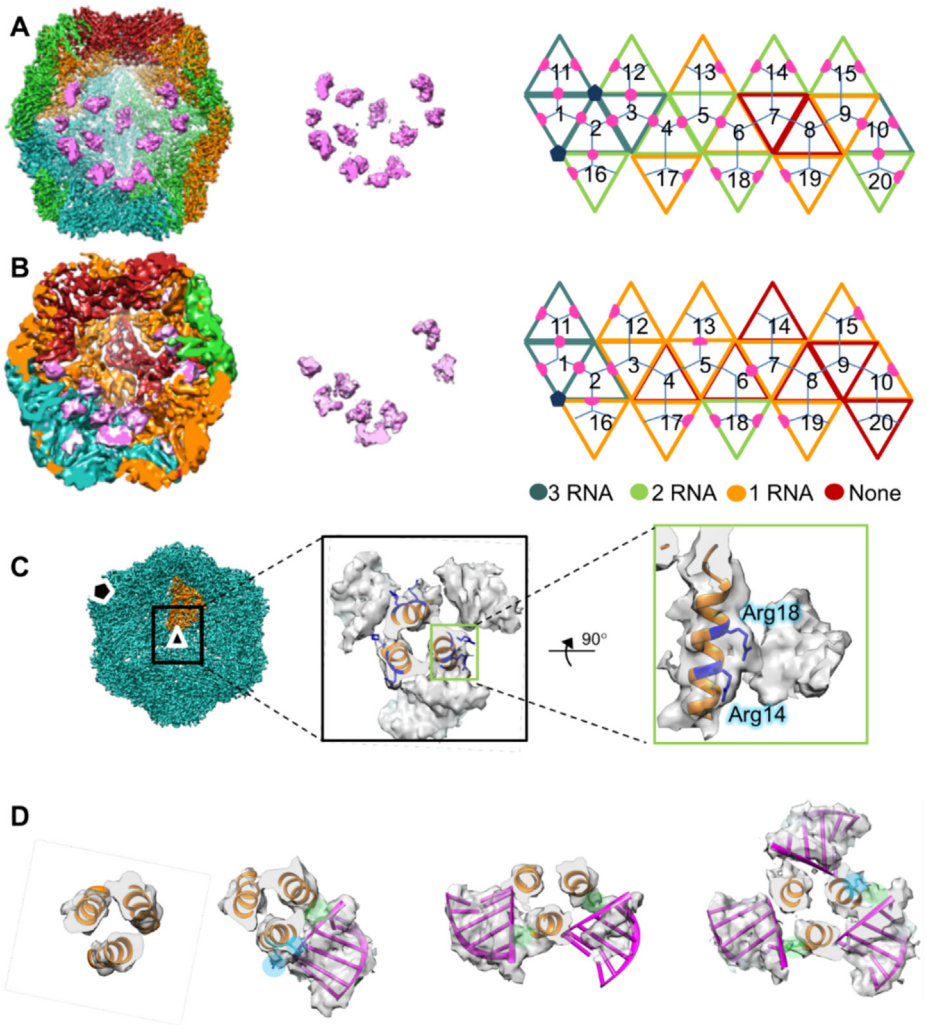

Figure 4. The asymmetric reconstructions of the VLPs. (A) Left, cross-sectional views along the 2-fold axes of VLPs assembled around PSs 1–5 **(A)** or 2/3-gRNA. **(B)**. Subunits are coloured according to the ordered RNA occupancy panel shown on the right as a series of triangles (labelled 1–20). Maps are visualised at 1.5 contour level. The distributions of the assumed, ordered RNA segments (pink) are shown in the middle, without their protein capsids. Right, representations of the triangular network forming each $T = 1$ capsid, coloured according to their assumed RNA occupancy. RNA segments are modelled as pink circles or semi-circles if segments are shared. **(C)** Left, asymmetric structure of STNV-1 VLP (teal) assembled around PS1-5 RNA (EMD-18466, Sup. Table 2), with one subunit shown in orange and the 3- and 5- fold axes indicated as symbols. Middle, cross-sectional view of the 3-fold axis looking towards the capsid interior showing the N terminal helices within their corresponding map density (grey) with positively charged residues which contact the non-CP density highlighted (blue). Right, side view of one of the N terminal helices. **(D)** As in (C; middle) the three-helical trimer is surrounded by 0 (left panel), 1 (left middle), 2 (right middle) and 3 RNA segments (right panel). Cryo-EM densities are shown in grey, STNV protein in orange, and modelled RNA segments in magenta.

CP subunit, beyond residue His23, is orientated towards the capsid interior, adopting an alpha helical conformation that extends to amino acid residue 10. Residues 1–9 are not seen in either type of reconstruction. The ordered section of the tails trimerise the CP subunits by interacting with each other via hydrophobic surfaces (Figures 3C and 4C). Residues Glu25 and Arg 55, form one of the calcium ion coordination sites in these capsids (Figure 3C) and sit at the “hinge” regions that connect the N-terminal residues to the globular body of each subunit. The CP capsids of both VLPs are also stabilised by bound calcium ions (Figure 3D) chelated at the 5-fold axes, as previously observed in the X-ray structure of native virus.[4] Asymmetric analyses reveal that the capsid shells of each VLP are complete (Figures 3 and 4), implying that a number of CP trimers within each shell cannot be in contact with the gRNA, i.e. those CP subunits assemble into the capsid without also binding to gRNA.

These VLPs also contain multiple inter-molecular contacts between their CP tails and non-protein electron densities (Figure 4C and D). The latter are present at relatively low resolution, are similarly located wherever they appear with respect to each CP trimer, and are a rough match to the shape expected for the electron density of an A-form RNA duplex at this resolution. A similar non-protein RNA density has been seen previously by X-ray diffraction in intact VLPs formed around either a simple stem-loop PS or a CP codon optimised mRNA.[18,26] Since the PSs in this system are expected to be SLs, we assume that we are seeing segments of the double-stranded gRNA used to trigger *in vitro* VLP assembly. The resolution of these segments is, however, insufficient to define their nucleotide sequences. CP trimers in these VLPs are surrounded by zero, one, two or three lozenges of dsRNA (Figure 4D). The number of ordered dsRNA segments differs in the differing VLPs, with the VLP encompassing PSs 1–5 having the most. Their differential occupancy may be a consequence of molecular frustration,[42] see also the Discussion). Where present, they are orientated such that each density contacts two CP helical tails. They do this by lying across patches of basic amino acids (Figure 4C), implying that these intermolecular contacts are not sequence-specific and further supporting their identification as gRNA fragments (Figure 4). The inferred loop sequences on these dsRNA PS fragments are unfortunately not visible in our cryo-EM maps, nor are the CP tails N-terminal to residue 10 which presumably encompass the PS loop binding motifs.

The structural basis of PS-mediated virus-like particle assembly

The reconstructions explain in part the roles of the PS-encompassing gRNA fragments in assembly initiation. They suggest that multiple RNA segments, encompassing multiple PSs, co-operate to stabilise CP trimers. Recombinant CPs in solution without RNA are uniformly monomers. We argued previously that electrostatic repulsion between STNV-1 CP subunits is overcome during assembly, e.g. by the contact of positively-charged amino acids to the backbone of gRNA fragments.[18]

Protein-gRNA interactions within the STNV-1 virion therefore have at least two components. CPs form sequence-specific contacts to RNA by recognising PS sites. This is the conclusion of our assembly competition data (Table 1). Since the asymmetric VLP structures do not reveal the molecular basis of this recognition, we assume that it occurs as the PS loops and the N-terminal 1–9 amino acids polypeptides in each CP interact. Such interactions are known to be altered by the nucleotide sequences in the loops of PS stem-loops[16,17] and may not be visible in our reconstructions because they form idiosyncratically with the non-identical gRNA sequences of the different PSs. In addition, there are also multiple non-sequence-specific contacts between the gRNA and CPs that help to stabilise the trimers of the CPs as they form the capsid.

Asymmetric reconstruction of the VLPs formed with multiple copies (7) of PSs 1–5 gRNA has non-CP densities at 17 of its 20 3-fold sites, implying multiple high-affinity contacts with PSs. The VLP assembled around the 2/3 gRNA fragment however shows many fewer such contacts, although this is inferred at medium resolution (5.26 Å). These outcomes are consistent with the unique sequence of the 2/3 gRNA fragment and the fact that it contains only a single copy of the PS1-5 sequence. The sequence of PS3, the central PS in PSs1-5, matches that of the most frequently obtained RNA aptamer selected *in vitro* against CP[18,25] therefore probably has the highest CP affinity for of any PS within the STNV-1 gRNA. The spacing to, and sequences of, its flanking PSs have also been shown to be important in *in vitro* assembly reactions.[16,17] This observation may be critical in understanding the biology of this virus. The high-affinity PS sites in the natural genome are limited to the 5′ end of its gRNA, whilst few if any PSs exist at the 3′ end of the genome facilitating its exit from the capsid and its contact with the replicase of the helper virus.

Discussion

The packaging of long polynucleotides into assembling or pre-formed isometric containers is a widespread requisite for viral function, and the role(s) of electrostatic interactions in this process has long been recognised as important in virus assembly and stability.[43–45] By working at low molecular concentrations, we have been able to show that many ssRNA viral genomes also regulate assembly of their protective protein shells by forming, in addition, defined sequence-specific interactions.[14,16,20,35] There are many potential advantages for virions that assemble via such genome contacts, e.g. preferential packaging of one strand over another, for example, since only the positive-sense strand can fold to display PS stem-loops. In some virions, PSs can be observed directly bound to CPs,[19,23,36] or such complexes can be detected via XRF,[20,29] e.g. on the interior of bacteriophage shells.

One consequence of our analysis is that there is a complex interdependence of gRNA and CP organisation, as might be expected in a system having multiple PSs formed from distinct nucleotide sequences. “Molecular frustration”[42] the conformational tension created as CPs interact with PSs from non-equivalent positions within viral capsids. Since PS sequences differ, we expect them to have differential CP affinities, a consequence of the packaging of a large gRNA molecule containing

those PSs into a viral capsid. These differential affinities will lead to some tightly bound PSs whilst others are more labile creating an apparently heterogenous viral particle. In ssRNA phages, it is known that sequence-specific CP-gRNA PS interactions switch the preferred conformation of an unliganded single polypeptide loop within their CPs. The combination of PS-bound and PS-unbound CP conformations triggers assembly of the $T = 3$ capsid.[33] The presence of CP-genomic RNA interactions within many other viral capsids,[19,32,36,46] suggests that this is a common aspect of many virion assembly processes. RNA PSs also appear to emerge spontaneously in artificial evolution experiments.[47] Such CP-genome contacts would be unlikely to occur if some virions did not use RNA regulation in assembly, at least in part. In the STNV-1 case described above, the asymmetric cryo-EM structures and our *in vitro* reassembly experiments suggest that capsid assembly is an RNA sequence-specific, nucleated event.[30] The bulk of the genome-regulated assembly occurs during assembly of "half-capsids",[14] i.e. in the initial phase of assembly, explaining why our VLPs form having packaged only 7 copies of the PSs1-5 sequence not the ~10 copies that one might have expected. Experiments that overwhelm this assembly mechanism, e.g. that occur at high CP concentrations, even *in vivo*,[48] fail to reveal this implied specificity.

The data described above explain most of what is known about the biology of STNV-1, and presumably many other plant satellite viruses. An assembled STNV-1 particle will enter a cell following an insect bite. The cytoplasm of such cells will naturally have low free-calcium ion concentrations, destabilising the protein shell of STNV-1, which is replete in chelated calcium ions (Figure 3).[4,41] Genome release will then likely occur in a staged way starting with the 3´ end of the gRNA. The fact that, as we show here, PSs in this part of the gRNA only make sparse contacts with CP subunits (Figure 4, Table 1), supports the view that the PS distribution is calibrated to support extrusion of the 3´ end of the gRNA. This is important for function as this substrate contains the binding site for the helper virus TNV RNA replicase,[49,50] leading to rapid disassembly. Virion expansion is a common phenomenon in ssRNA plant viruses[51–53] can be thought of a preliminary step to uncoating. *In vitro*, and probably *in vivo*, it is driven by loss of chelated calcium ions. It is possible that the lack of PS-CP contacts around the 3´ end of the gRNA further destabilises the protein shell sufficiently to trigger this event in this area of the capsid first, consistent with the expectation that such expansion events occur as a cascade,[54] promoting gRNA release at a defined locus in the capsid. In some viral systems, this might also mark the locus at which the capsid starts to disassemble, releasing the gRNA as a consequence. Late in infection, nascent CPs will bind replicated PSs 1–5 and thus nucleate the formation of novel virions, but only on the nascent positive-sense form of the gRNA.

In summary, the gRNA orchestrates the virion "life-cycle" from infection to the production of progeny virions. Furthermore, the CP architecture of the STNV-1 virion studied here is shared by many other plant virus CPs, suggesting that these aspects of their lifecycles, and in particular their assembly and genome release mechanisms are likely similar.

Materials & Methods

Assembly competition assays were carried out using recombinant STNV-1 CP, as described,[16,17,26,28,29] and *in vitro* transcribed RNA fragments were made by *in vitro* transcription of sequenced, commercial dsDNA templates (Figure 1).

VLPs for cryo-EM analysis were assembled in vitro around both RNAs using standard protocols, treated with RNase to remove un- or misassembled species and then purified over 15–45 % (w/v) sucrose density gradients.

Assembly substrates for competition assays were fluorescently-labelled using N-hydroxy-succinimide esters coupled to either Alexa Fluor-488 or Alexa Fluor-594, following transcription in the presence of 5´-amino-GMP which adds the fluorophore to the 5´-end of the transcript.

STNV VLP dissociation and recombinant CP purification

Recombinant STNV VLPs were purified from *E. coli*.[26] CP monomers were purified by disassembly in 500 mM Tris (pH 8.5), 100 mM EDTA, in the presence of 20 M Pepstatin A (Roche). STNV-1 CP was separated from nucleic acid by sequential Q-Sepharose and SP-Sepharose columns (Cytiva). STNV CP was eluted from the SP-Sepharose column using a 0.025–2 M NaCl gradient. Fractions were analysed by SDS-PAGE, and purified CP was dialysed into STNV reassembly buffer (50 mM HEPES, 3 mM CaCl2, pH7.5) at 4 ˚C and quantified using a Nanodrop (Thermofisher).

Preparation of genomic length templates

gRNA, C4-gRNA and AXXA templates were gifted by Dr. Nikesh Patel.[17] GBlocks coding for the other variants were cloned into the *Bam* HI and *Hin*dIII sites of a linearised pACYC184 vector using T4 DNA ligase (NEB) as per manufacturer's instructions. Plasmids were transformed into DH5 cells (NEB). Transcriptions were carried out using a Hiscribe T7 High yield RNA synthesis kit (NEB). Those requiring a 5´ fluorophore were labelled with an amino GMP during transcription and cross-linked to an Alexa Fluor 488 SDP ester (Invitrogen),

as described previously. Product integrity was checked using denaturing RNA gels.[16,55]

VLP reassembly reactions

(*Non-Competitive reactions*): The concentration of gRNAs was adjusted to 2 nM per RNA in a total volume of 18 mL 1X STNV reassembly buffer (50 mM HEPES, 3 mM $CaCl_2$ pH 7.5). VLP reassembly reactions were performed using a liquid-handling robot (Biomek 4000) in a low protein-binding 96-well plate.[29] The final volume of each well was 200 L after six STNV CP titration steps of 40 nM, yielding a final concentration of 240 nM CP, i.e. a 2-fold stoichiometric excess for a VLP. *(Competitive Reactions):* were as the non-competitive reactions but using titration steps of 20 nM CP monomer to a final concentration of 120 nM.

(Single-step CP addition): was accomplished as in the non-competitive reactions but 200 L of 120 nM STNV-1 CP solution was added in a single titration step. *(Excess reassembly titrations):* were carried out as for the non-competitive reactions but with 9x STNV CP titration steps of 40 nM, yielding a final concentration of 360 nM, i.e. a 3-fold excess with respect to the expected $T = 1$ STNV VLP.

After each addition of CP the reactions were mixed using the "pipette mixing" command. All VLP samples were concentrated to 500 L using *Vivaspin20* columns spun at 7500 xg in an FX6100 rotor.

Fractionation and Characterisation of VLPs

Concentrated reassembly reactions were loaded onto a pre-equilibrated (STNV reassembly buffer + 200 mM NaCl) TSK column at a flow rate of 0.35 mL/min on an AKTA pure with an associated DAWN8+, optilab T-rEX and QELS (SEC-MALLS, Wyatt technology). Absorbance of samples was measured at both 260 nm and 280 nm using unicorn software (version 7) and light scattering was measured and analysed using Astra software version 6.1.6.5. 1 mL fractions were collected and those with absorbance at 260 and 280 nm were concentrated using centrifugal filters with a 100 kDa MWCO (Millipore). Fluorescence of concentrated fractions was measured using a nanodrop 3000 (Thermofisher) at wavelengths of 510 and 617 nm. These were then normalised using the quantum yield of the corresponding fluorophores, Alexa Fluor 488 (0.92) and Alexa Fluor 594 (0.66).

Incubation with RNase and SEC-MALLS

Concentrated fractions were merged and incubated with 3 L RNase A (10 mg/mL) at 4 ˚C overnight and loaded onto a pre-equilibrated equilibrated (STNV reassembly buffer + 200 mM NaCl) TSK column as above.

PS folding analysis

Previously, we showed that the folding propensity of PSs1-5 regulated assembly of STNV VLPs by simply ensuring that the stems of each PS were fully Watson-Crick complementary and by changing A.U base pairs to G.C. A more sophisticated approach was used to analyse the folding propensity of all 30 putative PS sites in the STNV-1 gRNA. The STNV-1 gRNA sequence (Sup Figure 1) was folded locally using SFold[56] within 50 nucleotide (nt) windows, slid across the entire gRNA in increments of 5 nts from the 5´ end, or globally across the entire gRNA. 1000 sample folds were determined for both scenarios. Stem-loops (SL) with free energies of 2 kcal/mol or less, and which present an A.X.X.(X).A loop motif, were identified as putative PSs (Sup Figure 1). For each such putative PS, the frequency, the mean folding energy (with standard deviation on the mean) and the minimum energy of SL formation were calculated. Moving averages of these frequencies were then calculated using a sliding window of 100 nts slid in increments of 5 nts. For each window, the average frequency was taken from all the PSs within the window.

Structural Methods

Cryo-EM Structure Determinations of STNV VLPs. Purified STNV samples (3 L) were applied onto 400 mesh lacey grids coated in a 3 nm carbon film (Agar Scientific, UK). The sample was left to adsorb for 30 s before most of the sample was blotted away manually and this process was repeated 4 times. The grids were vitrified using a Vitrobot Mark IV freezing device (ThermoFisher). Chamber conditions were set at 4 ˚C and 95 % relative humidity. Grids were glow discharged for 30 sec. prior to application of the samples. PS1-5 and 2/3-gRNA data were collected on a FEI Titan Krios electron microscope operated at 300 kV and images recorded on a Gatan K2 detector operating in counting mode and on a Falcon III detector operating in linear mode, respectively. A total of 11,550 movies (STNV with PS1-5 gRNA) and 9,979 were recorded at a pixel size of 1.065 Å on the specimen. Each movie comprises 39 frames with a dose rate of 1.3 e^- $Å^2$ per frame, with a total exposure time of 1.3 s and an accumulated dose of 47.6 e^- $Å^2$. Data acquisition was performed with EPU Automated Data Acquisition Software for Single Particle Analysis (ThermoFisher) at 0.8 to 3.5 m (STNV with PS1-5 gRNA) and 1.0 to 3.75 m (STNV with 2/3-gRNA) defocus.

Image processing

Movies were motion-corrected and dose-weighted with MOTIONCOR2.[57] Aligned, dose-weighted micrographs were then used to estimate

the contrast transfer function (CTF) with GCTF.[58] All subsequent image processing steps were performed using RELION 3.[59,60] Particles were picked using the 'Autopick' programme in RELION (Log-based gaussian blob picker option). A total of 502,684 particles (PS1-5 dataset) and 509,873 (2/3-gRNA dataset) were picked. Particles were binned by 2 for initial rounds of 2D classification to speed-up calculations. 2D classification was performed from which STNV-VLP with RNA were selected. In total, 206,058 particles (PS1-5) and 115,419 particles (2/3-gRNA) were selected. The particle stacks were imported to cryosparc v3.0[61] allowing us to perform 'ab-initio 3D' map generation using one map as the output, giving C1 as the symmetry operator. The output maps were downloaded and used in RELION to perform 3D refinement, using selected particles that were extracted in full box size (252 x 252 pixels). The particles were refined against the initial maps using icosahedral symmetry, giving output maps at resolutions 2.8 Å (PS1-5 dataset) and 4.2 Å (2/3-gRNA) respectively. The refined particles were then separated based on image quality and refinement statistics using 3D classification with no alignment, using a STNV-1 capsid mask (soft padded to 10 pixels), sorted into five classes per dataset. From each dataset, particles belonging to highest resolution 3D map were selected; 78,988 particles (PS1-5 dataset) and 81,619 (2/3-gRNA dataset) respectively. Subsequent Icosahedral refinement of selected particles in each dataset resulted in refined maps at 2.6 Å (PS1-5) and 3.3 Å (2/3-gRNA) respectively. To improve the reconstructions, an iterative round of CTF refinements were done (Magnification anisotropy, per-particle defocus, beam tilt, per-particle defocus), the output particles from which were then imported into cryosparc to perform asymmetric and symmetric 3D refinements. Using non-uniform 3D refinement in cryosparc (with global CTF refinement option switched on), the maps refined to 2.39/2.92 Å (Icosahedral/C1; PS1-5 dataset) and 3.7 Å/4.56 Å (Icosahedral/C1; 2/3 gRNA dataset) respectively. For asymmetric 3D reconstruction, the refined icosahedral map of STNV-1 (from both datasets) was low-pass filtered to 30 Å to remove structural bias.

To visualize the ssRNA genome density inside the capsid structures, a 3D mask was generated by subtracting the capsid only 3D map from the RNA-filled capsid maps, giving only density for RNA genome packaged inside. The 3D mask was then applied to the PS1-5 gRNA and 2/3 gRNA maps, and were subsequently low-pass filtered to 6 Å. The capsid maps from icosahedral refinement were post-processed and sharpened using a 3D capsid mask to generate sharpened maps for capsid only model building.

Model building and refinement. The crystal structure of STNV (PDB:4BCU,[18] was manually docked as a rigid body into the density, followed by rigid-body docking with the 'Fit in Map' routine in UCSF Chimera.[62] The coordinates for a single monomer were saved and imported to Phenix[63] to run real space refinement against the individual cryo-EM maps using global minimization, morphing, simulated annealing, local grid search, atomic displacement parameter (ADP), secondary structure restraints, non-crystallographic symmetry (NCS) restraints, side chain rotamer restraints, and Ramachandran restraints. The output model was then manually inspected and rebuilt in Coot[64] to optimize the fit to the density. After icosahedral symmetrisation in Chimera to generate the entire capsid, a second step of real space refinement was performed in Phenix without morphing nor simulated annealing. Refinement statistics are listed in Sup. Table 2. For modelling RNA fragments, the masked RNA map was used to generate a single 12-nucleotide fragment, with nucleotides corresponding to the 'PS3 gRNA loop' using Coot. This fragment was then duplicated 60 times to model the gRNA densities across the capsid using rigid body docking in Chimera. After initial assignments of fragments to the densities, 19 fragments were assigned to PS1-5 density map and 17 fragments were assigned to 2/3-gRNA maps respectively. The position and orientation of the RNA fragments (across the capsid) based on the rigid-body fitting analysis were used for analysis and interpretation.

The quality of the atomic model, including basic protein geometry, Ramachandran plots, clash analysis, was assessed and validated with Coot, MolProbity[65] as implemented in Phenix, and with the Worldwide PDB (wwPDB) OneDep System (https://deposit-pdbe.wwpdb.org/deposition). Graphics were produced by UCSF Chimera[62,66].

CRediT authorship contribution statement

Emma Wroblewski: Investigation. **Nikesh Patel:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis. **Abid Javed:** Visualization, Formal analysis, Investigation, Writing – review & editing. **Carlos P. Mata:** Formal analysis. **Rebecca Chandler-Bostock:** Investigation. **Lekshmi B.G. Nair:** Formal analysis. **Sabine M. Ulamec:** Investigation. **Samuel Clark:** Investigation, Formal analysis. **Simon E.V. Phillips:** Writing – review & editing. **Neil A. Ranson:** Writing – review & editing. **Reidun Twarock:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Peter G. Stockley:** Writing – review & editing, Writing – original draft, Conceptualization.

DATA AVAILABILITY

Data will be made available on request.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank our colleagues for helpful discussions of these data. This work was supported initially by grants to PGS and RT from the UK BBSRC (BB/J00667X/1, BB/L022095/1 & BB/W017644/1), & from The Wellcome Trust (Joint Investigator Award Nos. 110145/10146 & 215062/Z/18/Z/215062/A/18/Z). RT also acknowledges funding via an EPSRC Established Career Fellowship (EP/R023204/1) and a Royal Society Wolfson Fellowship (RSWF/R1/180009). NP also acknowledges funding from the Medical Research Foundation and the UK BBSRC (MRF-044-0002-RG-PATEL & BB/W017644/1). SMU was funded by The Wellcome Trust (215062/Z/18/Z).

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.jmb.2024.168765.

Received 1 August 2024;
Accepted 23 August 2024;
Available online 29 August 2024

Keywords:
virus assembly;
genome regulated virus assembly and infection;
packaging signals;
molecular frustration

† These authors contributed equally to the manuscript.
‡ Current address: CPM: Biocomputing Unit, Department of Macromolecular Structures, National Centre for Biotechnology (CNB-CSIC), Darwin, 3 28049, Madrid, Spain.
§ Current address: Diamond Light Source Ltd., Diamond House, Harwell Science & Innovation Campus, Didcot, Oxfordshire, OX11 0DE, United Kingdom.
‖ Current address: Coriolis Pharma Research GmbH, Fraunhoferstr. 18 b, 82152 Martinsried, Germany.
¶ Current address: South View, Park Road, Combe, OX29 8NA, United Kingdom.

References

1. Cervia-Hasler, C., Brüningk, S.C., Hoch, T., Fan, B., Muzio, G., Thompson, R.C., Ceglarek, L., Meledin, R., Westermann, P., Emmenegger, M., Taeschler, P., Zurbuchen, Y., Pons, M., Menges, D., Ballouz, T., Cervia-Hasler, S., Adamo, S., Merad, M., Charney, A.W., Puhan, M., Brodin, P., Nilsson, J., Aguzzi, A., Raeber, M.E., Messner, Beckmann, N.D., Borgwardt, K., Boyman, O., (2024). Persistent complement dysregulation with signs of thromboinflammation in active Long Covid. *Science* **383**, 1–18. https://doi.org/10.1126/SCIENCE.ADG7942.
2. Legg, J.P., Owor, B., Sseruwagi, P., Ndunguru, J., (2006). Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Adv. Virus Res.* **67**, 355–418. https://doi.org/10.1016/S0065-3527(06)67010-3.
3. Henriksson, D., Tanis, R.J., Tashian, R.E., Nyman, P.O., (1981). Amino acid sequence of the coat protein subunit in satellite tobacco necrosis virus. *J. Mol. Biol.* **152**, 171–179. https://doi.org/10.1016/0022-2836(81)90101-7.
4. Liljas, L., Unge, T., Jones, T.A., Fridborg, K., Lövgren, S., Skoglund, U., Strandberg, B., (1982). Structure of Satellite tobacco necrosis virus at 3.0 Å resolution. *J. Mol. Biol.* **159**, 93–108. https://doi.org/10.1016/0022-2836(82)90033-X.
5. Ban, N., Larson, S.B., McPherson, A., (1995). Structural comparison of the plant satellite viruses. *Virology* **214**, 571–583. https://doi.org/10.1006/VIRO.1995.0068.
6. Larson, S.B., McPherson, A., (2001). Satellite tobacco mosaic virus RNA: Structure and implications for assembly. *Curr. Opin. Struct. Biol.* **11**, 59–65. https://doi.org/10.1016/S0959-440X(00)00166-4.
7. Casjens, S., King, J., (1975). Virus assembly. *Annu. Rev. Biochem* **44**, 555–611. https://doi.org/10.1146/ANNUREV.BI.44.070175.003011/CITE/REFWORKS.
8. Harrison, S.C., (1980). Protein interfaces and intersubunit bonding. The case of tomato bushy stunt virus. *Biophys. J .* **32**, 139. https://doi.org/10.1016/S0006-3495(80)84930-7.
9. Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., Bricogne, G., (1978). Tomato bushy stunt virus at 2.9 A resolution. *Nature* **276**, 368–373. https://doi.org/10.1038/276368A0.
10. Abad-zapatero, C., Abdel-meguid, S.S., Johnson, J.E., Leslie, A.G.W., Rayment, I., Rossmann, M.G., Suck, D., Tsukihara, T., (1980). Structure of southern bean mosaic virus at 2.8 A resolution. *Nature* **286**, 33–39. https://doi.org/10.1038/286033A0.
11. Hogle, J.M., Chow, M., Filman, D.J., (1979). Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **229** (1985), 1358–1365. https://doi.org/10.1126/SCIENCE.2994218.
12. Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D., Brown, F., (1989). The three-dimensional structure of foot-and-mouth disease virus at 2.9 Å resolution. *Nature* **337**, 709–716. https://doi.org/10.1038/337709a0.
13. Rossmann, M.G., Arnold, E., Erickson, J.W., Frankenberger, E.A., Griffith, J.P., Hecht, H.J., Johnson, J.E., Kamer, G., Luo, M., Mosser, A.G., Rueckert, R.R., Sherry, B., Vriend, G., (1985). Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* **317**, 145–153. https://doi.org/10.1038/317145A0.
14. Borodavka, A., Tuma, R., Stockley, P.G., (2012). Evidence that viral RNAs have evolved for efficient, two-stage packaging. *PNAS* **109**, 15769–15774. https://doi.org/10.1073/pnas.1204357109.
15. Twarock, R., Stockley, P.G., (2019). RNA-mediated virus assembly: Mechanisms and consequences for viral evolution and therapy. *Annu. Rev. Biophys.* **48**, 495–514. https://doi.org/10.1146/annurev-biophys-052118-115611.
16. Patel, N., Dykeman, E.C., Coutts, R.H.A., Lomonossoff, G. P., Rowlands, D.J., Phillips, S.E.V., Ranson, N., Twarock,

R., Tuma, R., Stockley, P.G., Johnson, J.E., (2015). Revealing the density of encoded functions in a viral RNA. *PNAS* **112**, 2227–2232. https://doi.org/10.1073/pnas.1420812112.
17. Patel, N., Wroblewski, E., Leonov, G., Phillips, S.E.V., Tuma, R., Twarock, R., Stockley, P.G., (2017). Rewriting nature's assembly manual for a ssRNA virus. *PNAS* **114**, 12255–12260. https://doi.org/10.1073/pnas.1706951114.
18. Ford, R.J., Barker, A.M., Bakker, S.E., Coutts, R.H., Ranson, N.A., Phillips, S.E.V., Pearson, A.R., Stockley, P.G., (2013). Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *J. Mol. Biol.* **425**, 1050–1064. https://doi.org/10.1016/j.jmb.2013.01.004.
19. Chandler-Bostock, R., Mata, C.P., Bingham, R.J., Dykeman, E.C., Meng, B., Tuthill, T.J., Rowlands, D.J., Ranson, N.A., Twarock, R., Stockley, P.G., (2020). Assembly of infectious enteroviruses depends on multiple, conserved genomic RNA-coat protein contacts. *PLoS Pathog.* **16** https://doi.org/10.1371/journal.ppat.1009146.
20. Chandler-Bostock, R., Bingham, R.J., Clark, S., Scott, A.J. P., Wroblewski, E., Barker, A., White, S.J., Dykeman, E.C., Mata, C.P., Bohon, J., Farquhar, E., Twarock, R., Stockley, P.G., (2022). Genome-regulated assembly of a ssRNA virus may also prepare it for infection. *J. Mol. Biol.* **434** https://doi.org/10.1016/J.JMB.2022.167797.
21. Gorzelnik, K.V., Cui, Z., Reed, C.A., Jakana, J., Young, R., Zhang, J., (2016). Asymmetric cryo-EM structure of the canonical Allolevivirus Q reveals a single maturation protein and the genomic ssRNA in situ. *PNAS* **113**, 11519–11524. https://doi.org/10.1073/PNAS.1609482113.
22. Koning, R.I., Gomez-Blanco, J., Akopjana, I., Vargas, J., Kazaks, A., Tars, K., Carazo, J.M., Koster, A.J., (2016). Asymmetric cryo-EM reconstruction of phage MS2 reveals genome structure in situ. *Nature Commun.* **7** https://doi.org/10.1038/ncomms12524.
23. Dai, X., Li, Z., Lai, M., Shu, S., Du, Y., Zhou, Z.H., Sun, R., (2017). In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**, 112–116. https://doi.org/10.1038/nature20589.
24. Meng, R., Xing, Z., Chang, J.Y., Yu, Z., Thongchol, J., Xiao, W., Wang, Y., Chamakura, K., Zeng, Z., Wang, F., Young, R., Zeng, L., Zhang, J., (2024). Structural basis of Acinetobacter type IV pili targeting by an RNA virus. *Nature Commun.* **15**, 1–9. https://doi.org/10.1038/s41467-024-47119-5.
25. Bunka, D.H.J., Lane, S.W., Lane, C.L., Dykeman, E.C., Ford, R.J., Barker, A.M., Twarock, R., Phillips, S.E.V., Stockley, P.G., (2011). Degenerate RNA packaging signals in the genome of satellite tobacco necrosis virus: Implications for the assembly of a T = 1 capsid. *J. Mol. Biol.* **413**, 51–65. https://doi.org/10.1016/j.jmb.2011.07.063.
26. Lane, S.W., Dennis, C.A., Lane, C.L., Trinh, C.H., Rizkallah, P.J., Stockley, P.G., Phillips, S.E.V., (2011). Construction and crystal structure of recombinant STNV capsids. *J. Mol. Biol.* **413**, 41–50. https://doi.org/10.1016/j.jmb.2011.07.062.
27. Zandi, R., Van Der Schoot, P., Reguera, D., Kegel, W., Reiss, H., (2006). Classical nucleation theory of virus capsids. *Biophys. J.* **90**, 1939–1948. https://doi.org/10.1529/BIOPHYSJ.105.072975.
28. Patel, N., Abulwerdi, F., Fatehi, F., Manfield, I.W., Le Grice, S., Schneekloth Jr., J.S., Twarock, R., Stockley, P.G., (2022). Dysregulation of hepatitis B virus nucleocapsid assembly in vitro by RNA-binding small ligands. *J. Mol. Biol.* **434**, https://doi.org/10.1016/j.jmb.2022.167557 167557.
29. Patel, N., Clark, S., Weiß, E.U., Mata, C.P., Bohon, J., Farquhar, E.R., Maskell, D.P., Ranson, N.A., Twarock, R., Stockley, P.G., (2021). In vitro functional analysis of gRNA sites regulating assembly of hepatitis B virus. *Commun Biol* **4**, 1–12. https://doi.org/10.1038/s42003-021-02897-2.
30. Wroblewski, E.L., (2018). Defining RNA Packaging Signals for Virus Assembly. University of Leeds https://etheses.whiterose.ac.uk/23243/.
31. Shakeel, S., Westerhuis, B.M., Domanska, A., Koning, R.I., Matadeen, R., Koster, A.J., Bakker, A.Q., Beaumont, T., Wolthers, K.C., Butcher, S.J., (2016). Multiple capsid-stabilizing interactions revealed in a high-resolution structure of an emerging picornavirus causing neonatal sepsis. *Nature Commun.* **7**, 1–8. https://doi.org/10.1038/ncomms11387.
32. Patel, N., White, S.J., Thompson, R.F., Bingham, R., Weiß, E.U., Maskell, D.P., Zlotnick, A., Dykeman, E.C., Tuma, R., Twarock, R., Ranson, N.A., Stockley, P.G., (2017). HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nature Microbiol.* **2**, 17098. https://doi.org/10.1038/nmicrobiol.2017.98.
33. Stockley, P.G., White, S.J., Dykeman, E., Manfield, I., Rolfsson, O., Patel, N., Bingham, R., Barker, A., Wroblewski, E., Chandler-Bostock, R., Weiß, E.U., Ranson, N.A., Tuma, R., Twarock, R., (2016). Bacteriophage MS2 genomic RNA encodes an assembly instruction manual for its capsid. *Bacteriophage* **6**, e1157666.
34. Dykeman, E.C., Stockley, P.G., Twarock, R., (2014). Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *PNAS* **111**, 5361–5366. https://doi.org/10.1073/pnas.1319479111.
35. Dykeman, E.C., Stockley, P.G., Twarock, R., (2013). Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *J. Mol. Biol.* **425**, 3235–3249. https://doi.org/10.1016/j.jmb.2013.06.005.
36. Shakeel, S., Dykeman, E.C., White, S.J., Ora, A., Cockburn, J.J.B., Butcher, S.J., Stockley, P.G., Twarock, R., (2017). Genomic RNA folding mediates assembly of human parechovirus. *Nature Commun.* **8**, 5. https://doi.org/10.1038/s41467-016-0011-z.
37. Rolfsson, Ó., Middleton, S., Manfield, I.W., White, S.J., Fan, B., Vaughan, R., Ranson, N.A., Dykeman, E., Twarock, R., Ford, J., Cheng Kao, C., Stockley, P.G., (2016). Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. *J. Mol. Biol.* **428**, 431–448. https://doi.org/10.1016/j.jmb.2015.11.014.
38. Stockley, P.G., Twarock, R., Bakker, S.E., Barker, A.M., Borodavka, A., Dykeman, E., Ford, R.J., Pearson, A.R., Phillips, S.E.V., Ranson, N.A., Tuma, R., (2013). Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism. *J. Biol. Phys.* **39**, 277–287. https://doi.org/10.1007/s10867-013-9313-0.
39. Zandi, R., Dragnea, B., Travesset, A., Podgornik, R., (2020). On virus growth and form. *Phys. Rep.* **847**, 1–102. https://doi.org/10.1016/J.PHYSREP.2019.12.005.

40. Van Lipzig, R., Gultyaev, A.P., Pleij, C.W.A., Van Montagu, M., Cornelissen, M., Meulewaeter, F., (2002). The 5´ and 3´ extremities of the satellite tobacco necrosis virus translational enhancer domain contribute differentially to stimulation of translation. *RNA* **8**, 229–236. https://doi.org/10.1017/S1355838202018071.
41. Lentz, P.J., Strandberg, B., (1974). Interpretation of the 10 Å rotation function of the satellite tobacco necrosis virus. *Acta Crystallogr. Sect. A* **30**, 552–559. https://doi.org/10.1107/S0567739474001318.
42. Twarock, R., Towers, G.J., Stockley, P.G., (2024). Molecular frustration: a hypothesis for regulation of viral infections. *Trends Microbiol.* **32**, 17–26. https://doi.org/10.1016/J.TIM.2023.07.003.
43. Van Der Schoot, P., Bruinsma, R., (2005). Electrostatics and the assembly of an RNA virus. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **71**, 61928. https://doi.org/10.1103/PhysRevE.71.061928.
44. Belyi, V.A., Muthukumar, M., (2006). Electrostatic origin of the genome packing in viruses. *PNAS* **103**, 17174–17178. https://doi.org/10.1073/pnas.0608311103.
45. Garmann, R.F., Comas-Garcia, M., Koay, M.S.T., Cornelissen, J.J.L.M., Knobler, C.M., Gelbart, W.M., (2014). Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *J. Virol.* **88**, 10472–10479. https://doi.org/10.1128/jvi.01044-14.
46. Brown, R.S., Anastasakis, D.G., Hafner, M., Kielian, M., (2020). Multiple capsid protein binding sites mediate selective packaging of the alphavirus genomic RNA. *Nature Commun.* **11**, 1–16. https://doi.org/10.1038/s41467-020-18447-z.
47. Tetter, S., Terasaka, N., Steinauer, A., Bingham, R.J., Clark, S., Scott, A.J.P., Patel, N., Leibundgut, M., Wroblewski, E., Ban, N., Stockley, P.G., Twarock, R., Hilvert, D., (1979). Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science* **372** (2021), 1220–1224. https://doi.org/10.1126/science.abg2822.
48. Kotta-Loizou, I., Peyret, H., Saunders, K., Coutts, R.H.A., Lomonossoff, G.P., (2019). Investigating the biological relevance of in vitro-identified putative packaging signals at the 5 = terminus of satellite tobacco necrosis virus 1 genomic RNA. *J. Virol.* **93**, https://doi.org/10.1073/pnas e02106-18.
49. Bringloe, D.H., Pleij, C.W.A., Coutts, R.H.A., (1999). Mutation analysis of cis-elements in the 3´- and 5´-untranslated regions of satellite tobacco necrosis virus strain C RNA. *Virology* **264**, 76–84. https://doi.org/10.1006/viro.1999.9994.
50. Bringloe, D.H., Gultyaev, A.P., Pelpel, M., Pleij, C.W.A., Coutts, R.H.A., (1998). The nucleotide sequence of satellite tobacco necrosis virus strain C and helper-assisted replication of wild-type and mutant clones of the virus. *J. Gen. Virol.* **79**, 1539–1546. https://doi.org/10.1099/0022-1317-79-6-1539.
51. Larsson, D.S.D., Liljas, L., van der Spoel, D., (2012). Virus capsid dissolution studied by microsecond molecular dynamics simulations. *PLoS Comput. Biol.* **8**, e1002502.
52. Robinson, I.K., Harrison, S.C., (1982). Structure of the expanded state of tomato bushy stunt virus. *Nature* **297**, 563–568. https://doi.org/10.1038/297563a0.
53. Unge, T., Montelius, I., Liljas, L., Öfverstedt, L.G., (1986). The EDTA-treated expanded satellite tobacco necrosis virus: Biochemical properties and crystallization. *Virology* **152**, 207–218. https://doi.org/10.1016/0042-6822(86)90385-5.
54. Indelicato, G., Cermelli, P., Twarock, R., (2019). A coarse-grained model of the expansion of the human rhinovirus 2 capsid reveals insights in genome release. *J. R. Soc. Interface* **16** https://doi.org/10.1098/RSIF.2019.0044.
55. Borodavka, A., Tuma, R., Stockley, P.G., (2013). A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol.* **10**, 481–489. https://doi.org/10.4161/rna.23838.
56. Ding, Y., Chan, C.Y., Lawrence, C.E., (2004). Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* **32**, W135. https://doi.org/10.1093/nar/gkh449.
57. Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., Agard, D.A., (2017). MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**, 331–332. https://doi.org/10.1038/nmeth.4193.
58. Zhang, K., (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12. https://doi.org/10.1016/j.jsb.2015.11.003.
59. Scheres, S.H.W., (2016). Processing of structurally heterogeneous Cryo-EM data in RELION. *Methods Enzymol.*, 125–157. https://doi.org/10.1016/bs.mie.2016.04.012.
60. Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J.H., Lindahl, E., Scheres, S.H.W., (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7** https://doi.org/10.7554/eLife.42166.
61. Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **14**, 290–296. https://doi.org/10.1038/nmeth.4169.
62. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., (2004). UCSF Chimera – A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. https://doi.org/10.1002/jcc.20084.
63. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221. https://doi.org/10.1107/S0907444909052925.
64. Emsley, P., Cowtan, K., (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132. https://doi.org/10.1107/S0907444904019158.
65. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21. https://doi.org/10.1107/S0907444909042073.
66. Tomasello, G., Armenia, I., Molla, G., (2020). The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics* **36**, 2909–2911. https://doi.org/10.1093/BIOINFORMATICS/BTAA009.