Revealing the density of encoded functions in a viral RNA

Nikesh Patel[a], Eric C. Dykeman[b,c], Robert H. A. Coutts[d,1], George P. Lomonossoff[e], David J. Rowlands[a], Simon E. V. Phillips[f], Neil Ranson[a], Reidun Twarock[b,c], Roman Tuma[a], and Peter G. Stockley[a,2]

[a]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom; [b]York Centre for Complex Systems Analysis, and [c]Departments of Mathematics and Biology, University of York, York YO10 5DD, United Kingdom; [d]Division of Biology, Imperial College London, London SW7 2AZ, United Kingdom; [e]Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norfolk NR4 7UH, United Kingdom; and [f]Research Complex at Harwell, Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxon OX11 0FA, United Kingdom

Edited by John E. Johnson, The Scripps Research Institute, La Jolla, CA, and accepted by the Editorial Board January 5, 2015 (received for review October 30, 2014)

We present direct experimental evidence that assembly of a single-stranded RNA virus occurs via a packaging signal-mediated mechanism. We show that the sequences of coat protein recognition motifs within multiple, dispersed, putative RNA packaging signals, as well as their relative spacing within a genomic fragment, act collectively to influence the fidelity and yield of capsid self-assembly in vitro. These experiments confirm that the selective advantages for viral yield and encapsidation specificity, predicted from previous modeling of packaging signal-mediated assembly, are found in Nature. Regions of the genome that act as packaging signals also function in translational and transcriptional enhancement, as well as directly coding for the coat protein, highlighting the density of encoded functions within the viral RNA. Assembly and gene expression are therefore direct molecular competitors for different functional folds of the same RNA sequence. The strongest packaging signal in the test fragment, encodes a region of the coat protein that undergoes a conformational change upon contact with packaging signals. A similar phenomenon occurs in other RNA viruses for which packaging signals are known. These contacts hint at an even deeper density of encoded functions in viral RNA, which if confirmed, would have profound consequences for the evolution of this class of pathogens.

virus assembly | single-molecule fluorescence correlation spectroscopy | satellite tobacco necrosis virus | packaging signal

Almost all natural nucleic acid sequences appear in the form of RNA polynucleotides at some point in their existence. The concept of "one RNA, one function" (e.g., for tRNA or mRNA) is increasingly seen as far too simplistic with RNAs known to alter their secondary/tertiary structures in regulated ways to control the appearance of distinct functional states (1). This multifunctionality is also true of the genomes of single-stranded (ss) RNA viruses, where refolding to promote replication, translation and control of gene expression are well known (2). Here we describe additional layers of previously unsuspected functions within the genome of the satellite plant virus, Satellite Tobacco Necrosis Virus (STNV). Aspects of these functions are potential drug targets for this class of viruses, which include major human pathogens and a significant fraction of emerging viral diseases (3, 4). Virus-like particles (VLPs) are also increasingly being used both for vaccines and for delivery of imaging, diagnostic, or therapeutic agents (5, 6). Thus, understanding the detailed molecular mechanisms used by these viruses for assembly is of widespread interest.

Previously we showed, using single molecule fluorescence correlation spectroscopy (smFCS), that at low (nanomolar) concentrations the coat proteins (CPs) of STNV and bacteriophage MS2 package their cognate genomes preferentially (7). Cognate packaging in vitro is accompanied by a characteristic drop in the hydrodynamic radius (Rh) of the viral RNA, making it smaller than its capsid as a subset of CPs bind. We have shown that this behavior is consistent with an assembly mechanism (8) controlled by the interaction of multiple CPs with degenerate RNA sequence/structure sites we term packaging signals (PSs), distributed across the genomes. PSs are bound sequence-specifically by their cognate CPs, increasing the yield, rate and/or fidelity of assembly. PS-mediated assembly has many potential advantages for the virus, including solving a viral version of the Levinthal Paradox (9). For the $T = 3$ MS2, there is a known assembly origin, the 19-nt-long coat protein gene translational operator (TR) stem-loop (10). TR functions as an allosteric effector, switching the dimeric viral capsomere between the two quasi-conformers required to build the capsid (11). This mechanism requires 60 TR-like sites within the genomic RNA for CP contacts (12). These defined RNA–CP contacts require the RNA in every particle be in a very similar conformation, and an asymmetric tomographic reconstruction is consistent with this view (13).

For STNV, there were no known PSs, so we used RNA SELEX combined with a theoretical analysis protocol we designed for the identification of PSs from SELEX data, to isolate preferred sequences for CP binding. This dual approach identified an aptamer whose sequence matches a region of the cognate genome (14). Both the aptamer (B3) and genomic sites (B3-like, PS3, Fig. 1 *A* and *B*) are predicted to fold into stem-loops. An X-ray structure of VLPs reassembled around B3 shows that the presence of multiple copies of a preferred RNA sequence increases the ordering of

Significance

Single-stranded RNA viruses self-assemble protective protein containers around their cognate genomes rapidly and efficiently at low concentrations. RNA encapsidation in vivo occurs preferentially with the cognate genome, in contrast to many in vitro reassembly experiments. We describe in molecular detail how this specificity and efficiency is accomplished using multiple contacts between coat proteins and dispersed packaging signals in the viral genome. The sequences and relative positioning of the packaging signals are important for this mechanism, creating a strong evolutionary constraint. Packaging signals overlap untranslated and coding regions ensuring assembly is in competition with other functions of the genome. Disrupting these contacts has deleterious consequences for capsid assembly identifying a novel antiviral drug target.

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.E.J. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[1]Present address: Geography, Environment and Agriculture Division, Department of Human and Environmental Sciences, School of Life and Medical Sciences, University of Hertfordshire, Hatfield AL109AB, United Kingdom.

[2]To 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.1420812112/-/DCSupplemental.

the N-terminal region of the STNV CP. The N-terminal helix extends by roughly one turn to include amino acid residue 8, in contrast to the virion or mRNA containing VLP, in which the electron density shows that the polypeptide is disordered before residue 12. These data suggest that B3 acts like a PS (15) (Fig. 1).

SmFCS confirms that B3 binds STNV CP with high affinity (low nM) to form a capsomere, probably a CP trimer with at least one B3 bound, that can then be assembled into $T = 1$ capsids in response to increasing concentrations of CP. This reaction is highly sequence-specific, because oligos containing B3 variant in the recognition motif fail to assemble under identical conditions. Despite the catalytic role of B3, assembly is highly dependent on electrostatic interactions: The VLPs formed are sensitive to both the ionic strength and the removal of structural calcium ions. Using a 5′ genomic fragment (127 nt long) encompassing the B3-like PS3 and two additional putative PSs on either side, we have confirmed that CP-induced reduction in Rh occurs as a consequence of binding PS3. The five PSs work collectively to promote successful capsid assembly and their spacing within the fragment is also important. These functions are encoded within a region heavily implicated in enhancing translation and viral RNA replication (16–19). In addition, PS3 is located within the CP region that encodes the amino acid sequence undergoing conformational change in response to PS binding. If similar contacts occur more widely for RNA viruses there may be even deeper embedding of function in viral RNAs.

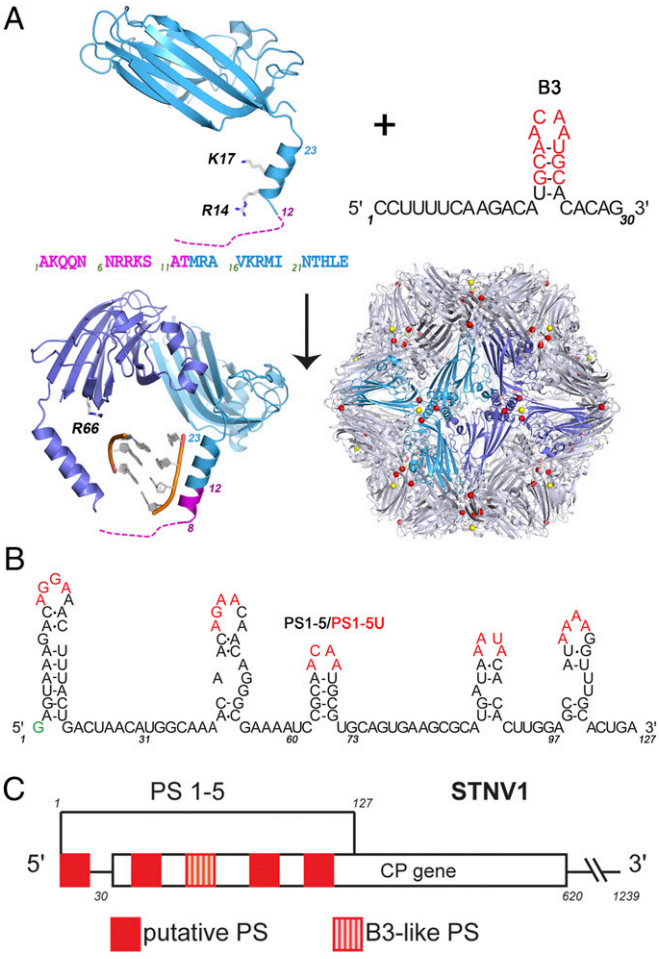

Fig. 1. The components of the STNV system. (*A*) (*Upper Left*) Ribbon diagram of the STNV CP subunit structure seen in virions [Protein Data Bank (PDB) ID 4V4M] with the disordered N-terminal amino acid sequence shown dashed, with the sequence of the first 25 amino acids below. When this assembles around the B3 aptamer, sequence and secondary structure shown in *Upper Right* (nucleotide identities with the STNV-1 genome in red), it forms a $T = 1$ capsid (*Lower Right*, PDB 3S4G), in which there is additional ordering of the CP N-terminal region (magenta helical turn) in response to binding the preferred RNA sequence. The virion is stabilized by multiple Ca^{2+} ions that bind around the differing symmetry axes (shown as color-coded dots). (*B*) The sequence and putative secondary structure of the 5′ genomic fragment encompassing five putative PSs, named in order 5′ to 3′ as PS1-5, respectively. Red nucleotides indicate the potential CP recognition motif (AXXA) and are the sites mutated to U to create variant fragments. The green G is a nonviral nucleotide added for ease of dye-labeling. (*C*) Map of the STNV1 genome showing the positions of these putative packaging signals (red) with respect to the CP gene.

Sequence-Specificity of Assembly

To study in vitro reassembly of STNV VLPs at low CP concentrations, we used fluorescently labeled oligonucleotides (see Fig. 1 *A* and *B* and Table S1 for details). This method enabled smFCS assembly assays at nanomolar RNA concentrations. Initially we monitored CP titrations into labeled B3 (1 nM) under reassembly conditions. This experiment reveals formation of a high affinity RNA–CP complex that begins to form at 1 nM CP and is fully formed by 5 nM CP. It has an apparent Rh of ∼5 nm (Fig. 2*B*), consistent with it being a CP trimeric capsomere based on the X-ray structure of the B3 VLP. Photon counting statistics (*SI Materials and Methods*) were used to estimate the affinity of the CP–B3 interaction, yielding a K_d of 0.8 ± 0.1 nM, and photon burst statistics suggest that the complex formed most likely contains a single B3 RNA (Fig. S1). Once formed, this complex is stable for extended periods (Fig. 2*C*) and indifferent to the addition of extra CP monomer until a threshold concentration ∼50 nM is reached. Thereafter assembly of $T = 1$ capsids occurs slowly in a CP concentration-dependent process that remains incomplete below 1 μM CP. These VLPs are also stable over extended periods, indifferent to the addition of more CP and protect the labeled RNA from RNase digestion (Fig. S2). This second phase of assembly must be driven by CP–CP interactions between the complex nucleated by the initial RNA–CP contact and incoming CPs. Note, at much higher concentrations (μM) the STNV CP is a monomer with no tendency to self-assemble in the absence of RNA (15). The slow kinetics of capsid assembly from the B3 capsomere suggests that the CPs must overcome a barrier to complex formation which we previously speculated was electrostatic, i.e., that the RNA reduces unfavorable interactions between the positively charged N-terminal regions of CP subunits (Fig. 1*A*).

To see whether the RNA effect is purely electrostatic or sequence-specific we set up an assembly titration with B34U, a sequence variant of B3 of the same length predicted to form a B3-like stem-loop, but with the CP recognition motif, -ACAA-, in the loop replaced by -UUUU-. RNA SELEX suggests that the minimal motif consists of a base-paired stem topped by a loop containing -AXXA-, where X is any nucleotide (14). The result (Fig. 2*A*, *Lower*) shows very clearly the importance of this sequence for CP recognition. At ∼twofold higher CP concentration than for B3, the variant shows an increased Rh suggesting formation of the capsomere. However, no further assembly occurs until the CP concentration reaches 3 μM, when there is a sharp increase in Rh followed by a steady decline, consistent with formation of unstable nonspecific aggregates rather than VLPs, and the RNA remains accessible to RNase (Fig. S2). The initial B34U-CP complex is not a viable on-pathway assembly intermediate despite the RNA being of the same length and therefore having the same negative charge. This result shows that capsomere assembly is sequence-specific and that only correctly formed species are intermediates in VLP formation. Because B34U RNA forms a complex with CP, sequence recognition is likely to involve binding of the base-paired stem, which is then orientated correctly for further assembly by sequence-specific contacts with the loop motif. B34U is able to make the first of these contacts, but not the latter. Transmission electron microscopy (TEM) images (Fig. S3*A*) and Rh distribution plots (Fig. 2*B*) support these interpretations. In conclusion, aptamer B3 has the properties expected of a PS.

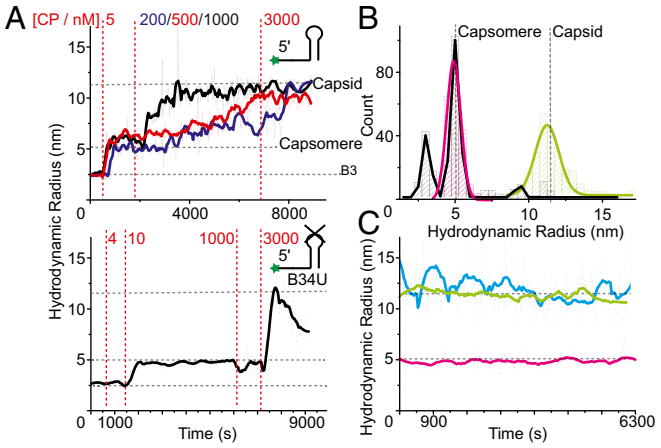

Fig. 2. RNA dependence and sequence-specificity of STNV assembly. (*A*) STNV CP was titrated into 1 nM of 5′ Alexa Fluor-488–labeled B3 (*Upper*) or B34U RNA (*Lower*) (see inset cartoons) in assembly buffer and the resulting hydrodynamic radius monitored using smFCS over time. Dotted red lines indicate titration points with the labels showing the final monomer CP concentrations reached. FFT smoothed data (thick lines) are shown for clarity, with horizontal gray dashed lines showing the sizes of capsid, B3, and capsomere. For B3 the curves for the differing CP concentrations post-formation of the capsomere are color coded and displaced slightly along the time axis. Samples at various points in the titrations were analyzed by TEM (Fig. S3). (*B*) Size distributions of the species in solution for capsomere (magenta) and B3 containing VLP (green) together with the STNV CP + B34U (black) containing species in plot *A*. Gaussian peak fits are shown by thick lines, with vertical gray dashed lines showing expected sizes. (*C*) 1 nM B3 containing capsomeres (magenta) and VLPs (green) were assembled using 4 and 1,000 nM STNV CP, respectively, and their Rh measured over 2 h. The results are compared with similar measurements of 300 nM mRNA VLP (cyan) labeled on the CP with Alexa Fluor-488.

Natural PSs Act Cooperatively

To determine whether assembly with the genomic STNV PSs occurs similarly, and explore potential cooperative effects of having multiple PSs on the same RNA fragment, we carried out CP titrations using a dye-labeled 127 nt RNA. This fragment starts at the 5′ end of the STNV-1 genome (Fig. 1 *B* and *C*) and encompasses five putative PSs, namely sequences that can be folded into stable stem-loops displaying -AXXA- motifs (14). Aptamer B3 has the highest sequence identity with nucleotides 59–72, which is the middle PS on this fragment. We term this fragment PS1-5. We compared reassembly with identical RNA concentrations (10 nM) of PS1-5 and B3 alone (Fig. 3*A*). B3 forms the capsomere at low concentration, ~5 nM and begins to assemble VLPs at 200 nM, although the kinetics of capsid formation are relatively slow and remain incomplete (>4,000 s). On titration to 3,000 nM CP, assembly goes slowly to completion. In contrast, PS1-5 appears to start binding CP only at 120 nM, resulting in a drop of its Rh by 20–30%, reminiscent of the behavior of the full-length genome (7). At 200 nM it rapidly assembles (~250 s) into intact capsids that are nuclease resistant (Fig. S2), suggesting that the multiple PSs promote cooperative formation of the CP–CP contacts found in the capsid.

This difference in behavior between B3 and PS1-5 could be due to the differences in the sequences/structures of their PSs, or simply because of the length of the PS1-5 fragment. To explore this behavior by the longer fragment, we examined the properties of variant fragments in which all of the PSs were mutated by conversion of 4 nt in their loop regions to uridines, or mutating just PS3, or ablating all of the PSs except PS3 (Fig. 3*B*). The all mutant PS1-5 binds CP at 120 nM like the wild-type fragment, but this leads to an increase in its Rh, not a decrease. It does not form $T = 1$ capsids or become nuclease resistant, even when the CP concentration is raised to 3 μM (Fig. 3 *B* and *D*). TEMs suggest that the complexes produced are aggregates rather than partially formed capsids (Fig. S4*A*). This outcome confirms that the loop sequence of PS3 at least is important for CP recognition. When only the putative flanking PSs (1 and 2, and 4 and 5) are mutated, binding again occurs at 120 nM CP and the Rh decreases, showing that this effect is due to initial binding at PS3. However, subsequent assembly is very slow and formation of complete capsids is only achieved at 3,000 nM CP. The result confirms that PS sites other than PS3 play important roles in making capsid assembly efficient, and supports the idea that multiple PSs act cooperatively. When only PS3 is mutated in the PS1-5 fragment, there is no decrease in Rh, confirming that CP binding to PS3 triggers this event, and assembly does not occur until 200 nM CP and does not result in formation of capsids (Fig. 3 *B* and *D*).

A hierarchy of PS CP affinities is an essential requirement of the PS-mediated assembly model (9) and is consistent with the sequence/structure variations in PSs PS1-5 on the genomic fragment. PS3 is the highest affinity site and the other PSs contribute to assembly efficiency but cannot trigger VLP formation alone. The putative STNV PS sites were identified by looking at the short range folding potential of genomic RNA, specifying a stem-loop containing the -AXXA- motif (14), and all such sites with a negative free energy of formation were identified as putative PSs. This simple approach appears successful. Of course, PSs with small folding free energies may only form active conformations in response to CP binding.

Another predicted property of PSs is that their relative locations within the genome is such that they facilitate the CP–CP interactions seen in the capsid, implying that their relative separation in the RNA sequence, which determines their spatial locations in three-dimensions, should be important. To test this, we produced three additional variants of PS1-5 in which 10-nt spacers, chosen in each case to avoid altering the potential secondary structures, were added either 5′ to PS3, i.e., between PS2 and PS3, or 3′ to it, between PS3 and PS4, or at both of these sites. These RNAs are designated 5′ or 3′ spacer PS1-5 and 5′ and 3′ spacer PS1-5, respectively. The results of their titrations with CP are shown in Fig. 3*E*. All three support fairly rapid assembly beyond 120 nM CP, but without an initial Rh decrease. The resulting smFCS traces are very noisy and their distributions suggest that these species form mixtures with few correctly assembled capsids. TEMs (Fig. S4*B*) suggest that although some $T = 1$ capsids are formed by these RNAs, there are many partially formed and malformed species present as well. RNase treatment leads to large changes in these Rh distributions (Fig. S2), confirming that the natural spacing of the five PS sites in the 5′ genomic fragment is important for accurate and efficient capsid assembly. The lack of Rh decreases in fragments with altered PS spacing but containing wild-type PS3 sequences suggests a more complex explanation of the phenomenon in the PS1-5 fragment. CP binding to B34U shows that capsomere-like species can form with high affinity for stem-loops lacking the loop recognition motif. The PS1-5 variants with a wild-type PS3 could allow CPs interacting at that site to contact the neighboring RNA stem-loops, each equivalent to B34U. Although these complexes are not efficient at ensuring cooperative capsid assembly, they could explain the initial decrease in Rh. The spacer variants suggest that both neighboring stem-loops must be contacted to cause this effect, emphasizing the importance of PS positioning in the genome.

Roles of Electrostatics in Assembly and Stability

The sequence-specificity of assembly initiation revealed by the experiments with B3, PS1-5, and their variants, is consistent with PS-mediated assembly, but not with expectations that electrostatic interactions are the principal driving forces behind assembly in this class of viruses (20–25). The PS fragments and the variant sites, with the exception of the insert variants, are of the same length, have the same net charge and are predicted to share common secondary structures. Under ensemble assembly conditions (CP ≥ 10 μM) B34U triggers $T = 1$ capsid formation,

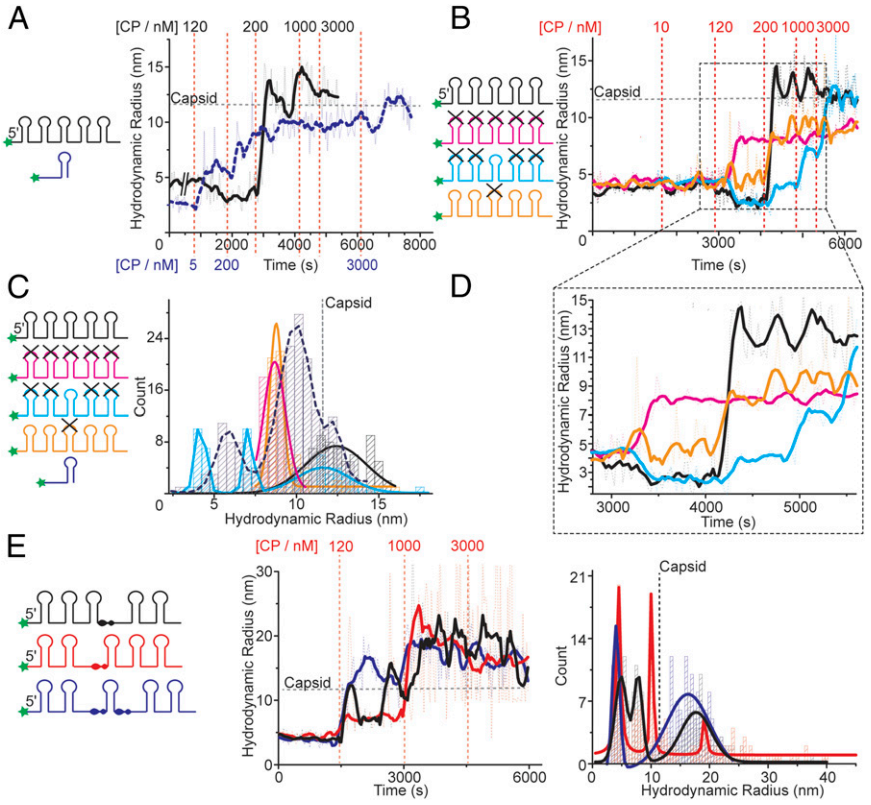

Fig. 3. Packaging signals function cooperatively. (*A*) CP was titrated into 10 nM Alexa Fluor-488–labeled PS1-5 (black) or B3 (dashed - blue). Details as in legend Fig. 2*A*. Titration points are color coded for each RNA. (*B*) CP titration into 10 nM of Alexa Fluor-488–labeled PS1-5 and its sequence variants (X), shown as cartoons (*Left*). (*C*) Color-coded size distributions of species present in the titrations shown in *A* and *B* (substrate key, *Left*). Data shown are taken from the titration of 200 nM CP onwards in *A* and *B*. For PS1-5 and its variants this is 4,000–6,000 s, and for B3 from 2,000 to 8,000 s. (*D*) Expanded section of *B* showing the effects of CP addition on each variant. (*E*) CP titration into spacing variants (see cartoon; *Left*) and their size distribution plots (*Right*). Data shown are from 0 to 5,000 s.

although much less efficiently than B3 (15), i.e., the sequence-specific RNA discrimination is dramatically reduced at higher concentration, and reassembly can be driven almost exclusively by electrostatic interactions (26). In contrast, the outcomes of single-molecule in vitro reassembly, carried out at nanomolar concentrations, are more sensitive to sequence-specific effects and may reflect more accurately conditions in vivo. To examine the effects of charge neutralization on assembly, we repeated the B3 titration at higher RNA concentration (10 nM vs. 1 nM, Fig. S3*B*). More RNA increases the proportion of $T = 1$ capsid and larger intermediates present at lower CP concentrations, but assembly is still relatively slow and requires CP > 1 μM to be complete. It seems therefore that increasing the numbers of phosphodiesters lowers the barrier to assembly, consistent with it being electrostatic in nature.

As a test of this idea, we compared the salt stability of VLPs assembled at either concentration of B3 or with PS1-5, with that of the VLP from *Escherichia coli*, which packages a modified mRNA encoding the CP gene (Fig. 4*A* and Fig. S5*A*). Addition of NaCl to the reassembled particles leads to rapid disassembly of both the 1 and 10 nM B3 VLPs, which are fully disassembled by ∼150 and ∼400 mM, respectively, whereas the PS1-5 and mRNA VLPs remain stable up to 700 mM. STNV virions contain structural calcium ions that bind to three distinct sites located at the particle symmetry axes (Fig. 1*A*). Note, not all these sites are occupied in the crystal structure of the B3 VLP (15). Because these should contribute to electrostatic stability, we repeated the salt titration in the presence of 5 mM EDTA (Fig. 4*A* and Fig. S5*B*). Both B3 VLPs are fully disassembled under these conditions before addition of extra salt, whereas the PS1-5 VLP partially disassembles at 50 mM NaCl, with only the mRNA VLP remaining stable throughout the titration range. These results confirm that particle stability is strongly dependent on formation of favorable, relatively local, electrostatic contacts, following a sequence-specific assembly initiation. Clearly, the presence of longer RNAs encompassing multiple PS sites stabilizes the CP–CP contacts relative to the VLPs formed with the short RNA.

The results described above are consistent with a two-stage assembly model following a free energy path similar to that shown in Fig. 4*B*. Capsomere formation is rapid, high-affinity, and spontaneous, but for the RNA–CP complex to be on-pathway to capsid assembly, the PS stem-loop must be orientated precisely via sequence-specific interactions with the loop motif. In longer RNAs, this initial event is coupled to a decrease in the size of the viral RNA driven by CP binding to correctly placed additional PSs, a vital step in preparing it for specific encapsidation (7, 8). On-pathway capsomeres with short RNAs are prevented from self-assembly to capsid by an electrostatic activation barrier created by the repulsion of clustering six positively charged amino acid side chains into the small volume around the particle threefold helices (Fig. 1*A*). This inhibitory effect can be overcome using more RNA, either in the form of higher concentrations of single PSs or as natural fragments encompassing multiple, appropriately positioned PS sites. The latter are more effective at lowering the activation barrier, allowing assembly to be cooperative once it initiates. Note, the cooperativity for the complete genomic fragment, with up to 30 PSs, would be even more pronounced than seen here for the 5′ fragment with only five such sites.

The Relationship of the Viral RNA to its Coat Protein

By definition there is a close relationship between the genomic RNA of ssRNA viruses and their coat proteins; one is encoded by the other. For satellite viruses, this is heightened by the fact that the only "viral" RNA encapsidated is an extended mRNA encoding the CP subunit. The polymerase responsible for replication is supplied by the helper virus, Tobacco Necrosis Virus. The STNV genome therefore encompasses at least two functions: It is a substrate for both replication and translation. Previous studies have shown that sequences within the 5′ and 3′ untranslated regions (UTRs) are responsible for making both these processes efficient via complementary base pairing between the UTR residues 24–55 and 979–1009 (27) within the RNA. The site in the 5′UTR overlaps the PS containing region. It is therefore

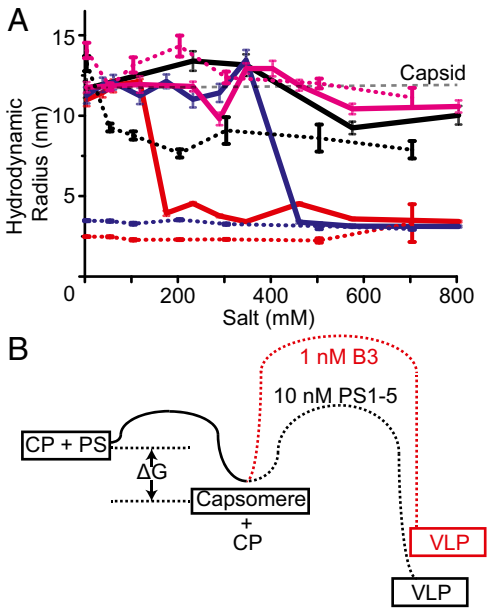

Fig. 4. The roles of electrostatic interactions in assembly of capsomere and capsid. (*A*) Salt titration into mRNA VLPs (magenta), B3 VLPs, assembled at 1:1,000 nM (red) or 10:1,000 nM (blue) RNA:CP, and PS1-5 VLPs assembled at 10:1,000 nM (black). Rh was measured over 10 min at each NaCl concentration. The average and SE of the Rh's are shown. The same experiments, performed in the presence of 5 mM EDTA, are shown as dotted lines. (*B*) Schematic of relative free energy barriers associated with formation of the capsomere and $T = 1$ capsid.

clear that replication/translation and virion assembly are distinct processes brought about by differing RNA conformations.

A global secondary structure of the entire genome calculated using Mfold suggests that only PS1 in the 5′ fragment used here would be formed spontaneously (14). However, if the RNA is folded allowing only for short-range interactions, i.e., in 50-nt sections, then all of the predicted PSs can be formed. The data with the 5′ fragment suggests that STNV packaging occurs on newly replicating RNAs as they emerge from the polymerase. The PS-mediated assembly mechanism also implies that there is a defined spatial relationship between the protein shell of the capsid and the PSs within the genomic RNA. Because ssRNA viruses use RNA-dependent RNA polymerases that lack proof reading functions, the maintenance of such PS sites is a significant challenge. In other systems it has been shown that the coding variants within a viral quasi-species are likely to increase the misincorporation of noncognate RNAs, leading to the suggestion that ssRNA can be horizontal vectors for gene transfer (28, 29). It is thus likely that strategies have evolved that favor conservation of PSs in viral evolution, thus enhancing packaging specificity.

Discussion

There are different ways in which ssRNA viruses may counter the challenge of enhancing packaging of their cognate genomic RNAs against the background of cellular competitor RNAs. One solution could be the physical separation of viral components from cellular factors, but the evidence of significant misincorporation even in viruses believed to operate this way (29) suggests that other mechanisms are likely to have selective advantages. An alternative possibility is that packaging of cognate genomes is the result of specificity between viral genomes and their cognate CPs. Such a mechanism would mirror a phenomenon occurring more widely, in particular in early events in the development of the RNA–protein world: a bias toward interactions between encoded proteins and their mRNAs, even when those proteins are not thought to function via RNA binding (30–32).

We speculate that a similar phenomenon may give rise to PS conservation, and thus ensure packaging specificity, in STNV. In particular, the molecular details of the PS3–CP interaction illustrate how this can be achieved. PS3 is located at nt60 in the STNV-1 genome, and forms part of the codon for amino acid residue 10 in the CP, or 9 neglecting the N-terminal methionine (Fig. 1*A*). Thus, the strongest PS forms a section of the gene encoding the amino acids in the CP that alter their conformation upon PS binding (Fig. 5). This match may of course be coincidental, but a statistical analysis (*SI Materials and Methods*) suggests that it would only occur ∼9% of the time by chance. PS3 is also one of only three PS sites, with PS11 and PS12, conserved in all three known STNV strains (Fig. S6). Testing this idea experimentally would be challenging, so we compared STNV with the two other ssRNA viruses for which we have detailed CP–RNA structural data, namely MS2 and TMV (33, 34) (Fig. S7). The former assembles via a mechanism mediated by multiple PSs, whereas the latter is a helical virus that assembles from a unique PS-RNA contact. There are four known PSs within the MS2 CP gene. One of them encodes a 6-aa peptide that plays a significant role in recognizing the strongest PS in that genome, TR (Fig. 5*B* and Fig. S7 *A* and *C*). The probability that this match occurs by chance in that case is 3.1% (*SI Materials and Methods*). In Group 2 Tobacco Mosaic Viruses, the origin of assembly, a PS, encodes one of the two peptides forming the RNA-binding site in the CP (35–37), suggesting similar relationships occur even in a helical virus (Fig. S7*D*). The occurrence of this phenomenon in a number of viruses suggests that PS encoding features of the protein sites they are in contact with could be an unappreciated aspect of viral CP–genome interactions.

STNV CP shares a common fold with many other ssRNA viruses, most of which are more complex, having quasi-equivalent capsids, in part defined by the conformations of their positively charged N-terminal regions that often contain arginine-rich motifs, ARMs (38). Removal of this region in STNV prevents assembly (39), whereas in other viruses it promotes RNA-independent assembly of $T = 1$ capsids (40). Here we have used smFCS assays to probe the roles of RNA in STNV assembly at nanomolar concentrations, showing that it follows a two-stage PS-mediated mechanism, thus preventing assembly around noncognate RNAs. Even with relatively short PS fragments we have recreated assembly features seen with the intact genome, showing that CP recognition of degenerate PSs, and the latter's relative spacing along the genomic RNA, both contribute to efficient cooperative capsid assembly (9, 41). These assays also validate the use of RNA SELEX against viral CPs, combined with theoretical analysis techniques pioneered by us, as a route for identifying PSs (15).

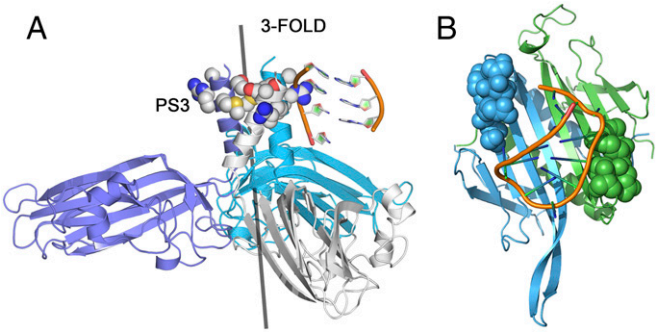

Fig. 5. PSs encode critical features of viral CPs involved in PS RNA recognition. (*A*) View of the contacts between the ordered portion of the B3 RNA (cf. PS3) and the PS3-encoded amino acids (space filling representation) in the B3 VLP structure (PDB ID 4BCU). (*B*) The MS2 coat protein dimer in A/B conformation in complex with the highest affinity genomic PS, the TR stem-loop (34). The amino acids in the A (blue) and B (green) chains that are encoded by one of the four PSs within the CP gene (Fig. S7) are shown in space-filling representation.

MICROBIOLOGY

It is tempting to speculate that, because of the multiple selective advantages for viruses assembling via such mechanisms, viruses with similar CP folds will also assemble by similar routes (14). The requirement to encode PSs, and perhaps the peptide functions they encode, introduces a previously unsuspected constraint on viral evolution, especially given the density of functions located within the same RNA sequences. Disrupting this assembly mechanism would therefore be a novel antiviral strategy. Previously, we showed that hydrodynamic collapse of viral genomes is a signature of cognate encapsidation, and has many similarities to the assembly of the ribosome. The roles of the multiple genomic PSs revealed here also have similarities to protein sequestration by noncoding (nc) RNAs, "RNA sponges" (42), where proteins bind sequentially, specifically, and cooperatively to stem-loops, leading to protection of a ncRNA from degradation. Assembly of ssRNA viruses may therefore have more in common with cellular RNA biology than previously suspected.

Materials and Methods

STNV CP was prepared by dissociation of recombinant VLPs produced in *E. coli* (15). RNAs were either chemically synthesized with a dye attached to the 5′ end, (DNA technology A/S, Denmark) or transcribed in the presence of 5′-amino-GMP which was subsequently chemically coupled to a dye using an amine reactive Alexa Fluor 488 SDP ester (Invitrogen). SmFCS assays and data analysis were performed as described (8).

ACKNOWLEDGMENTS. We thank Prof. Arwen Pearson for access to the X-ray density and model of the B3 VLP and for helpful comments, and Prof. Cheng Kao and his colleagues at Indiana University for access to their unpublished CLIP-SEQ data on MS2. We thank the University of Leeds for its support of the Single Molecule Facility within the Astbury Centre. This work was supported by UK Biotechnology and Biological Science Research Council Grants BB/J00667X/1, BB/L022095/1, BB/L021803/1, BB/J004596/1 and the Wellcome Trust (089311/Z/09/Z and 090932/Z/09/Z). R. Twarock acknowledges funding via a Royal Society Leverhulme Trust Senior Research fellowship (LT130088), and E.C.D. acknowledges funding via an Early Career Leverhulme Trust fellowship (ECF-2013-019).

1. Woodson SA (2011) RNA folding pathways and the self-assembly of ribosomes. *Acc Chem Res* 44(12):1312–1319.
2. Simon AE, Gehrke L (2009) RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochim Biophys Acta* 1789(9-10):571–583.
3. Schneemann A (2006) The structural and functional role of RNA in icosahedral virus assembly. *Annu Rev Microbiol* 60:51–67.
4. Enserink M (2014) Infectious diseases. Crippling virus set to conquer Western Hemisphere. *Science* 344(6185):678–679.
5. Lehtinen M, Dillner J (2013) Clinical trials of human papillomavirus vaccines and beyond. *Nature reviews. Clin Oncol* 10(7):400–410.
6. Yan R, Hallam A, Stockley PG, Boyes J (2014) Oncogene dependency and the potential of targeted RNAi-based anti-cancer therapy. *Biochem J* 461(1):1–13.
7. Borodavka A, Tuma R, Stockley PG (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc Natl Acad Sci USA* 109(39):15769–15774.
8. Borodavka A, Tuma R, Stockley PG (2013) A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol* 10(4):481–489.
9. 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(14): 5361–5366.
10. Ling CM, Hung PP, Overby LR (1969) Specificity in self-assembly of bacteriophages Q beta and MS2. *Biochemistry* 8(11):4464–4469.
11. Stockley PG, et al. (2007) A simple, RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. *J Mol Biol* 369(2):541–552.
12. Dykeman EC, Stockley PG, 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(17):3235–3249.
13. Dent KC, et al. (2013) The asymmetric structure of an icosahedral virus bound to its receptor suggests a mechanism for genome release. *Structure* 21(7):1225–1234.
14. Bunka DH, et al. (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(1):51–65.
15. Ford RJ, et al. (2013) Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *J Mol Biol* 425(6):1050–1064.
16. van Lipzig R, et al. (2002) The 5′ and 3′ extremities of the satellite tobacco necrosis virus translational enhancer domain contribute differentially to stimulation of translation. *RNA* 8(2):229–236.
17. Bringloe DH, Gultyaev AP, Pelpel M, Pleij CW, Coutts RH (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(Pt 6):1539–1546.
18. Bringloe DH, Pleij CW, Coutts RH (1999) Mutation analysis of cis-elements in the 3′- and 5′-untranslated regions of satellite tobacco necrosis virus strain C RNA. *Virology* 264(1):76–84.
19. Meulewaeter F, Danthinne X, Van Montagu M, Cornelissen M (1998) 5′- and 3′-sequences of satellite tobacco necrosis virus RNA promoting translation in tobacco. *Plant J* 14(2):169–176.
20. Rudnick J, Bruinsma R (2005) Icosahedral packing of RNA viral genomes. *Phys Rev Lett* 94(3):038101.
21. van der Schoot P, Bruinsma R (2005) Electrostatics and the assembly of an RNA virus. *Phys Rev E Stat Nonlin Soft Matter Phys* 71(6 Pt 1):061928.
22. Belyi VA, Muthukumar M (2006) Electrostatic origin of the genome packing in viruses. *Proc Natl Acad Sci USA* 103(46):17174–17178.
23. Balint R, Cohen SS (1985) The incorporation of radiolabeled polyamines and methionine into turnip yellow mosaic virus in protoplasts from infected plants. *Virology* 144(1):181–193.
24. Bruinsma RF (2006) Physics of RNA and viral assembly. *Eur Phys J E Soft Matter* 19(3): 303–310.
25. Yoffe AM, et al. (2008) Predicting the sizes of large RNA molecules. *Proc Natl Acad Sci USA* 105(42):16153–16158.
26. Garmann RF, et al. (2014) Role of electrostatics in the assembly pathway of a single-stranded RNA virus. *J Virol* 88(18):10472–10479.
27. Kaempfer R, van Emmelo J, Fiers W (1981) Specific binding of eukaryotic initiation factor 2 to satellite tobacco necrosis virus RNA at a 5′-terminal sequence comprising the ribosome binding site. *Proc Natl Acad Sci USA* 78(3):1542–1546.
28. Routh A, Domitrovic T, Johnson JE (2012) Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci USA* 109(6): 1907–1912.
29. Routh A, Domitrovic T, Johnson JE (2012) Packaging host RNAs in small RNA viruses: An inevitable consequence of an error-prone polymerase? *Cell Cycle* 11(20): 3713–3714.
30. Polyansky AA, Zagrovic B (2013) Evidence of direct complementary interactions between messenger RNAs and their cognate proteins. *Nucleic Acids Res* 41(18):8434–8443.
31. Polyansky AA, Hlevnjak M, Zagrovic B (2013) Proteome-wide analysis reveals clues of complementary interactions between mRNAs and their cognate proteins as the physicochemical foundation of the genetic code. *RNA Biol* 10(8):1248–1254.
32. Hlevnjak M, Polyansky AA, Zagrovic B (2012) Sequence signatures of direct complementarity between mRNAs and cognate proteins on multiple levels. *Nucleic Acids Res* 40(18):8874–8882.
33. Namba K, Pattanayek R, Stubbs G (1989) Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 A resolution by X-ray fiber diffraction. *J Mol Biol* 208(2):307–325.
34. Valegård K, Murray JB, Stockley PG, Stonehouse NJ, Liljas L (1994) Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* 371(6498):623–626.
35. Fukuda M, Meshi T, Okada Y, Otsuki Y, Takebe I (1981) Correlation between particle multiplicity and location on virion RNA of the assembly initiation site for viruses of the tobacco mosaic virus group. *Proc Natl Acad Sci USA* 78(7):4231–4235.
36. Guilley H, Jonard G, Richards KE, Hirth L (1975) Sequence of a specifically encapsidated RNA fragment originating from the tobacco-mosaic-virus coat-protein cistron. *Eur J Biochem* 54(1):135–144.
37. Takamatsu N, Ohno T, Meshi T, Okada Y (1983) Molecular cloning and nucleotide sequence of the 30K and the coat protein cistron of TMV (tomato strain) genome. *Nucleic Acids Res* 11(11):3767–3778.
38. Rao AL (2006) Genome packaging by spherical plant RNA viruses. *Annu Rev Phytopathol* 44:61–87.
39. Lane SW, et al. (2011) Construction and crystal structure of recombinant STNV capsids. *J Mol Biol* 413(1):41–50.
40. Sorger PK, Stockley PG, Harrison SC (1986) Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. *J Mol Biol* 191(4):639–658.
41. Stockley PG, et al. (2013) Packaging signals in single-stranded RNA viruses: Nature's alternative to a purely electrostatic assembly mechanism. *J Biol Phys* 39(2):277–287.
42. Duss O, et al. (2014) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* 509(7502):588–592.

Supporting Information

Patel et al. 10.1073/pnas.1420812112

SI Materials and Methods

STNV Purification. Recombinant STNV VLPs were purified from *E. coli* (1) and the CP monomers purified by disassembly in 50 mM Tris (pH 8.5), 10 mM EDTA, in the presence of Complete Protease Inhibitor Mixture (Roche). STNV CP was separated from the mRNA by sequential Q-Sepharose, and SP-Sepharose columns (GE Healthcare). 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 analyzed by SDS/PAGE and its concentration determined by UV absorbance. Fractions with an A_{260}:A_{280} ratios of 0.6 or lower were used in assembly assays. CP monomer was stored in 50 mM Hepes (pH 7.5), 25 mM NaCl and 50% (wt/vol) glycerol at −20 °C. STNV VLPs were labeled using Alexa Fluor-488 SDP ester (Invitrogen) over 4 h at room temperature in carbonate buffer (pH 8.3) and the modified VLPs purified using a NAP5 column, followed by a 15–45% (wt/vol) linear sucrose gradient. Peak fractions were then dialyzed into a buffer containing 50 mM Hepes (pH 7.5) and 100 mM NaCl.

RNA Synthesis. Aptamers B3 and B34U were chemically synthesized and HPLC purified by DNA technology A/S, Denmark. PS1-5 and variants (Table S2) were produced using in vitro transcription using a T7 Megascript kit (Ambion) using dsDNA templates obtained from Eurofins MWG and a Phusion high fidelity DNA polymerase kit (NEB). These transcripts were made using a 4:1 molar excess of aminohexyl-G-monophosphate (Tebubio) to GTP to create a chemically reactive 5′ amino group, which was then coupled to Alexa Fluor 488 SDP ester (Invitrogen). RNAs > 100 nt in length were purified using an QIAgen RNeasy Mini elute spin column, adjusting the ethanol concentration when appropriate to ensure binding. RNAs < 100 nt were purified by phenol:chloroform extraction and isopropanol precipitation. RNAs were analyzed by 10% (wt/vol) denaturing polyacrylamide gel electrophoresis. The variants of PS1-5 with an insert 3′ and/or 5′ to PS3 were designed by adding one nucleotide at a time either side of PS3, checking the Mfold predictions after each iteration to ensure no undesirable structures are formed. The concentration of fluorescently labeled RNA was calculated using the following formula: $C\ (\mu g/mL) = 40 \times (A_{260} - (0.3 \times A_{494}))$ (Invitrogen). All kits were used according to the manufacturers' protocols unless stated. The efficiencies of labeling protein/RNA were calculated using the equation: $(\text{moles dye per mole protein/RNA}) = A_{494}/(\varepsilon_{488} \times [\text{Target}])$ (Invitrogen) where A_{494} is the absorbance at 494 nm and ε_{488} is the extinction coefficient of Alexa Fluor 488 at 494 nm.

Assembly Assays. Assembly reactions were performed by adding RNA in water to STNV CP monomer in a buffer containing 50 mM Hepes (pH 7.5), 25 mM NaCl and 50% (wt/vol) glycerol. Using a 10× stock of reassembly buffer, the assay conditions were then adjusted to 50 mM Hepes (pH 7.5), 3 mM $CaCl_2$, 1 mM DTT, and 0.05% (vol/vol) Tween-20 (and where appropriate NaCl) at 21 °C. Successive additions of monomer were then added as necessary, with CP never exceeding 5% of total reaction volume. Manual mixing throughout the reactions caused an approximate 1-min delay at the start of time-resolved FCS data collection. FCS measurements were made using a custom-built FCS setup with 30-s data accumulation per autocorrelation function (CF). Individual CFs were decomposed into triplet state relaxation and diffusion (characterized by diffusion time, TD) components and the latter was converted into an apparent hydrodynamic radius, Rh (2). Samples for TEM were taken at the end of each measurement. Plots of Rh over time were smoothed using the FFT filter in Origin Pro-8 with a cutoff percentage of 35%. Plots of Rh distribution were also fitted using Origin Pro-8 software, to a normal single or multiple peak Gaussian function.

Photon Counting Statistics and Burst Size Analysis. Bursts from the passage of individual molecules through a confocal volume were measured at a concentration of 500 pM Alexa Fluor-488–labeled B3 RNA to which increasing amounts of coat protein were added (0.5, 2.5, 12.5, and 22.5 nM).The laser power was kept low at 30 μW. Photon arrival times were recorded and binned (0.5 ms). Several (5–10) repeated runs accumulating at least 1×10^6 photons each were performed. Bursts were identified as one or more consecutive bins with photon counts above the background level (set by buffer recording) containing an accrued minimum of 20 photons.

Statistical Analysis of PS Locations. To determine the probability of one of the four MS2 PSs overlapping with the coat protein codons for at least four contact sites in the amino acid sequence, we performed the following analysis. We randomly placed four PSs coding for six amino acids in a window corresponding to the 130 CP codons and monitored how many times they overlap with the positions of the 11 codons of amino acids within the contact map in Fig. S7*A*. Based on 10 million simulations, we established the chance of a single PS overlapping with at least four amino acid–TR contacts is 3.1%.

Similarly, for STNV, 17 PSs of an average length of 5 codons were randomly positioned within the 196 codons of the CP gene. After 10 million simulations, we obtained an 8.6% chance that a single PS overlaps with the known contact sites (3).

TEM. Each sample was added to a glow discharge-irradiated Carbon/Formvar 300-mesh grid (Agar Scientific) and stained using 2% (wt/vol) uranyl acetate, and then viewed with a FEI Spirit G2 microscope at 49,000× magnification.

1. Lane SW, et al. (2011) Construction and crystal structure of recombinant STNV capsids. *J Mol Biol* 413(1):41–50.
2. Borodavka A, Tuma R, Stockley PG (2012) Evidence that viral RNAs have evolved for efficient, two-stage packaging. *Proc Natl Acad Sci USA* 109(39):15769–15774.
3. Bunka DH, et al. (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(1): 51–65.

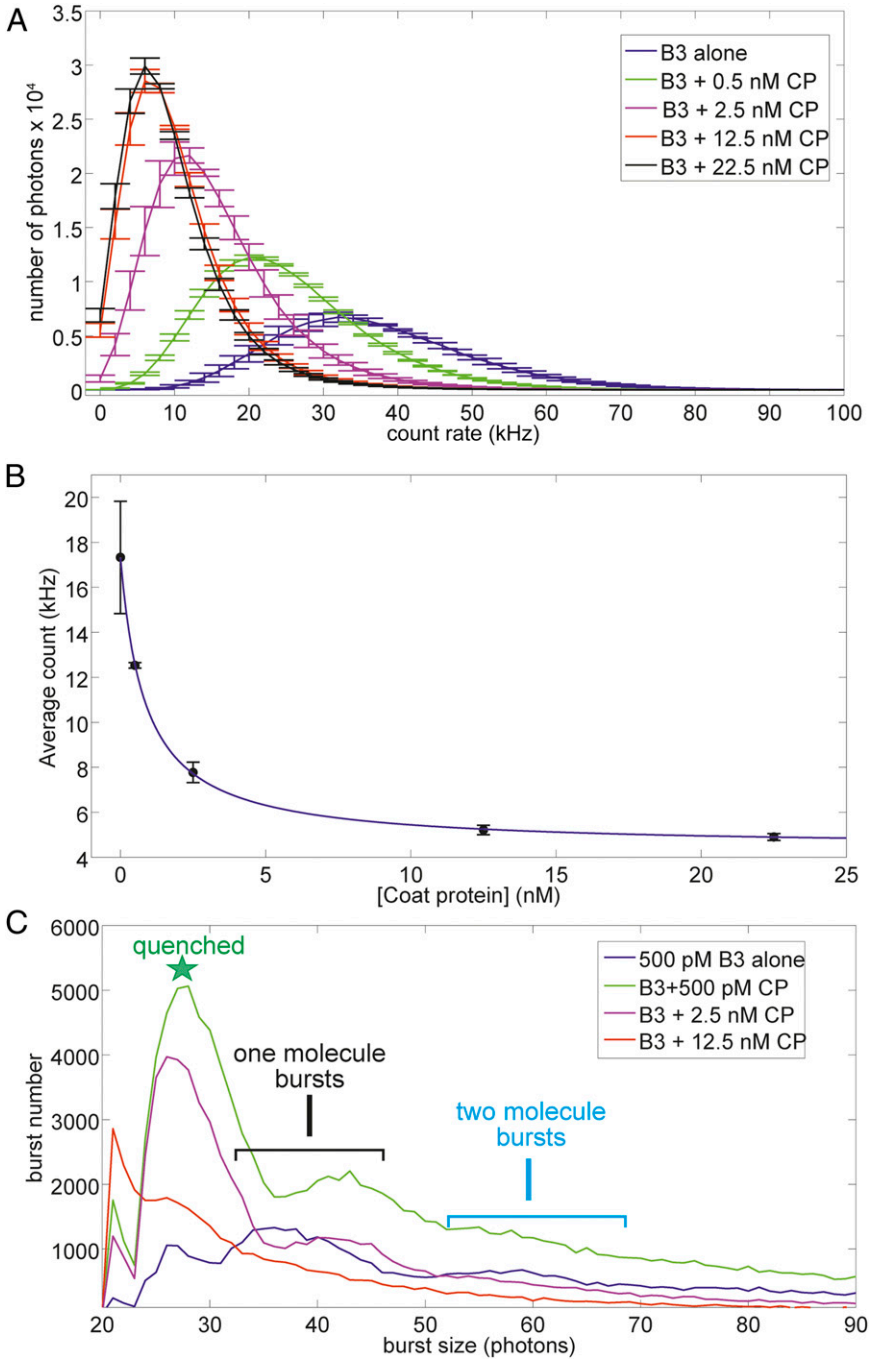

Fig. S1. Photon counting statistics and apparent stoichiometry of the B3 RNA:CP complex. (*A*) Photon count distributions are shown for several coat protein concentrations (see color coding in inset legend). The average photon counts decrease due to quenching upon binding. Bin counting errors estimated from repeated runs. Bin size: 0.5 ms. (*B*) Plot of the average count rate estimated from the histograms in *A* as a function of coat protein concentration. A single saturation binding model (black solid line) was fitted to the data with apparent $K_d = 0.8 \pm 0.1$ nM. (*C*) Burst size distributions of B3 RNA:CP complexes. Three principal classes of bursts were identified: (*i*) single molecule bursts (35–45 photons), (*ii*) two molecule bursts (55–70 photons), and (*iii*) quenched bursts (<30 photons). Number of quenched photon bursts increases with the binding of CP as shown in *A* and *B*. Note that single molecule bursts are significantly populated, whereas hardly any two molecule bursts appear at 0.5 nM and 2.5 nM coat, for which substantial numbers (>50%) of B3 molecules are expected to be bound by CP (see *B*).

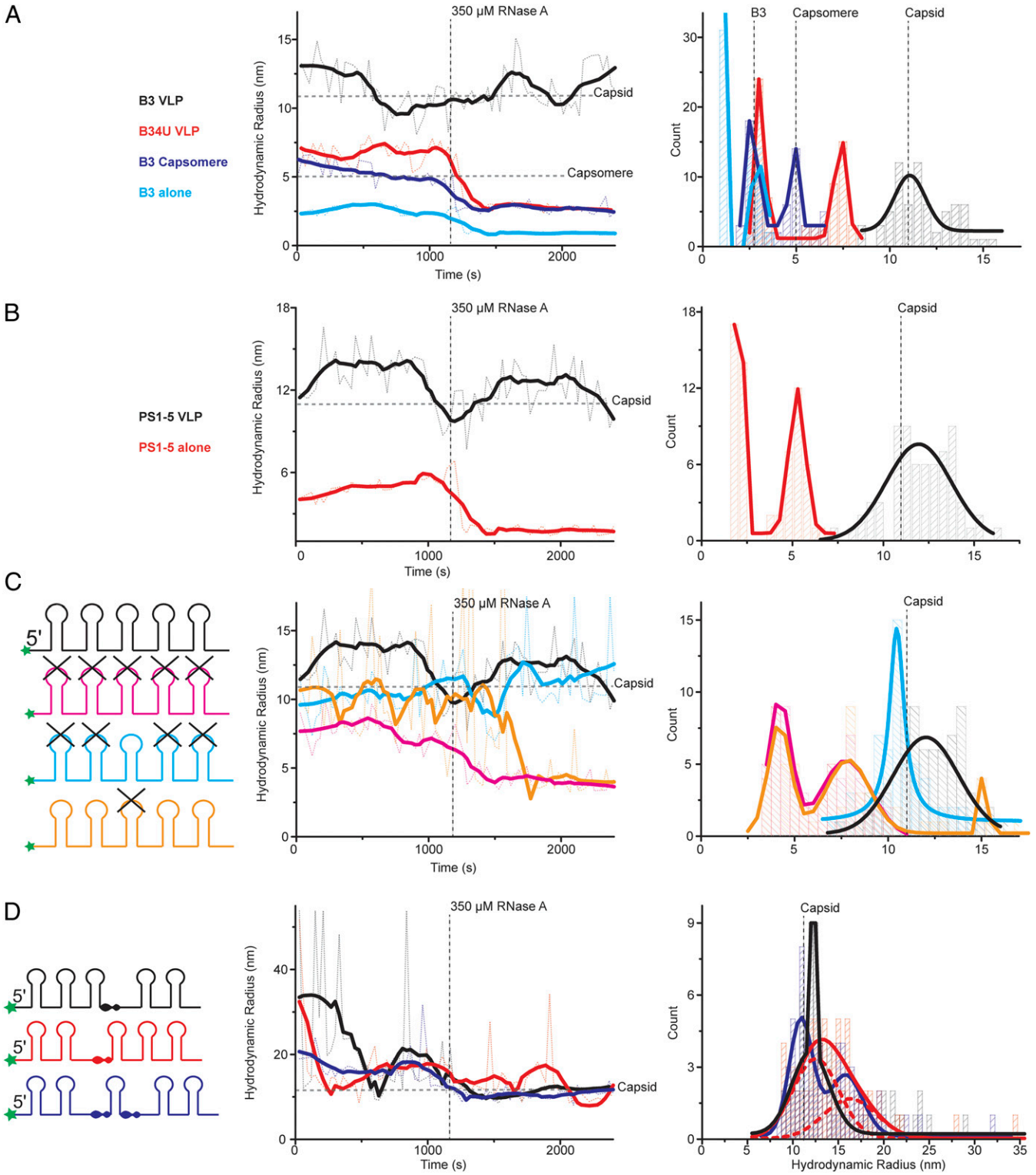

Fig. S2. RNase treatment of RNA:CP complexes monitored by smFCS. (*A*, *Left*) 350 μM RNase A was added (shown via vertical dotted gray line) into B3 VLP (1 nM RNA: 3 μM CP) (black), B34U VLP (1 nM RNA: 3 μM CP) (red), B3 capsomere (1 nM RNA: 5 nM CP) (blue) and 1 nM B3 RNA alone (cyan), and the resulting hydrodynamic radius was monitored using smFCS over time. FFT smoothed data (thick lines) are shown over raw data (thin dotted lines), with horizontal gray dashed lines showing the sizes of capsid and capsomere. (*A*, *Right*) Size distributions of species in solution. Data shown are taken from 0 to 2,400 s. Gaussian peak fits are shown by thick lines, with vertical gray dashed lines showing expected sizes. (*B*) PS1-5 VLP (10 nM RNA: 3 μM CP) (black) and 10 nM PS1-5 RNA alone (red) were RNase treated as in *A*. (*C*) PS1-5 VLP (10 nM RNA: 3 μM CP) (black) and its sequence variants were RNase treated as in *A*. (*D*) PS1-5 spacing variant VLPs shown as cartoons (*Left*) were RNase treated as in *A*.

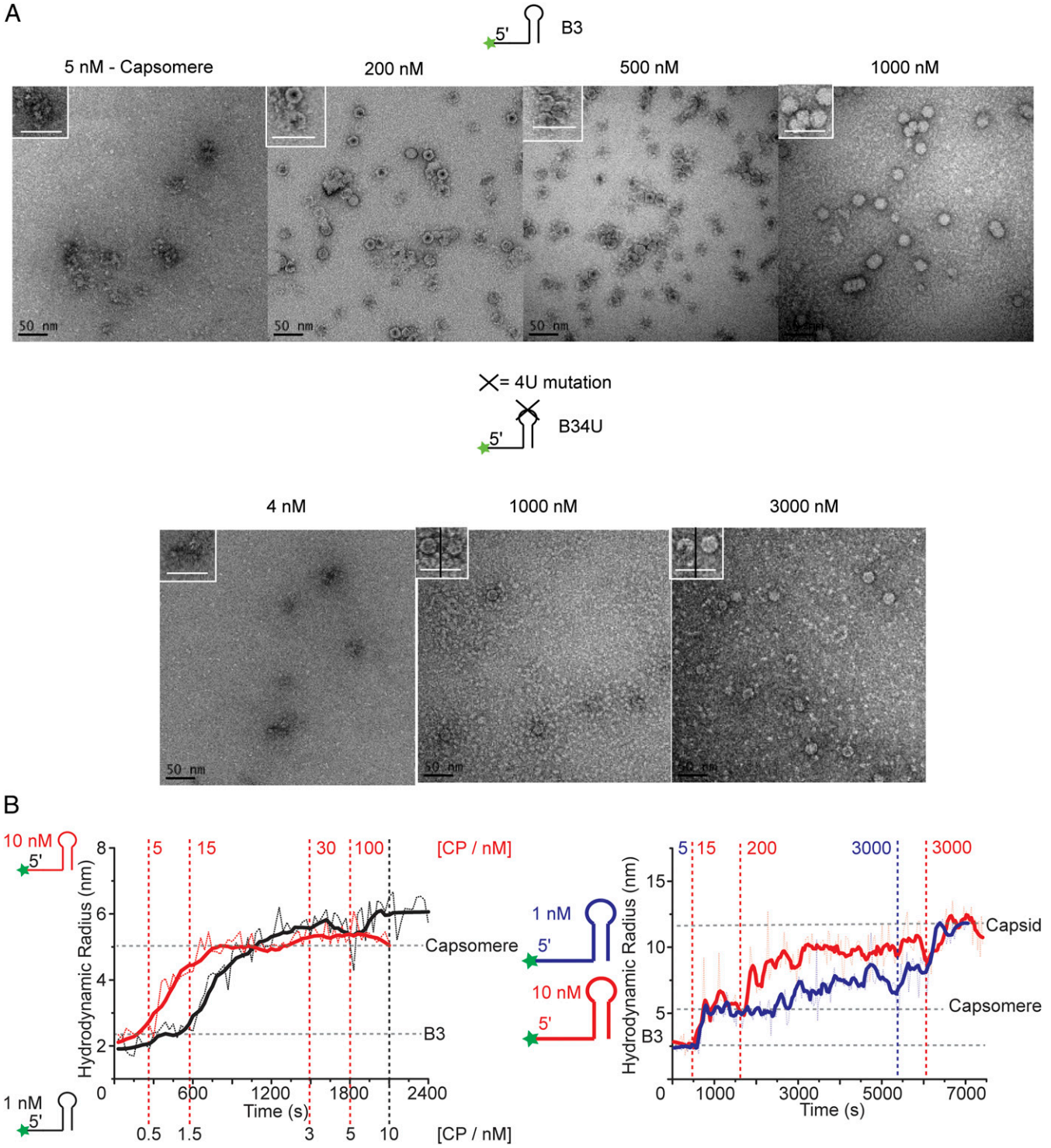

Fig. S3. TEMs of assembly products with B3 and B34U. (*A*, *Upper*) (Left to right) reassembly of STNV using RNA aptamer B3 at 5, 200, 500, and 1,000 nM CP. (*A*, *Lower*) (Left to right) reassembly of STNV using RNA aptamer B34U at 4, 1,000, and 3,000 nM of CP. (Scale bars, 50 nm.) (*B*, *Left*) CP was titrated in the amounts stated (top/bottom of graph) at discrete points indicated by the vertical red lines into 1 nM (black) or 10 nM (red) of Alexa Fluor-488–labeled RNA aptamer B3 in an attempt to form capsomere. The resulting hydrodynamic radius was monitored using (smFCS) over time. FFT smoothed data (thick lines) are shown over raw data (thin dotted lines), with gray dashed lines showing expected sizes of different species seen in solution. (*B*, *Right*) CP was titrated in the amounts stated (top of graph) at discrete points indicated by the vertical red lines into 1 nM (blue) or 10 nM (red) of Alexa Fluor-488–labeled RNA aptamer B3 and the resulting hydrodynamic radius was monitored using (smFCS) over time. Data were analyzed as described above.

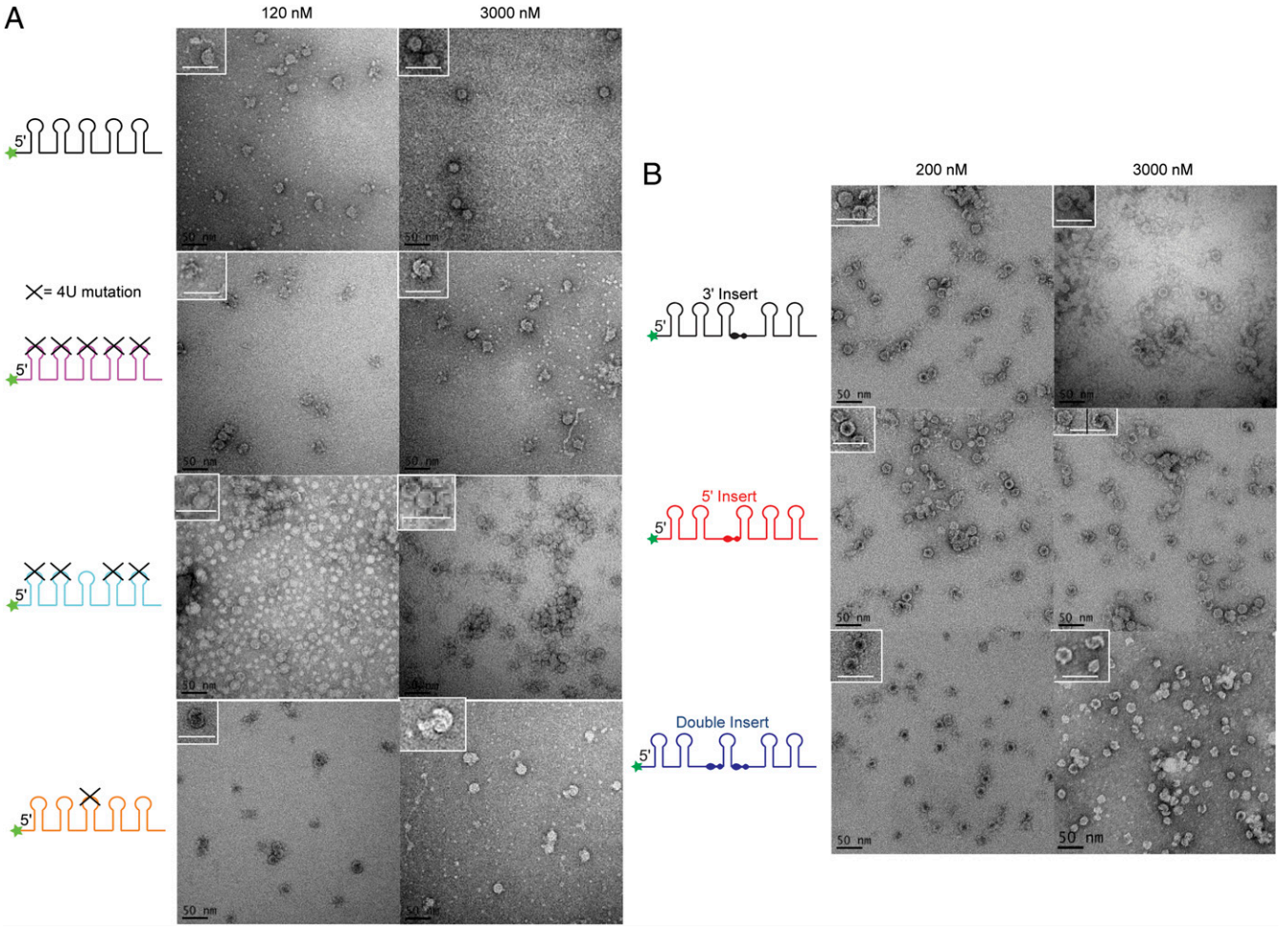

Fig. S4. TEMs of assembly reactions with the genomic fragment, PS1-5, and its variants. (*A*) EMs of PS1-5 mediated assembly, and the assembly of its mutant variants as indicated by the substrate key. Left to right, the reassembly at 120 and 3,000 nM of CP. (*B*) EMs of the 10/20 nucleotide sequence insert variants of RNA oligo PS1-5 as indicated by the substrate key. Left to right, the reassembly of STNV at 120 and 3,000 nM of CP. (Scale bars, 50 nm.)

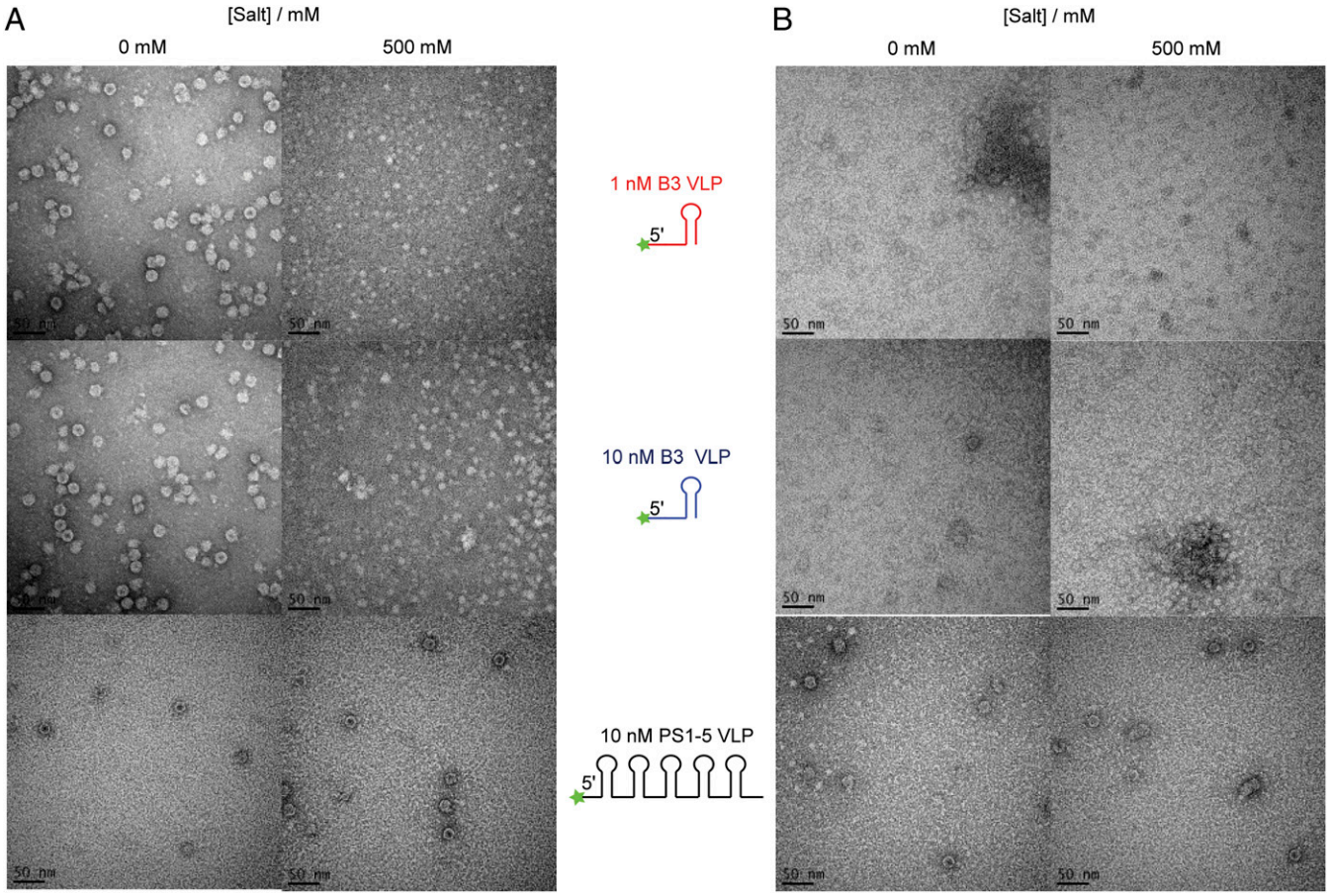

Fig. S5. Analysis of the roles of electrostatic interactions in STNV VLP stability. (*A*) TEMs of smFCS assembled VLPs (as denoted in substrate key) in 0 mM and 500 mM NaCl. (*B*) TEMs of smFCS assembled VLPs (as denoted in substrate key) in the presence of 5 mM EDTA and 0 mM or 500 mM NaCl. (Scale bars, 50 nm.)

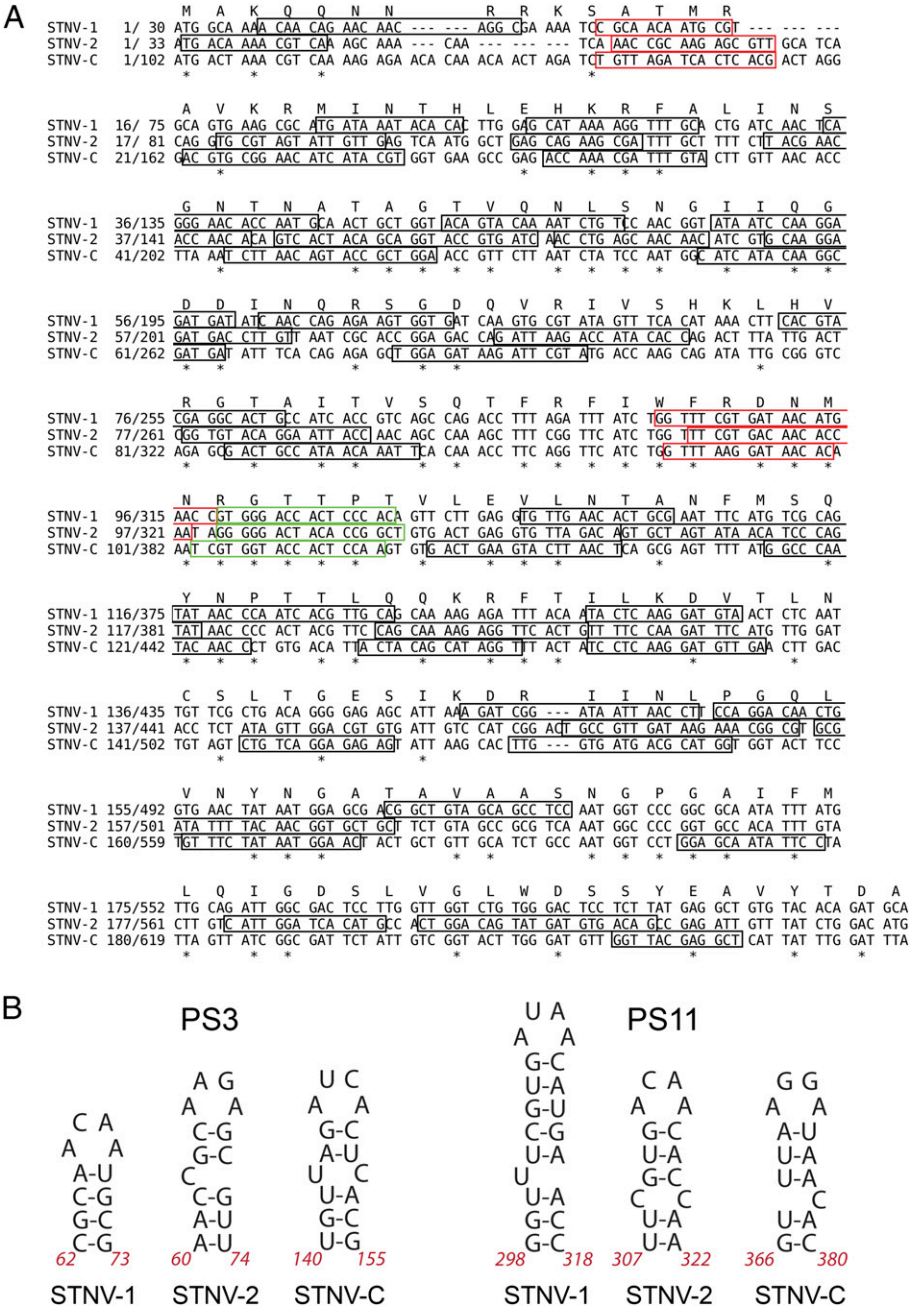

Fig. S6. Conservation of PS sites between the three known genomes of STNV-1, STNV-2, and STNV-C. (*A*) Translation of the coat protein gene is shown for the STNV-1 genome and is indicated by the amino acid letter sequence above the codon. Amino acids which are conserved across all three strains are indicated by an asterisk below the sequences. Amino acid number and nucleotide number for the first amino acid/nucleotide in each row in the genome are indicated by the aa#/nt# numbers. Locations of the putative PSs identified from Bunka et al. (1) for each STNV strain are indicated by a box. Those which overlap with at least four amino acids in all three strains are colored red and green, and their potential secondary structures in all three genomes are shown below (*B*).

1. Bunka DH, et al. (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(1):51–65.

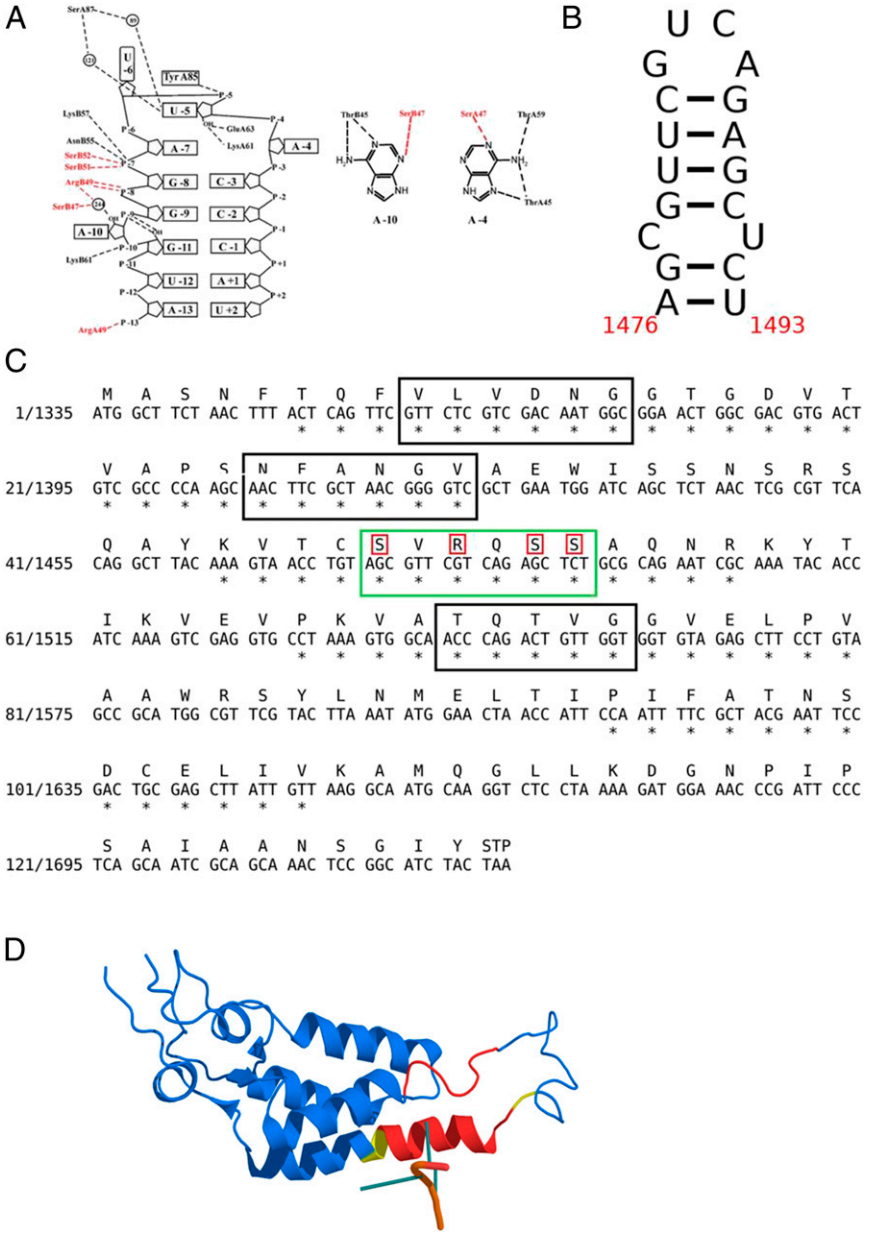

Fig. S7. The relationship between PSs and the sequences they encode. An MS2 packaging signal within the CP gene encodes a significant fraction of amino acid the residues involved in PS recognition. Four of the 60 PSs identified in MS2 via CLIP-SEQ analysis overlap with the coat protein (CP) gene. (*A*) One of these encodes four of the known contact sites between CP and RNA. The contact map, adapted from Horn et al. (1), shows PS-encoded contacts in red. (*B*) The secondary structure of the PS encoding these contact residues. (*C*) The MS2 CP gene sequence with the positions of its four PSs shown boxed, the green box indicates the PS shown in *B*; red boxes around the corresponding amino acids indicate residues that contact the TR PS. (*D*) TMV origin encodes a peptide in the CP it binds to. The structure of the TMV CP (PDB ID 2TMV) adapted from Namba et al. (2), determined by fiber diffraction methods at 2.9 Å resolution shown as a ribbon cartoon, in which the peptides which bind the RNA (stick model, front) are shown in red, whereas the peptides encoded by the complete origin of assembly is colored yellow.

1. Horn WT, et al. (2004) The crystal structure of a high affinity RNA stem-loop complexed with the bacteriophage MS2 capsid: Further challenges in the modeling of ligand-RNA interactions. *RNA* 10(11):1776–1782.
2. Namba K, Pattanayek R, Stubbs G (1989) Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 A resolution by X-ray fiber diffraction. *J Mol Biol* 208(2):307–325.

Table S1. Dye-labeling of RNA oligomers

RNA	ε_{260}	A_{260}	A_{494}	Degree of labeling, %
B3	296.7	1.2	0.085	98.5
B34U	292.3	0.8	0.06	99
PS1-5	1316.2	3.3	0.015	27
3′ spacer PS1-5	1749.9	2.73	0.012	36
5′ spacer PS1-5	1760.8	1.63	0.011	56
5′ & 3′ spacer PS1-5	1895.1	1.46	0.015	91
PS1-5 (U mutants)	1285	4	0.024	34
PS1-5 (B34U mutant)	1311.6	1.05	0.01	59
PS1-5 (Flanking mutants)	1289.6	6.6	0.083	76

Table S2. RNA sequences

RNA oligonucleotide	Sequence (5′ – 3′)
B3	CCUUUUCAAGACAUGCAACAAUGCACACAG
B34U	CCUUUUCAAGACAUGCAUUUUUGCACACAG
PS1-5	GAGUAAAGACAGGAAACUUUACUGACUAACAUGGCAAAACAACAGAACAACAGGCGAAAAUCCGCAACAAUGCGUGCAGUGAAGCGCAUGAUAAAUACACACUUGGAGCAUAAAAGGUUUGCACUGA
5′ & 3′ spacer PS1-5	AGUAAAGACAGGAAACUUUACUGACUAACAUGGCAAAACAACAGAACAACAGGCGAAAAUCGAUAGAAAAACGCAACAAUGCGAAAAAACGAAAUGCAGUGAAGCGCAUGAUAAAUACACACUUGGAGCAUAAAAGGUUUGCACUGA
5′ spacer PS1-5	AGUAAAGACAGGAAACUUUACUGACUAACAUGGCAAAACAACAGAACAACAGGCGAAAAUCGAUAGAAAAACGCAACAAUGCGUGCAGUGAAGCGCAUGAUAAAUACACACUUGGAGCAUAAAAGGUUUGCACUGA
3′ spacer PS1-5	AGUAAAGACAGGAAACUUUACUGACUAACAUGGCAAAACAACAGAACAACAGGCGAAAAUCCGCAACAAUGCGAAAAACGAAAUGCAGUGAAGCGCAUGAUAAAUACACACUUGGAGCAUAAAAGGUUUGCACUGA
PS1-5 (U mutants)	AGUAAAGACUUUUUACUUUACUGACUAACAUGGCAAAACAACAUUUUUACAGGCGAAAAUCCGCAUUUUUGCGUGCAGUGAAGCGCAUGAUAUUUUCACACUUGGAGCAUUUUUGGUUUGCACUGA
PS1-5 (Flanking mutants)	AGUAAAGACUUUUUACUUUACUGACUAACAUGGCAAAACAACAUUUUUACAGGCGAAAAUCCGCAACAAUGCGUGCAGUGAAGCGCAUGAUAUUUUCACACUUGGAGCAUUUUUGGUUUGCACUGA
PS1-5 (B34U mutant)	AGUAAAGACAGGAAACUUUACUGACUAACAUGGCAAAACAACAGAACAACAGGCGAAAAUCCGCAUUUUUGCGUGCAGUGAAGCGCAUGAUAAAUACACACUUGGAGCAUAAAAGGUUUGCACUGA

Nucleotides within stem loops subject to mutation are highlighted red. Inserts are highlighted blue.