The HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly

Nikesh Patel [*], Simon J. White [*], Rebecca F Thompson, Richard Bingham[1], Eva U. Weiß[1], Daniel P. Maskell, Adam Zlotnick[2], Eric Dykeman[1], Roman Tuma, Reidun Twarock[1]✉, Neil A. Ranson✉& Peter G. Stockley✉.

Astbury Centre for Structural Molecular Biology, University of Leeds,
Leeds, LS2 9JT, UK.

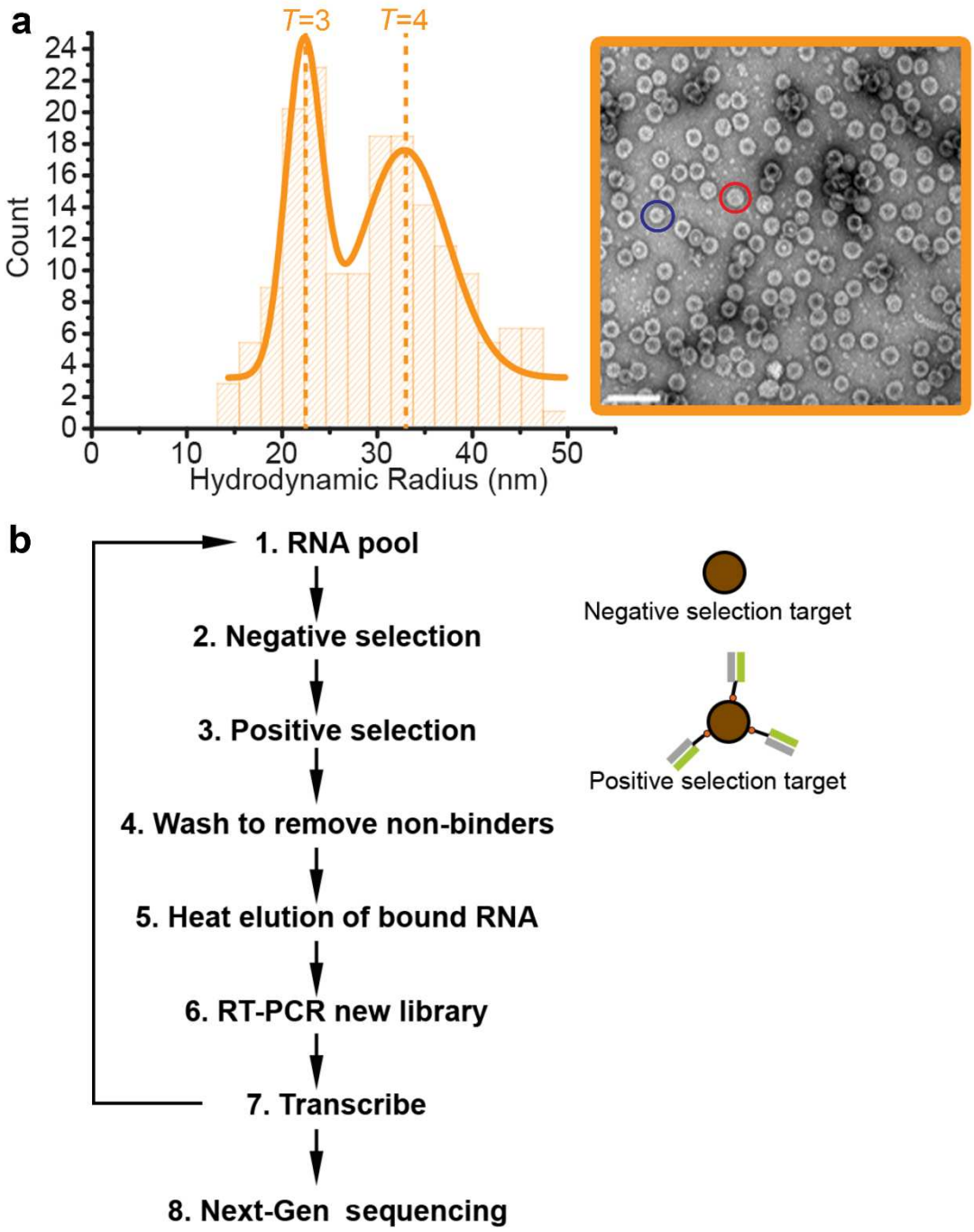
[1]Departments of Biology and Mathematics & York Centre for Complex Systems Analysis, University of York, York, YO10 5DD, UK

[2]Department of Molecular & Cellular Biochemistry, Indiana University, Bloomington, IN 47405, USA.

*These authors contributed equally to this work.

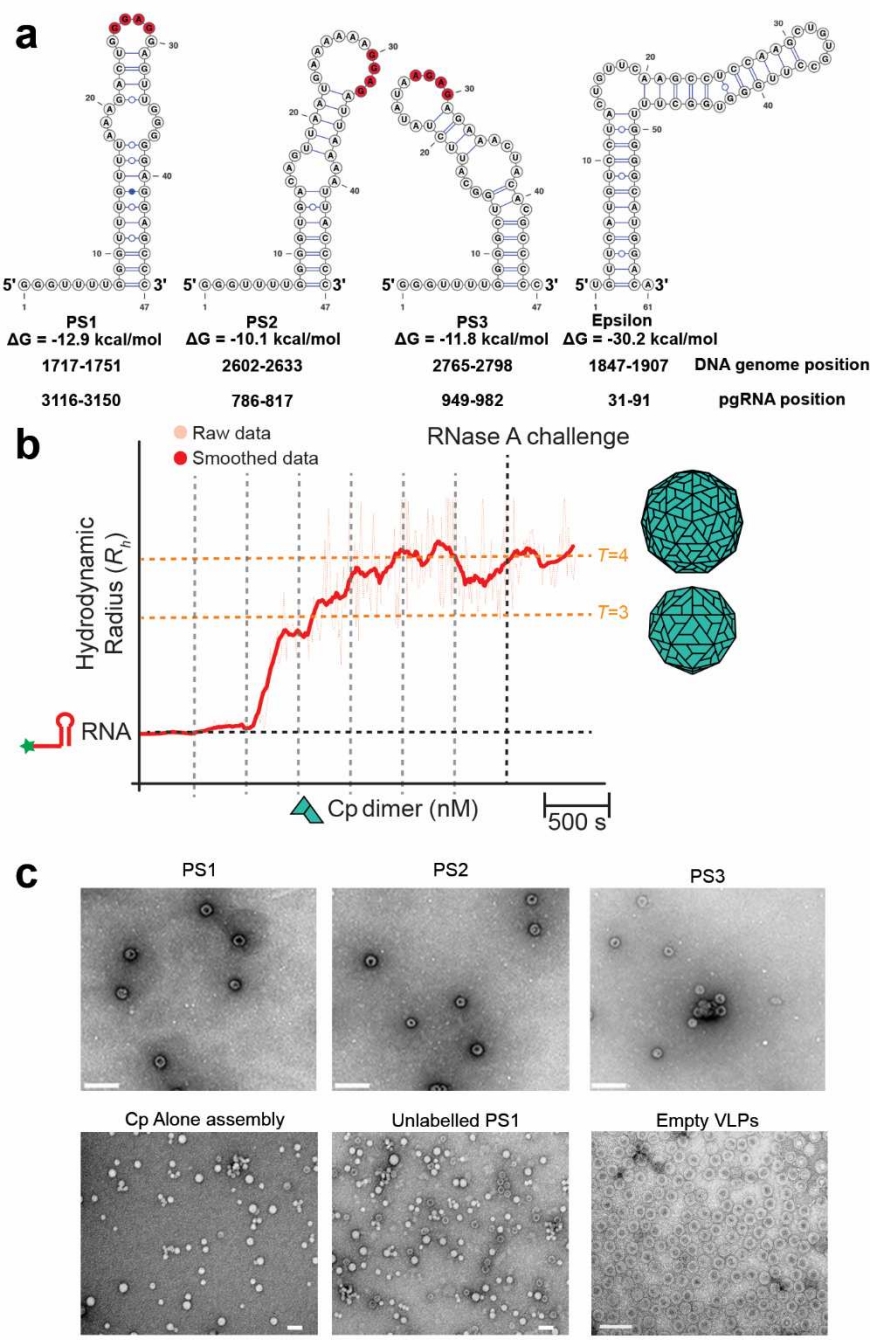
✉Joint communicating authors

Supplementary Figures



Supplementary Figure 1. Characterising HBV VLPs from *E.coli* and SELEX protocol. (a) Hydrodynamic radial distribution and negative stain EM image of Alexa Fluor-488 labelled HBV VLPs purified from *E.coli*. Integration of peak yields suggests a roughly 2:1 ratio of *T*=4 (red circle, 63%) and *T*=3 (blue circle, 37%) VLPs. Scale bar represents 100 nm. (b) SELEX protocol showing selection for aptamers with high affinity to HBV 185 Cp. HBV VLPs were immobilised onto carboxylic magnetic beads (brown circles) and dissociated into Cp dimers (green and grey rectangles) using guanidinium chloride. An RNA pool encompassing a random region (40N) was enriched for sequences with affinity for Cp by repeated cycles of binding to these beads, partitioning and amplification. Negative selections at each round used carboxylic acid beads which had been treated with NHS-EDC and inactivated with Tris. Stringency was increased after

round 5 by decreasing the number of positive beads by half and increasing the number of washes from 8 to 10. The reverse transcriptase-PCR products at the end of each round were analyzed by native PAGE to confirm the isolation of products for the next round of selection. The 10th round products were converted to DNA and sequenced.



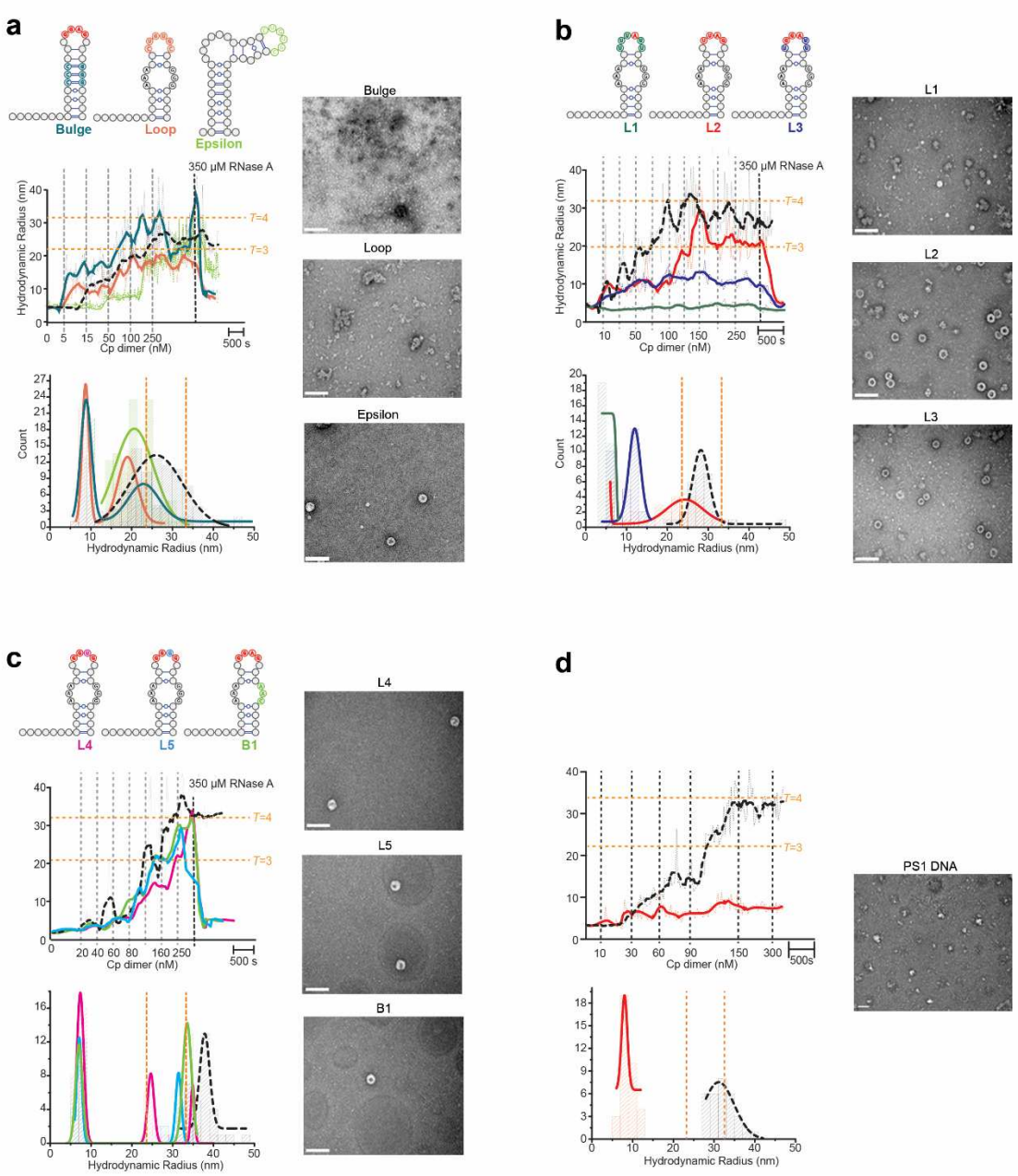
Supplementary Figure 2. PS oligo structures, example smFCS trace and EMs of PS containing VLPs.

(a) PS1-3 and ε secondary structures, made using VARNA software[1], were predicted in Mfold. Preferred sites were taken from the HBV genome, NC_003977.1, at positions: $PS1_{(1717\text{-}1751)}$, $PS2_{(2602\text{-}2633)}$ and $PS3_{(2765\text{-}2798)}$. In order to make them all the same length (47 nucleotides) to avoid effects of charge differences the following additions were made; PS1, 5' - GGGUUUUGG and CCC - 3'; PS2, 5' - GGGUUUUGGGG and CCCC - 3';

PS3, 5'- GGGUUUUGG and CCCC - 3'. The consensus motif RGAG is highlighted in red in each of the loops. The stability of each RNA fold, as predicted by Mfold, is shown below each structure.

(b) Example smFCS assay. R_h values for, fluorescently labelled RNAs are determined before and after Cp is titrated in at fixed time points (vertical dashed lines), allowing the R_h values to equilibrate after each step. The faint red trace represents real time, raw signal while the thick red line represents smoothed data. PS1 R_h initially climbs slowly, until a threshold Cp concentration, which triggers rapid assembly into a T=3 or T=4 VLP (R_h ~24-32 nm, orange dashed lines) as determined by measurements of Alexa-Flour 488 labelled HBV particles from *E.coli* (Supplementary Figure 1a). At the end of each titration, the complexes formed are challenged by addition of RNase A. An unchanged R_h is assumed to mean that the test RNA has been encapsidated in a closed VLP. The time scale on which this occurs is indicated in the bottom right.

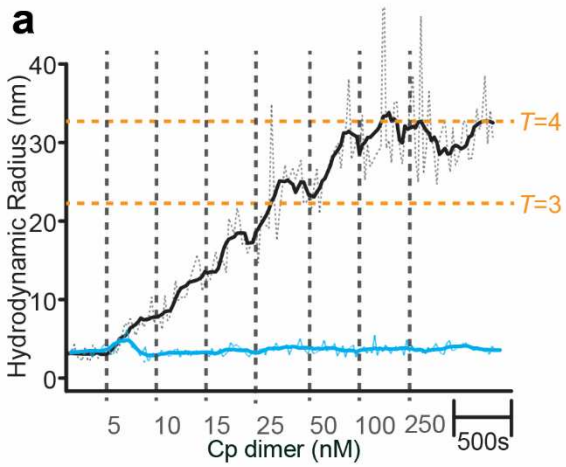
(c) TEMs from assembly reactions of PS1, 2, 3 and Cp alone and unlabelled PS1 in Figure 3. Large white particles in Cp alone and Unlabelled PS1 TEMs are latex beads. Also present is TEM from empty particle assembly described in Sup Table 2. These empty HBV particles were assembled at much higher concentrations of Cp (1.5 µM) and in the absence of RNA. Scale bars represent 100 nm.

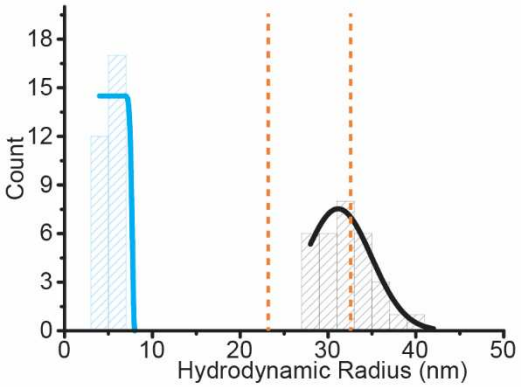


Supplementary Figure 3. smFCS assays of PS1 variants

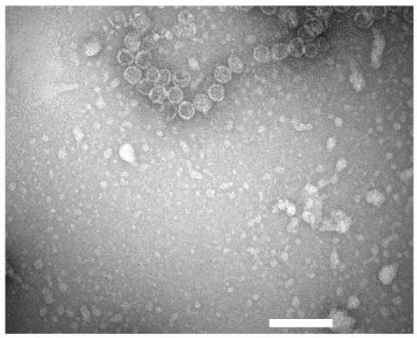
(a) smFCS assays of the PS1 variants (structures top left) and accompanying hydrodynamic radial distributions plotted in 2 nm bins and fitted with Gaussian peaks below, as colour coded in the key. 15 nM PS1 (black), PS1 loop mutant (dark orange) bulge mutant (dark blue) and epsilon (green) RNAs were tested for their ability to form VLPs under single molecule conditions. Vertical dotted lines indicate points of addition of Cp with the final concentrations shown in nM. Samples were allowed to equilibrate between additions. RNase A was added to check for correctly formed particles. Samples were taken prior to RNase A addition for analysis by TEM shown right, both here and throughout this figure. (b) - as (a) with RNA oligos PS1 (dashed black), L1 (green), L2 (red) and L3 (blue). (c) - as (a) with RNA oligos PS1 (dashed black), L4 (magenta), L5 (cyan) and B1 (light green). (d) as (a) with RNA oligos PS1 (dashed black) and DNA oligo PS1. Scale bars represent 100 nm. PS1 controls (dashed black) in each

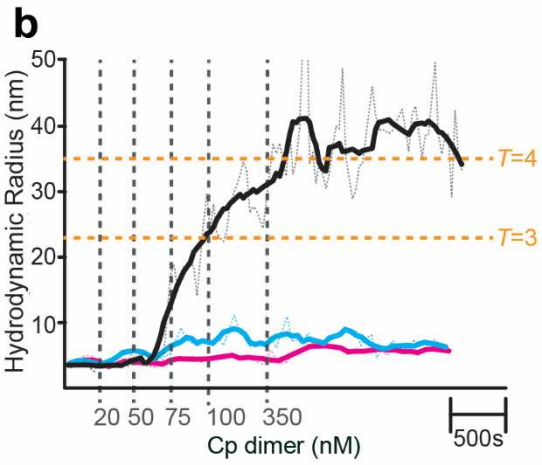
panel were repeated for individual batches of purified Cp, accounting for the variations in assembly efficiency seen. smFCS and TEM were repeated in triplicate.



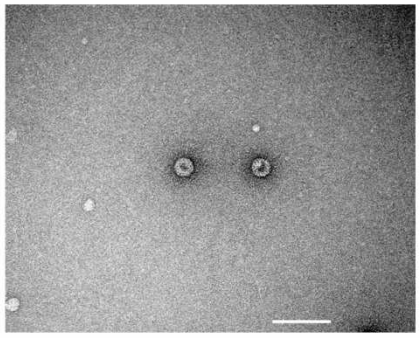


Cp 149

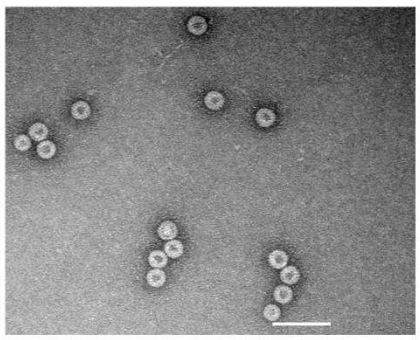


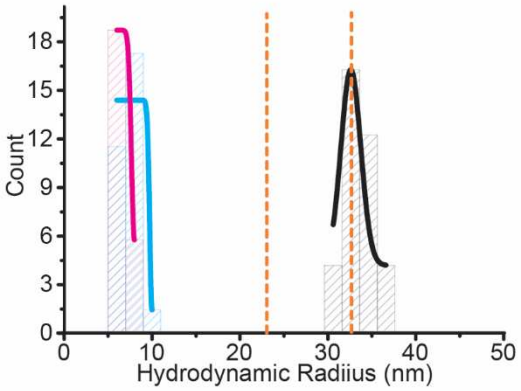


Kinase and Cp added to PS1

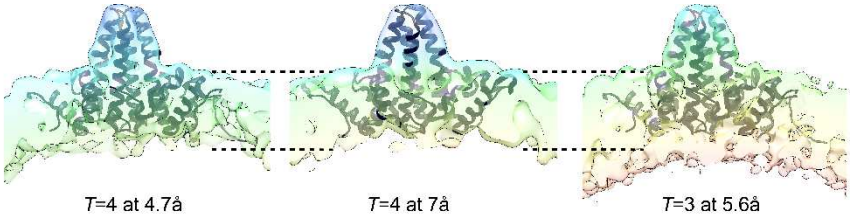


Kinase and Cp pre-equilibrated and added to PS1





Supplementary Figure 4. Role(s) of ARD and its charge on assembly.
(a) smFCS assays of 15 nM PS1 with Cp (grey) and Cp_{149} (cyan) and accompanying hydrodynamic radial distributions plotted in 2 nm bins and fitted with Gaussian peaks below. EM images of particles are shown (right). (b) as (a) with PS1 and Cp (black), kinase and Cp pre equilibrated and added to PS1 (cyan) and PS1 and Cp with kinase added simultaneously (magenta). TEMs are shown right. Scale bars represent 100 nm. PS1 controls (black) in each panel were repeated for individual batches of purified Cp, accounting for the variations in assembly efficiency seen. smFCS and TEM were repeated in triplicate.



Supplementary Figure 5. ARD structure in *T*=4 and *T*=3 VLPs.
Slabs (~30 Å thick) through the structures of the icosahedrally-averaged *T*=4 particle at 4.7 Å (left), the same *T*=4 structure low pass filtered to 7 Å (middle), and the *T*=3 particle at 5.6 Å (right). A Cp dimer is fitted into each. Even at a slightly lower resolution than the *T*=3 VLP, there is no equivalent density for the ARD in the *T*=4 VLP, confirming that it has different conformations in each particle.

Supplementary Tables

	Expected mass (Da)	**Observed mass (Da)**
SRPKΔ	45615.4	45614.7 ± 1.37
Phosphorylated Cp_{185}	21995.4	21995 ± 0.71
Cp_{185}	21395.3	21395.6 ± 0.86
Cp_{149}	16852.3	16851.7 ± 0.06

Supplementary Table 1: Masses of the different forms of Cp and kinase (SRPKΔ) used, as determined by ESI-MS mass spectrometry

	Fluorescence Polarisation		*Total Fluorescence*	
Sample	*-RNase*	*+RNase*	*-RNase*	*+RNase*
PS1 oligo	72	43.5	73637	74102
PS1 VLP	130	128	30187	33564
PS1 + empty VLP	52.8	15.5	69336	70672

Supplementary Table 2: Association of Alexa-Fluor-488 labelled PS1 with Cp. Anisotropy was used to determine if 15 nM of Alexa-Fluor-488 labelled RNA PS oligos can bind to, or enter, 125 nM of preformed shells of Cp. The latter were formed by reassembly in the absence of RNA at high concentration[2] (Supplementary Fig 2c). Fluorescence polarisation values are influenced by the mass of the dye-labelled species[3]. The polarisation value for PS1 oligo goes down following addition of RNase, as expected but remains unchanged when incorporated in VLPs assembled in the presence of the oligo. When labelled PS1 is added to the empty Cp VLP its fluorescence emission is unaffected, suggesting that it is not quenched, and it remains RNase sensitive confirming that it does not bind the outside of the protein shell or get internalised.

RNA Oligo	**Loop**	**Bulge**	**Assembly behaviour**	**Comment**
PS1	**GGGAGG**	**GGG**	**+++**	
L1	**UUUAUU**	**GGG**	**+ - -**	**Loop G's are important**
L2	**GUUAGG**	**GGG**	**+ - -**	**Loop G's are important**
L3	**UGGAUU**	**GGG**	**+ + -**	**Loop G's are important**
L4	**GGGUGG**	**GGG**	**+ + -**	**Loop A is important**
L5	**GGGGGG**	**GGG**	**+ + -**	**Loop A is important**
B1	**GGGAGG**	**AAC**	**+ + -**	**Bulge sequence/structure is important**

Supplementary Table 3: Sequence changes and corresponding assembly behaviour of PS1 variant oligonucleotides, L1-5 and B1. Assembly behaviour is indicated as follows, the first "+" indicates RNA-Cp binding, the second signifies formation of *T*=3/ *T*=4 sized species, and the third indicates RNase protection. "-" indicates failure in that assay.

Supplementary References

1. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25,** 1974–1975 (2009).
2. Porterfield, J. Z. *et al.* Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J. Virol.* **84,** 7174–7184 (2010).
3. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. Ch. 1. page 12. (Springer US, 2006).