

Published in final edited form as:

*J Mol Biol.* 2022 March 24; 434(10): 167557. doi:10.1016/j.jmb.2022.167557.

## Dysregulation of Hepatitis B Virus Nucleocapsid Assembly *in vitro* by RNA-binding Small Ligands

Nikesh Patel<sup>1,†</sup>, Fardokht Abulwerdi<sup>2,‡</sup>, Farzad Fatehi<sup>3,4</sup>, Iain W. Manfield<sup>1</sup>, Stuart Le Grice<sup>2</sup>, Jr. John S. Schneekloth<sup>2</sup>, Reidun Twarock<sup>3,4,5</sup>, Peter G. Stockley<sup>3,†</sup>

MF Summers

<sup>1</sup>Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

<sup>2</sup>Center for Cancer Research, National Cancer Institute, Frederick, MD 21702-1201, United States

<sup>3</sup>Department of Mathematics, University of York, York, YO10 5DD, UK

<sup>4</sup>York Cross-disciplinary Centre for Systems Analysis, University of York, York, YO10 5GE, UK

<sup>5</sup>Department of Biology, University of York, York, YO10 5DD, UK

### Abstract

RNA sequences/motifs dispersed across the genome of Hepatitis B Virus regulate formation of nucleocapsid-like particles (NCPs) by core protein (Cp) *in vitro*, in an epsilon/polymerase-independent fashion. These multiple RNA Packaging Signals (PSs) can each form stem-loops encompassing a Cp-recognition motif, -RGAG-, in their loops. Drug-like molecules that bind the most important of these PS sites for NCP assembly regulation with nanomolar affinities, were identified by screening an immobilized ligand library with a fluorescently-labelled, RNA oligonucleotide encompassing this sequence. Sixty-six of these "hits", with affinities ranging from low nanomolar to high micromolar, were purchased as non-immobilized versions. Their affinities for PSs and effects on NCP assembly were determined *in vitro* by Surface Plasmon Resonance. High-affinity ligand binding is dependent on the presence of an -RGAG-motif within the loop of the PS, consistent with ligand cross-binding between PS sites. Simple structure-activity relationships show that it is also dependent on the presence of specific functional groups in these ligands. Some compounds are potent inhibitors of *in vitro* NCP assembly at nanomolar concentrations. Despite appropriate logP values, these ligands do not inhibit HBV replication

This work is licensed under a [CC BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Correspondence to: Nikesh Patel; Peter G. Stockley.

**Correspondence to:** @FBSResearch @BindingLeeds (N. Patel), @AstburyCentre @MedResFdn (P.G. Stockley).

<sup>†</sup>Joint communicating authors.

<sup>‡</sup>Current address: Center for Drug Evaluation and Research (CDER), US Food & Drug Administration, Washington DC, USA.

### CRedit authorship contribution statement

**Nikesh Patel:** Investigation, Formal analysis, Conceptualization, Writing – original draft, Writing – review & editing. **Fardokht Abulwerdi:** Investigation. **Farzad Fatehi:** Investigation, Methodology. **Iain Manfield:** Methodology. **Stuart Le Grice:** Methodology, Supervision, Writing - review & editing. **John S. Schneekloth Jr.:** Methodology, Supervision, Writing - review & editing. **Reidun Twarock:** Conceptualization, Supervision, Writing - review & editing. **Peter G. Stockley:** Conceptualization, Supervision, Writing review & editing.

### Conflict of Interest Statement

The authors declare no conflicts of interest.

in cell culture. However, modelling confirms the potential of using PS-binding ligands to target NCP assembly as a novel antiviral strategy. This also allows for computational exploration of potential synergic effects between antiviral ligands directed at distinct molecular targets *in vivo*. HBV PS-regulated assembly can be dysregulated by novel small molecule RNA-binding ligands opening a novel target for developing directly-acting anti-virals against this major pathogen.

## Keywords

packaging-signal mediated assembly; directly-acting anti-virals; hepatitis B virus

---

## Introduction

As of 2016, the WHO estimated that over 2 billion individuals were infected with Hepatitis B Virus (HBV), which is currently the main risk factor for liver cancer. Its heavy annual death toll and resultant economic burdens arise within the cohort of ~260 million people who suffer from chronic infections,<sup>1,2</sup> for which there is currently no cure. These patients typically suffer cycles of asymptomatic liver inflammation, which over decades can lead to hepatocellular carcinoma, cirrhosis and death. Despite the availability of an effective vaccine,<sup>3</sup> it is not universally deployed, allowing ~1 million new infections to occur annually. Many of these cases are due to vertical transmission from mother to child, a route that frequently allows the infection to progress to the chronic state. Current clinical therapy for HBV infections targets the reverse transcriptase (RT) function of the virally-encoded polymerase (Pol).<sup>4,5</sup> Treatment with nucleot(s)ide analogues is not curative, is required throughout the lifetime of the patient, and only lowers the eventual death toll via liver disease by 2–4 fold.<sup>6</sup> Interferon therapy is also used as an adjuvant for the immune system but has had limited success in reducing HBV levels, and produces undesirable side effects.<sup>4,5</sup> Cp-binding allosteric modulators (CPAMs) that inhibit correct assembly of the NCP are also under development,<sup>7–9</sup> as are inhibitors of the RNase active site of Pol.<sup>10,11</sup> The WHO has declared a Global Challenge to make HBV a fully treatable disease by 2030.<sup>2</sup>

HBV is a para-retrovirus (Sup Figure 1), i.e. a dsDNA virus which initially packages a positive-sense, single-stranded (ss) ~3500 nt RNA, termed the pre-genomic RNA (pgRNA), into a  $T=4$  NCP.<sup>12–14</sup> NCPs with  $T=3$  architecture are also produced but whether they contribute to infection is not known. The pgRNA carries a 5' cap and a 3' poly-A tail,<sup>15,16</sup> and forms a complex with a molecule of the virally-encoded Pol bound to the *e* RNA stem-loop.<sup>17–20</sup> The Pol enzyme reverse transcribes the pgRNA, whilst degrading it, and then copies the ssDNA strand formed, all within the confines of the NCP.<sup>21–23</sup> Virion formation, budding and infection are complex processes,<sup>24</sup> resulting in formation of a relaxed, circular dsDNA genome (rcDNA) that is delivered to new host cells. This rcDNA is repaired by host enzymes, generating a covalently-closed circular DNA (cccDNA),<sup>25,26</sup> from which the pgRNA and sub-viral mRNAs for expression of HBV proteins are transcribed.<sup>12,20</sup> Persistence of the cccDNA ensures the longevity of HBV infection.

Given the initial requirement to package a ssRNA genome, we previously explored whether this virus assembles its NCP using the RNA Packaging Signal (PS)-mediated mechanism

that we identified for spherical ssRNA viruses from a number of different viral families<sup>27–33</sup>. These viral genomes encompass multiple, dispersed sequences, often in the form of stem-loops, that act collectively in binding CPs, thus regulating formation of the infectious virion<sup>27–29,31,32,34,35</sup>. The differing PS sites within a gRNA have differential affinities for their cognate CPs defining a preferred assembly pathway (a Hamiltonian path).<sup>33</sup>

RNA SELEX<sup>31</sup> against Cp from HBV strain NC\_003977.1, and bioinformatics analysis of the resulting Cp-binding aptamers, was used to identify 9 putative PSs that are conserved across the pgRNA sequences of 14 randomly selected HBV strains from GenBank.<sup>27</sup> All of these sites are highly conserved in each of these genomes, consistent with their functions as PSs, and the three PSs in strain NC\_003977.1 best matched by the aptamers are completely conserved (PSs1-3, Figure 1A). When presented within oligonucleotides, each triggers sequence-specific assembly of Cp dimers *in vitro* into mostly  $T=4$  NCPs. Sequence variants of homologous PS sites within genome length transcripts (gRNA) of strain JQ707375.1, which was not part of the previous analysis, also regulate efficient NCP assembly (Figure 1B–D) in the absence of the terminal modifications of the pgRNA and bound Pol.<sup>36</sup> PS1 is most important for NCP assembly regulation in these *in vitro* assays. A combination of cryo-EM reconstruction and RNA X-ray footprinting of the reassembled NCPs<sup>36,37</sup> confirms that the pgRNA-Cp interaction includes the PS1 Cp-recognition motif (–RGAG–), as expected for PS-mediated assembly.

Here we describe the identification of uncharged, small molecule, PS1-binding ligands, some with nanomolar affinities. Several of these are potent inhibitors of NCP formation *in vitro*, implying that its assembly regulation via gRNA-Cp interactions could be a drug target.

## Methods

### Cloning, expression and purification of HBV Cp dimer

HBV Cp was expressed from a pET28b plasmid in *E. coli* BL21(DE3) cells and purified as NCPs.<sup>31,38</sup> NCPs were dissociated into Cp dimers using 1.5 M guanidinium chloride as previously described.<sup>31,38</sup> The HBV Cp dimer concentration was determined spectrophotometrically ( $\epsilon_{280}$  of Cp dimer = 55,920 L mol<sup>-1</sup> cm<sup>-1</sup>) in a Nanodrop<sup>TM</sup> One. Fractions with an  $A_{260/280}$  ratio of ~0.65 or lower were used in reassembly assays. Absorbance values at 260 and 280 nm were corrected for light-scattering throughout, using the absorbance values at 310 and 340 nm, as previously described.<sup>39</sup>

### Preparation of genomic RNA

A wild-type HBV gRNA construct was assembled using a clone purchased from ATCC<sup>®</sup> (pAM6 39630<sup>TM</sup>, strain acc. no JQ707375.1). The pgRNA precursor sequence was copied in a modular fashion using PCR and cloned into the correct order within a pACYC184 vector, between the *BspHI* and *HindIII* sites using a Gibson reassembly<sup>®</sup> Master Mix, according to the manufacturer's protocol (New England Biolabs). Transcription of the RNA was carried out using a Highscribe T7 High-yield RNA synthesis kit (NEB), after linearisation of the DNA plasmid using *HindIII*. RNA was annealed by heating to 70 °C for 90 sec followed by cooling slowly to 4 °C in a buffer containing 50 mM NaCl, 10 mM HEPES and 1 mM DTT

at pH 7. Products were assessed using a 1% (w/v) denaturing formaldehyde agarose gel (Sup Figure 2). RNA concentration was estimated using  $A_{260}$  values ( $\epsilon_{260}$  of gRNA = 32, 249  $\text{mM}^{-1} \text{cm}^{-1}$ ).

### Small molecule microarray screening

SMM screening was performed as previously described.<sup>40</sup> Briefly, Corning GAPS II slides were functionalized with an isocyanate group using an established procedure.<sup>41</sup> Slides were printed using an ArrayIt NanoPrint LM60 microarrayer, and compounds were printed as 10 mM stock solutions in DMSO with two replicate spots. After printing, slides were incubated with fluorescently-labelled PS1 RNA purchased from Dharmacon in RNase-free 1 × PBS buffer (12 mM  $\text{NaH}_2\text{PO}_4$ , 137 mM NaCl, 3 mM KCl, pH 7.4). After washing, slides were imaged with an Innopsys 1100 scanner. Composite Z scores were generated for each compound. Hits were characterized as (1) having a Z score >3; (2) having an increase of Z-score of 3 relative to a buffer-incubated control; and (3) having a coefficient of variance (CV) of <120% for replicate spots.<sup>40</sup> Finally, hits were filtered for selectivity against all other RNAs screened previously in the SMM format in the Schneekloth laboratory.

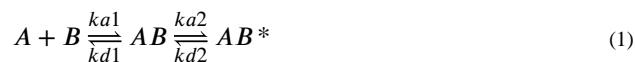
### Assaying ligand affinities for PS RNAs

**RNA preparation**—5'-Amino-labelled RNA oligonucleotides (Sup Table 1, Figure 3) were purchased from IDT and biotinylated covalently by rolling at room temperature for 4 h in 100 mM sodium borate buffer (pH 8.0), with a 100-fold molar excess of EZ-Link Sulfo-NHS-LC-LC-Biotin (Thermo Fisher). Biotinylated oligonucleotides were extracted from a 10% (v/v) denaturing acrylamide gel via a 'crush and soak' method. Gel slices corresponding to the desired oligonucleotides were excised and eluted three times into separate aliquots (100  $\mu\text{L}$ ) of 10 mM TE buffer over 3 h. Aliquots were combined and precipitated overnight using 50% (v/v) isopropanol, 0.3 M NaOAc and 1/100 volume RNA grade glycogen at  $-20^\circ\text{C}$ . Biotinylated, precipitated RNA was eluted into 100  $\mu\text{L}$  ddH<sub>2</sub>O.

### Determining ligand affinities using surface plasmon resonance (SPR)

Biotinylated RNA oligonucleotides were immobilised onto the surface of streptavidin-coated chips (GE Healthcare) in a Biacore T200 at a concentration of 100 nM. Buffer flow was set at 5  $\mu\text{L}/\text{min}$  over 15 min contact time, and the surface of the chip saturated with the RNA oligonucleotides, as determined by a plateau in the resulting refractive index change. Small molecular weight compounds stored in DMSO at a concentration of 10 mM, were used at initial concentrations of 75, 100, 150, 200 and 300  $\mu\text{M}$ . Further measurements at lower concentrations were taken where  $K_D$  measurements were lower than 75  $\mu\text{M}$ , dropping as low as 1 nM for compound #63. These were stored in a chilled compartment at  $9^\circ\text{C}$  within a skirted, 384-well plate (GE Healthcare) before being washed over the chip surface in a buffer containing 20 mM HEPES (pH 7.5), 250 mM NaCl, 2% (v/v) DMSO and 0.1% (v/v) Tween20, at a rate of 30  $\text{mL}/\text{min}$  for 6 min at  $37^\circ\text{C}$ , allowing for an association time of 1 min followed by dissociation over 5 min. Solvent corrections to account for variable concentrations of DMSO were performed according to manufacturer's instructions. Data were plotted with the best-fitting model as determined by the  $\chi^2$  values. All fits used the two-state model available in the Biacore T200 evaluation software, where A is the analyte

(compound), and B is the immobilised RNA. Complex formation between A and B results two kinetically distinct products, denoted as AB and AB\* below (GE Healthcare).



A two-state model (1) was used to fit the SPR curves, so that the association ( $k_{a1}$ ,  $k_{a2}$ ) and dissociation ( $k_{d1}$ ,  $k_{d2}$ ) rate constants of complexes AB and AB\* are given by:

$$\frac{dR_{AB}}{dt} = (k_{a1} \cdot C_A) \cdot (R_{\max} = R_{AB} - R_{AB^*}) - (k_{d1} \cdot R_{AB}) - (k_{a2} \cdot R_{AB}) + (k_{d2} \cdot R_{AB^*}) \quad (2)$$

$$\frac{dR_{AB^*}}{dt} = (k_{a2} \cdot R_{AB}) + (k_{d2} \cdot R_{AB^*}) \quad (3)$$

Eq. (3) defines the rate of formation of each complex, AB (2) and AB\* (3), directly from the SPR response curves obtained. R is the concentration of the complex (subscript) and  $C_A$  is the concentration of the analyte.

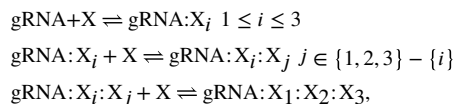
$$K_D = (k_{a1}/k_{d1}) \cdot (1 + k_{a2}/k_{d2}) \quad (4)$$

Dissociation and association constants from (2) and (3) were used to calculate an affinity constant (4).<sup>42,43</sup>

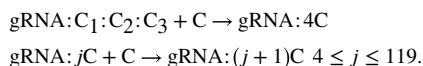
**HBV NCP reassembly**—178  $\mu$ L of 1.1 nM heat-annealed gRNA in a buffer containing 20 mM HEPES (pH 7.5), 250 mM NaCl, and 5 mM DTT was incubated in each well of a 96-well plate at room temperature for 30 min. 2  $\mu$ L of DMSO  $\pm$  compound was then added, and a further equilibration was performed for 30 min. Cp dimer in dissociation buffer (as above) was then titrated into the RNA using a Biomek 4000 liquid-handling robot (Beckmann Coulter), step-wise up to a ratio of 1200:1 (10 discrete 2  $\mu$ L solution aliquots, using the following Cp dimer concentrations (in brackets), were then added: 1 (100 nM), 10 (1  $\mu$ M), 25 (2.5  $\mu$ M), 75 (7.5  $\mu$ M), 120 (12  $\mu$ M), 240 (12  $\mu$ M), 480 (24  $\mu$ M), 720 (24  $\mu$ M), 960 (24  $\mu$ M) and 1200 nM (24  $\mu$ M) Cp dimer). The final volume in each well was 200  $\mu$ L, i.e. a cumulative volume of 19.2 mL/plate. The final concentrations of RNA and Cp were 1 nM and 1.2  $\mu$ M, respectively. Cp aliquots were calculated such that they reached a maximum of 10% of the final volume, maintaining the GuHCl concentration <0.15 M. Following incubation at room temperature for 1 h, the samples were pooled and concentrated to a final volume of 2 mL. The  $A_{260/280}$  ratio was measured using a Nanodrop™ One (ThermoFisher Scientific), and the RNA concentrations calculated using the corrected absorbance value at 260 nm. Each sample was split into 2, with one half treated with 1  $\mu$ M RNase A, and incubated overnight at 4 °C. After incubation, 5  $\mu$ L of each sample was visualised by negative stain electron microscopy (nsEM) to assess particle yield, shape and "completeness". The remaining samples were analysed by application to a TSK G6000 PWXL column (Tosoh), in a buffer containing 20 mM HEPES (pH 7.5), 250 mM NaCl and 5 mM DTT, attached to a SEC-MALLS system (ÄKTA Pure (GE Healthcare)

connected to an Optilab T-REX refractometer and miniDAWN<sup>®</sup> multiple angle laser light-scatterer fitted with a Wyatt QELS DLS module (Wyatt Technology). Light-scattering peaks were collected and concentrated to ~1 ml, where their RNA content, was estimated using their  $A_{260}$  relative to the starting material.

**Computational modelling of ligand inhibition of NC assembly**—Assembly kinetics were modelled via a set of reactions adapted from our recent intracellular model for HBV infection,<sup>44</sup> that includes the roles of the three evolutionarily conserved packaging signals (PS1-PS3) in the gRNA during virion assembly.<sup>31</sup> We assume that all three PSs interact with Cp dimers and PS-targeting compounds according to the following reactions:



where X denotes either a Cp dimer (C) or a compound (d), and an index refers to the PS site where binding occurs. For example, gRNA :C<sub>1</sub> : C<sub>2</sub> : d<sub>3</sub> indicates C bound to PSs 1 and 2, and a drug molecule to PS3. The gRNA :C<sub>1</sub> : C<sub>2</sub> : C<sub>3</sub> complex will then recruit 117 Cp dimers at a rate  $\kappa$  to build an intact particle via recruitment of further CP dimers:



In order to model malformations caused by drug molecules, we introduce reactions in which the number of Cp dimers to be recruited is limited to  $n_c$ . The latter is estimated based on the experimentally determined hydrodynamic radii ( $R_h$ ) of the malformed particles (Table 2). In particular in the compound-free case we have  $R_h = 18.2\text{nm}$  and each nucleocapsid contains 120 Cp dimers. As the surface of a sphere is equal to  $4\pi r^2$ , we assume that the number of Cp dimers is in a good approximation proportional to  $R_h^2$ . Therefore, we use  $n_c = 120 \times \left(\frac{R_h}{18.2}\right)^2$  for the average number of Cp dimers in a malformed particle. For example,  $n_c \approx 114$  for  $R_h = 17.7\text{ nm}$ . A Gillespie algorithm was used to perform stochastic simulations of the system, resulting in the time-course of virus build up in the absence and presence of the drug molecules.

**Parameter value estimation**—178  $\mu\text{l}$  of 1.1 nM gRNA in the drug-free scenario corresponds to  $1.179 \times 10^{11}$  copies of gRNA. The simulation mimics the experimental protocol for Cp dimer titration with step-wise addition in 10 discrete steps with 10 minutes intervals up to a ratio of 1200:1. The numbers of Cp dimers added at consecutive steps were;  $1.2 \times 10^{11}$ ,  $9 \times 1.2 \times 10^{11}$ ,  $15 \times 1.2 \times 10^{11}$ ,  $50 \times 1.2 \times 10^{11}$ ,  $45 \times 1.2 \times 10^{11}$ ,  $120 \times 1.2 \times 10^{11}$ ,  $240 \times 1.2 \times 10^{11}$ ,  $240 \times 1.2 \times 10^{11}$ ,  $240 \times 1.2 \times 10^{11}$ , and  $240 \times 1.2 \times 10^{11}$ . After the equivalent of one hour incubation time, assembly reactions are started with recruitment rate  $\kappa$  equal to  $10^6 \text{ M}^{-1} \text{ s}^{-1}$ <sup>45</sup> and PS binding affinity for Cp of 4 nM.<sup>44</sup> As 1  $A_{260}$  Unit = 40  $\mu\text{g/ml}$ , and assuming a volume of 100  $\mu\text{l}$  and an average weight of a nucleocapsid of 4 MDa (6.6422 x

$10^{-18}$  g),<sup>46</sup> we obtain  $n_p = 0.04 \times OD \times \frac{1}{6.6422 \times 10^{-18}} \times 10^{-4} = 6 \times 10^{11} \times OD$  for conversion of OD values to the number of assembled particles ( $n_p$ ). Since after nuclease treatment (160 minutes post the experiment start)  $A_{260}$  is 0.12, the number of fully formed nucleocapsids is equal to  $7.2 \times 10^{10}$ . We fitted the forward rate of Cp dimers for binding to PSs ( $f_{nuc}$ ) to match this value, giving us  $336 \text{ M}^{-1} \text{ s}^{-1}$ .

In the presence of 1 nM of Compound #63,  $A_{260}$  is 0.028 after nuclease treatment. Using the formula above,  $n_p \approx 1.7 \times 10^{10}$ . We explored different forward rates for the binding of Compound #63 to PSs ( $k_f$ ). For  $k_f < 5 \mu\text{M s}^{-1}$ , the number of fully formed particles is  $> 1.7 \times 10^{10}$ . Increasing  $k_f$  from 5 to  $10 \mu\text{M}^{-1} \text{ s}^{-1}$  reduces the number of nucleocapsids to  $1.4 \times 10^{10}$ . However, further increasing  $k_f$  does not reduce the number of nucleocapsids. We therefore show results for  $k_f = 10 \mu\text{M}^{-1} \text{ s}^{-1}$ , and use this value for other compounds, which is consistent with an estimate of  $1 - 10 \mu\text{M}^{-1} \text{ s}^{-1}$  for other molecules.<sup>47</sup> This model, combined with a post HBV-infection intercellular model,<sup>48</sup> allows the impacts of various antiviral ligands to be estimated without recourse to animal experiments. In particular, it allows us to look at the effects of synergy between DAAs.

**Fluorescence anisotropy NCP assembly assay**—178  $\mu\text{L}$  of 1.1 nM heat-annealed gRNA or 16.5 nM heat-annealed PS1(NC\_003977.1) in a buffer of 20 mM HEPES (pH 7.5), 250 mM NaCl and 5 mM DTT was incubated in wells of a 96 well plate at room temperature for 30 min. Each nucleic acid was fluorescently labelled at the 5' end using an AlexaFluor SDP ester as previously described.<sup>31</sup> 2  $\mu\text{L}$  of DMSO  $\pm$  compound was added, and a further equilibration was performed for 30 min. Cp dimer in dissociation buffer (as above) was then titrated into the RNA using a Biomek 4000 liquid-handling robot (Beckmann Coulter), step-wise up to a ratio of 1200:1 (10 discrete 2  $\mu\text{L}$  solution aliquots, using the following Cp dimer concentrations (in brackets), were then added: 1 (100 nM), 10 (1  $\mu\text{M}$ ), 25 (2.5  $\mu\text{M}$ ), 75 (7.5  $\mu\text{M}$ ), 120 (12  $\mu\text{M}$ ), 240 (12  $\mu\text{M}$ ), 480 (24  $\mu\text{M}$ ), 720 (24  $\mu\text{M}$ ), 960 (24  $\mu\text{M}$ ) and 1200 nM (24  $\mu\text{M}$ ) Cp dimer). RNA only controls were performed, with dissociation buffer added without Cp dimer. The final volume in each well was 200  $\mu\text{L}$ . Fluorescence anisotropy was measured after each addition using a POLARstar OMEGA (BMG Labtech, Aylesbury). The final concentrations of RNA and Cp were 1 nM and 1.2  $\mu\text{M}$ , respectively. Cp aliquots were calculated such that they reached a maximum of 10% of the final volume, which keeps the GuHCl concentration  $< 0.15 \text{ M}$ . Each well was then treated with 1  $\mu\text{M}$  RNase A, and incubated at room temperature for 1 hour, after which a final fluorescence anisotropy reading was taken. These values were normalised with respect to the initial reading, and the anisotropy change calculated as the final RNase treated sample – the starting fluorescence anisotropy of the RNA.

## Results

### Identification & characterization of PS1 binding compounds

A 47-nt long oligonucleotide, capable of folding into a stem-loop encompassing the NC\_003977.1 PS1 sequence (Figure 1A), was 5'-labelled with the red fluorescent dye (AlexaFluor 647 for SMM screening. A library of  $\sim 20,000$  drug-like molecules covalently-linked onto an isocyanate-functionalized glass surface<sup>40</sup> was incubated with a solution of

labelled RNA. A second copy of the same slide was incubated with the buffer used to anneal the RNA. These signals were used to correct for auto-fluorescence by immobilised compounds. After two buffer washes, slides were scanned at 635 nm and PS1-binding compounds identified, as described in Methods, identifying 72 ligands that bind HBV PS1 (Sup Figure 3).

Sixty-six, non-immobilized versions of these PS1-binding compounds are commercially-available. These were purchased, resuspended in DMSO (to 10 mM), and stored frozen until diluted for binding assays. Each compound was numbered (1, 2, 3, etc.) based on their relative positions in the initial SMM array. SPR<sup>49,50</sup> was used to determine the binding affinities of each ligand for a series of RNA oligonucleotides, including the PS homologues from both NC\_003977.1 & JQ707375.1 (Figures 2B and 3); as well as sequence/structure variants. These latter include NC\_003977.1 PS2 & PS3, bulge & loop variants of the JQ707375.1 PS1 (PS Bulge and PS Motif, respectively), and an unrelated stem-loop ( PS) lacking an –RGAG-motif and the epsilon stem-loop (Figures 2 and 3). 5'-biotinylated RNAs encompassing these sequences were immobilized onto streptavidin-coated sensorchips and their affinities and stoichiometries for the test compounds determined.

The apparent association ( $k_a$ ) and dissociation ( $k_d$ ) rate constants were calculated from their sensorgrams (Methods). The best fits to the data are with the two-state binding model within the T200 Biacore Evaluation Software (Table 1, Figure 2, Sup Table 2 and Sup Figure 4). Most compounds possess higher affinities for the selection target, PS1, although roughly a third bind better to the PS2 or PS3 oligonucleotides. These results are consistent with the functional conservation of the PSs despite sequence/structural changes. The worst binders were compounds 27, 54 and 55, which have affinities in the millimolar range. All compounds bound preferentially to PSs 2 and 3 compared to negative controls, implying recognition of a common feature. Only 5 of the 66 compounds tested show any binding to the negative control RNAs with affinities in the high millimolar range (compounds 1, 4, 25, 49 and 50, data not shown). The affinities of the compounds ( $K_D$ ) for PS sites range from nanomolar to millimolar. The RU changes on binding are consistent with formation of complexes with a stoichiometry of 2:1 ligand: PS RNA for the highest affinity binders, consistent with the best-fitting model applied to analyze the data. Lower affinity binders appear to bind less specifically with apparent saturations ~5:1 ligand: PS RNA.

**Inhibition of NCP assembly**—To test whether ligand binding by the PSs inhibits assembly of the HBV NCP *in vitro*, we used an established assay with a genome length transcript gRNA (~3400 nt), based on strain JQ707375.1. This RNA lacks the terminal modifications but still re-assembles *in vitro* to nuclease-resistant, mostly  $T=4$  NCPs.<sup>36</sup> At low nanomolar concentrations, assembly of NCPs is dependent on sequence-specific interactions between Cp and gRNA,<sup>31,36</sup> i.e. it mimics *in vivo* assembly outcomes.<sup>32,51</sup> The titration of Cp into this gRNA was automated,<sup>36</sup> allowing low concentration reassembly reactions to be carried out in 96-well plates. The products were then pooled and concentrated before characterization. Importantly, for the ligands tested here their affinities for PS1 from strain NC\_003977.1, used for ligand identification, and JQ707375.1, used for *in vitro* reassembly are very similar (Table 1), implying that these PSs share the molecular features bound by the smaller molecules.



To assess the inhibitory effects of PS-binding ligands, small-scale reactions (90  $\mu$ L) at nanomolar concentrations were performed, titrating 1.2  $\mu$ M Cp in 10 steps over a 2 h period into a solution of gRNA transcript (1.1 nM, final concentration 1 nM). Ligands were added (final concentration 10  $\mu$ M, unless stated otherwise) to the gRNA solution prior to Cp titration, allowing the full range of ligand affinities to be tested. Pooled samples from 96-well plates were split into two aliquots. RNase A was added to one of these to degrade unencapsidated gRNA and then samples of each aliquot were visualized by negative-stain electron microscopy (nsEM) (Figure 4). The remainder of each aliquot was then fractionated by size-exclusion chromatography, monitoring the outflow via a multiple-angle, laser light-scattering detector (SEC-MALLS). These steps allow the integrity, shape, and hydrodynamic radius ( $R_h$ ) of assembled particles to be assessed, and they provide an estimate of assembly efficiency (Figure 4). Peaks eluting from the gel filtration column were re-concentrated to determine their RNA and Cp content via their corrected  $A_{260/280}$  ratios (Table 2). Note, DMSO at the concentration used here has no effect on reassembly. A dominant, single symmetrical light-scattering peak (red) elutes from the column at  $\sim$ 8-8.5 mLs post-injection. These NCPs have a  $R_h$  of 18.2 nm (Table 2), slightly smaller than that of NCPs produced by Cp expression in *E.coli* in the absence of gRNA (19.4 nm).<sup>36</sup> Their indifference to nuclease treatment (Figure 4, Table 2) implies that the assembled Cp shells are complete.

Compound #3 has one of the lowest affinities for PS1 ( $K_D \sim$  210  $\mu$ M) and binds PS3 poorly. However, it binds PS2 reasonably tightly ( $K_D \mu$  300 nM), implying that this PS will be fully saturated at the test concentration. Reassembly in the presence of Compound #3 [10  $\mu$ M] reveals the formation of largely nuclease-resistant NCPs (Figure 4A) eluting in the same position as those from ligand-free controls. Note, the filter used to concentrate samples removes unassembled Cp from the traces. The  $A_{260/280}$  ratio of this peak is also similar to the control, implying that Compound #3 has very little effect on NCP assembly, consistent with reassembly in the presence of gRNA transcripts lacking PS3.<sup>36</sup> However, the yield of NCP is reduced ( $>$ 5-fold) in the presence of Compound #3 compared to the uninhibited control. This level of variation is striking (Table 2) and cannot be accounted for unless the ligand behaves as an assembly inhibitor. EM images offer support for this conclusion. Prior to nuclease treatment, most reassembled particles lack the smooth circular appearance of the uninhibited control, and stain penetration is variable. In addition, many particles appear to have a unique, bright structural feature on their peripheries (see arrowheads on the micrograph, Figure 4A). These mostly disappear with nuclease treatment implying that they are due to incompletely protected gRNA.

Although the impact of Compound #3 appears limited, this result shows that it partially inhibits correct encapsidation, consistent with the drop in NCP yield.

Compound #63 has the highest affinity for PS1 ( $K_D \sim$  19 nM), the most important PS for regulation of *in vitro* NCP formation.<sup>36</sup> Consistent with this, reassembly in the presence of this ligand shows the strongest inhibitory effects. Prior to nuclease treatment the NCPs formed have the same apparent  $R_h$  as the control, with a yield similar to that obtained with Compound #3. However, nuclease treatment degrades the majority of the NCPs formed, leaving only minor peaks on the gel filtration profile. EMs of the NCPs formed in the presence of Compound #63 are strikingly different than the control. Post-RNase treatment,

most particles are heterogeneous and heavily stain-penetrated, and have formed large clumps with relatively few remaining as separate shells of the correct size (Figure 4B). Absorption values confirm these interpretations (Table 2). Similar inhibitory effects were also observed with Compounds #10 and #46, both of which have intermediate affinities for PS sites (Table 1, Figure 4C and D). Both compounds result in low yields of nuclease-sensitive, particulate material that elutes later than the NCP peak in the uninhibited control. In each case there are multiple additional peaks, implying that assembly is significantly dysregulated by these PS-binding compounds.

To explore the concentration dependence of these effects the final concentration of Compound #63 was serially diluted 10-fold to a final concentration of 1 nM, i.e. until it was equimolar with the gRNA assembly substrate (Figure 5; Table 3). Nuclease-resistant NCP yields increase from ~15% at the highest ligand concentration to ~90% of the control value at the lowest concentration. Even at 1 nM ligand concentration, i.e. well below the apparent affinities for all three PSs (Table 1), some assembly dysregulation occurs. Reductions in NCP yield are accompanied by significant changes in the appearance of the samples in nsEM (Figure 5). At the lowest ligand concentration (1 nM) particles are largely separate with circular cross-sections, although most of these show stain-penetration with some small variations in apparent diameter. At higher ligand concentrations, they become much less regular and start to clump together, and this tendency increases with the level of assembly inhibition. Binding of Compound #63 to the RNA PSs, which the SPR data suggests occurs at all three sites, appears to be the direct cause of these effects. Since clumping occurs it suggests that even when reassembled at low concentration there is a tendency for Cp to bind additional gRNAs. That would be consistent with the appearance of separate particles after nuclease treatment.

**Structure activity relationship (SAR) of Compound #63**—The PS-binding ligands possess many different chemophores. To understand which structural features of Compound #63 contribute to its ability to bind PS sites tightly, a structure—activity relationship (SAR) of its pendent functional groups was carried out (Figure 6, Sup Figure 5). Five commercial ligand variants were used to explore the role(s) of the halogens in the C6 ring and the amino group on the C5 ring. There is a >4000-fold drop in affinity for PS1 upon removal of the latter, highlighting it as a key feature for RNA recognition. Removal of both halogens appears fully compensated by inclusion of an aldehyde instead of chorine. All other variants that including swapping the positions of chlorine and fluorine, or the substitutions and removals, cause 200–1000-fold drops in PS1 affinity. Binding of these ligand variants to JQ707375.1 PS1 lacking a bulged stem (PS1 Bulge) (Figure 2) seem only modestly affected by these ligand substitutions (Figure 6). In contrast, removal of the Cp-recognition loop motif (PS1 Motif) leads to significant drops in affinity. These results suggest that the primary ligand binding site is the Cp-recognition motif in the stem-loop.

These experiments suggest that the PS-binding ligands inhibit Cp-binding to HBV gRNA, and thus NCP assembly regulation. We therefore explored whether these compounds would inhibit HBV infections in cell culture. Although we did see examples of such effects with some of the ligands (Dorner, *pers. comm.*), they were not reproducible at concentrations up to 100  $\mu$ M. This is surprising given the known *logP* values of the compounds (Table

1).<sup>52,53</sup> However, by preventing the formation of NCPs *in vitro*, the ligands identified here possess the properties required for a directly-acting antiviral (DAA) for HBV. In addition since they target multiple sites across the pgRNA. Since these are highly unlikely to undergo simultaneous mutation and also relatively conserved across strain variants, they represent a stable and novel molecular target.<sup>54</sup> The preliminary SAR data reported above can be used by synthetic groups to identify inhibitors with appropriate *in vivo* characteristics. In order to facilitate such studies the *in vitro* NCP assembly assay has been adapted to a semi-high-throughput, multi-well plate format. This assay monitors changes to the anisotropy of dye-labelled PS-encompassing oligonucleotides during assembly (Sup Figure 6).

**Modelling the anti-viral effects of PS-binding compounds**—Given the urgent clinical need for additional HBV inhibitors, we also explored the effects of PS binding ligands on HBV replication using an intracellular computational model. This is useful because there are very few susceptible organisms for HBV infection, other than primates. Ethical considerations therefore inhibit widespread testing of synergic effects between potential DAAs, whereas these can be assessed readily in the mathematical model.

The model used<sup>44</sup> describes different aspects of the viral life cycle in the context of several discrete reactions (Sup Figure 1), describing processes from viral entry to envelopment, thus facilitating the comparison of different antiviral strategies. Previously, virion release kinetics from a single infected cell were modelled both in the absence of therapy, and in the presence of different antiviral strategies, including the PS-targeting compounds analysed here. These compounds were<sup>44</sup> comparable in performance to other assembly inhibitors, such as the Cp allosteric modulators (CAMs), currently in clinical trials in chronic HBV patients.<sup>55</sup>

Here, we focus on modelling the impacts of the assembly inhibition observed *in vitro* on this full cellular model, (Sup Figure 1, green highlights) enabling direct comparison of the effects of various compounds. The outcomes of this reduced model, directly inform the full cellular model, as well as within host models of a viral infection<sup>56</sup> that also take the host immune response into account. The reaction kinetics of NCP assembly around gRNA were modelled<sup>44,57</sup> to mimic the conditions used in the titration assay described here (see Methods). Figure 7 shows the predicted time course of intact particle formation, without ligand (black) and for different concentrations of Compound #63 (1 nM, blue; 5 nM, red; 10 nM magenta). In the absence of inhibitory ligands, the number of complete particles 160 min post-initiation is  $7.2 \times 10^{10}$ . This decreases to  $1.4 \times 10^{10}$ ,  $1.5 \times 10^7$ , and  $6.4 \times 10^3$  in the presence of 1, 5 and 10 nM of Compound #63, respectively.

A comparative analysis of the effects of the other compounds, at concentrations of 1, 3 and 30 nM was also performed, assuming a gRNA concentration of 1 nM (Table 4). For Compounds #3, #10 and #46, when going from 1 to 30 nM ligand we see modest drops in particle numbers of 2, 5 and  $4 \times 10^{10}$ , respectively. However, at 30 nM, Compound #63 is predicted to inhibit particle formation completely (Table 4).

## Discussion

As expected for PS-mediated assembly regulation,<sup>31,36</sup> individual PSs vary in their importance for regulation of HBV NCP assembly *in vitro*.<sup>35</sup> PS1 is the major regulatory requirement for assembly around the genome length ssRNA transcript used here. Disruption of the PS1 Cp-recognition site by mutation of 6/3400 nts leads to severe inhibition of NCP assembly *in vitro*.<sup>36</sup> High-affinity PS1 sequence binding ligands, are also potent assembly inhibitors of NCPs *in vitro*, opening up a novel antiviral target.<sup>44</sup>

These data are the first example of DAAs targeting HBV PS-mediated NCP assembly. Other DAA treatments for HBV infection are being developed, such as CAMs. CAMs have been shown to disrupt HBV NC assembly at micromolar concentrations, and are based on the initial discovery of heteroaryldihydropyrimidine (HAP) compounds.<sup>58–62</sup> These compounds show early promise in clinical trials<sup>60,63–66</sup> and it will be interesting to see if the potential synergy with the anti-PS-mediated DAAs probed here by modelling can be made to work *in vivo* in cells. Development of drugs targeting multiple distinct steps of the HBV life cycle may be required to achieve an effective cure, as it has been with HIV.<sup>67</sup> The mathematical modelling of infection is important in such cases because it allows exploration of conditions that might be difficult to achieve experimentally.

There are currently few clinical drugs that target RNA. Risdiplam, a small molecule targeting RNA for treatment of spinal muscular atrophy, has however recently gained FDA approval.<sup>68</sup> This development is an important milestone, since the human transcriptome is considerably larger than the proteome, which contains most current drug targets. Targeting RNA with oligonucleotides is established but using small ligands is advantageous because of their easy administration.<sup>69</sup> Previously, anti-HIV ligands that prevent its assembly by binding to the assembly initiation signal in its gRNA were isolated, and shown to inhibit gRNA incorporation.<sup>70,71</sup>

Here we show that it is possible to dysregulate HBV NC assembly *in vitro* using drug-like ligands that bind to the RNA PSs of its pgRNA. The best of these inhibitors binds with nanomolar affinity and SAR (Figure 6) suggests that these are good starting points for further development of directly-acting anti-viral drugs (DAAs). The logP values of these ligands are such that they should gain access to the cytoplasm, where NC assembly occurs, with ease. Understanding why the *in vitro* inhibitory effects are absent in cell culture is not trivial. Nevertheless, the data presented here describe the first example of isolation and characterisation of ligands and their effects *in vitro*, for an assembly regulatory element, i.e., the PSs of the HBV pgRNA.

Since we<sup>28,29</sup> and others<sup>72</sup> have shown that ssRNA viruses, including major human pathogens in the picorna- and alphaviruses mostly lacking antivirals or vaccines,<sup>28–34</sup> also use this form of assembly regulation, the approach described here to identify novel DAAs could be useful. The ease of the selection of these ligands is proof of principle that potential DAAs for virions that use PS-mediated assembly regulation could be rapidly identified. Comparison of strain variants in these cases suggests that the multiple PSs regulating virus

assembly are highly conserved. As drug targets this makes them resilient to error-prone genome replication.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This paper is dedicated to the memory of our colleague Professor Marcus Dorner, Imperial College, London, who passed away in 2019.

We thank Prof. Adam Zlotnick, Indiana University for gifts of an HBV Cp expression clone and insights on soluble Cp preparation. We also thank Professor John Tavis, Pennsylvania State University, for attempts to assay inhibitory effects of these PS-binding ligands *in vivo* (cell culture); the Medical Research Foundation for the award of a career development grant to NP; and the UK MRC for previous grant funding to study HBV assembly (MRF-044-0002-RG-PATEL & MR/N021517/1, respectively). RT & PGS thank The Wellcome Trust (Joint Investigator Award Nos. 110145 & 110146 to PGS & RT, respectively) for funding. We also acknowledge the financial support of The Trust of infrastructure and equipment in the Astbury Centre, University of Leeds (grants 089311/Z/09/Z; 090932/Z/09/Z & 106692), and the support of the Astbury Biostructure Facility by the University of Leeds. RT acknowledges additional funding via an EPSRC Established Career Fellowship (EP/R023204/1) and a Royal Society Wolfson Fellowship (RSWF\R1\180009). JS, FA and S Le G were supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

## Data Availability

Data will be made available on request.

## Data availability

Data available upon request. Requests should be directed towards NP or PGS.

## References

1. Mathers BM, Degenhardt L, Phillips B, Wiessing L, Hickman M, Strathdee SA, Wodak A, Panda S, et al. Injecting Drug. Global epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. *Lancet*. 2008; 372: 1733–1745. [PubMed: 18817968]
2. World Health Organization. Combating Hepatitis B and C to Reach Elimination by 2030. World Health Organization; 2016. 1–16.
3. Murray K, Bruce SA, Wingfield P, van Eerd P, de Reus A, Schellekens H. Protective immunisation against hepatitis B with an internal antigen of the virus. *J Med Virol*. 1987.
4. Tillmann HL. Antiviral therapy and resistance with hepatitis B virus infection. *World J Gastroenterol*. 2007.
5. Lok AS, Lai CL, Leung N, Yao GB, Cui ZY, Schiff ER, Dienstag JL, Heathcote EJ, et al. Long-term safety of lamivudine treatment in patients with chronic hepatitis B. *Gastroenterology*. 2003; 125: 1714–1722. [PubMed: 14724824]
6. Li Q, Lomonosova E, Donlin MJ, Cao F, O’Dea A, Milleson B, Berkowitz AJ, Baucom JC, et al. Amide-Containing  $\alpha$ -Hydroxytropolones as Inhibitors of Hepatitis B Virus Replication. *Antiviral Res*. 2020; 177 104777 [PubMed: 32217151]
7. Packianathan C, Katen SP, Dann CE, Zlotnick A. Conformational Changes in the Hepatitis B Virus Core Protein Are Consistent with a Role for Allostery in Virus Assembly. *J Virol*. 2010.
8. Venkatakrishnan B, Katen SP, Francis S, Chirapu S, Finn MG, Zlotnick A. Hepatitis B Virus Capsids Have Diverse Structural Responses to Small-Molecule Ligands Bound to the Heteroaryldihydropyrimidine Pocket. *J Virol*. 2016.

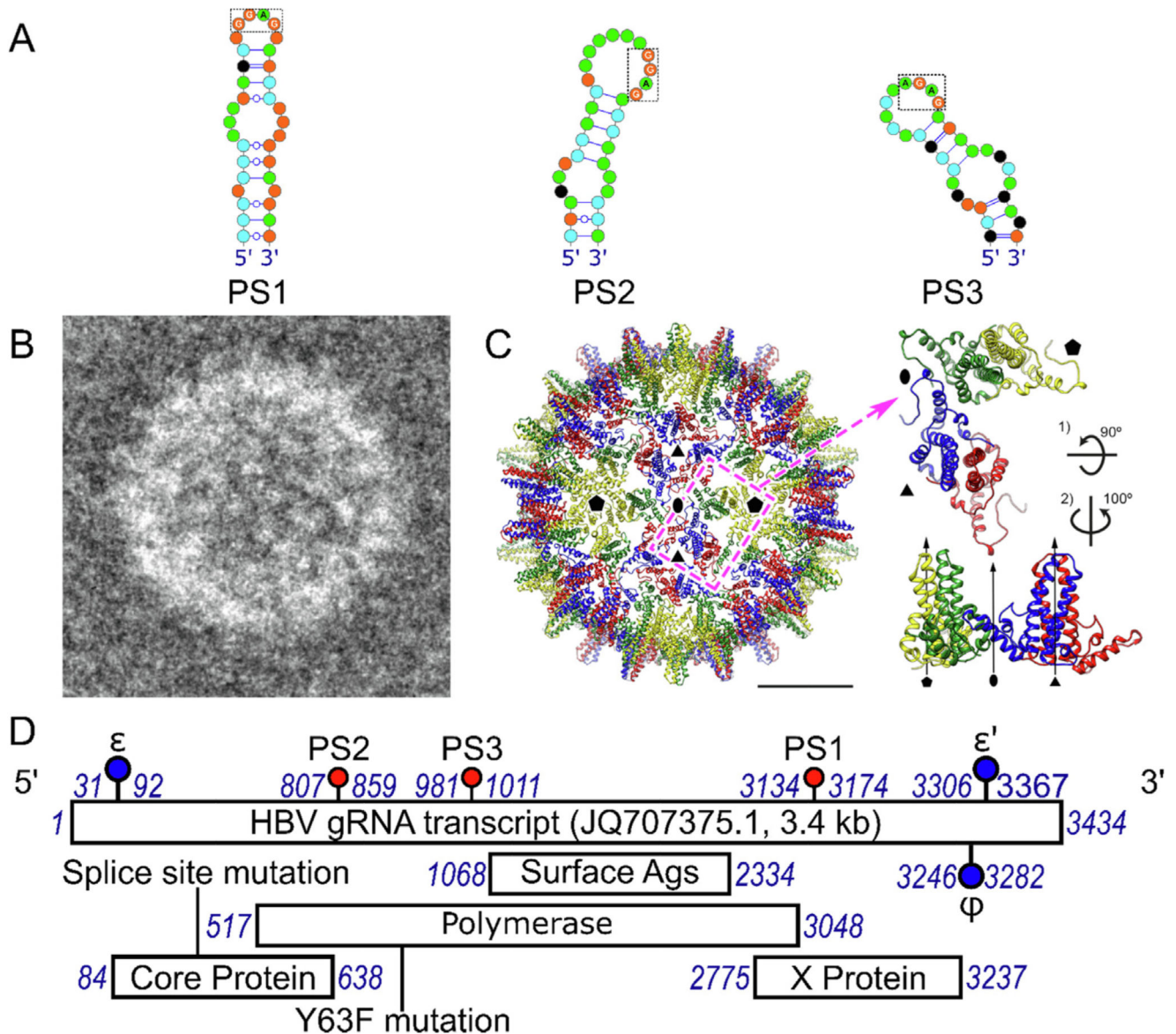
9. Zhang X, Cheng J, Ma J, Hu Z, Wu S, Hwang N, Kulp J, Du Y, Guo J-T, Chang J. Discovery of Novel Hepatitis B Virus Nucleocapsid Assembly Inhibitors. *ACS Infect Dis.* 2019; 5: 759–768. [PubMed: 30525438]
10. Hu Y, Cheng X, Cao F, Huang A, Tavis JE.  $\beta$ -Thujaplicinol inhibits hepatitis B virus replication by blocking the viral ribonuclease H activity. *Antiviral Res.* 2013; 99: 221–229. [PubMed: 23796982]
11. De Clercq E, Férrir G, Kaptein S, Neyts J. Antiviral treatment of chronic hepatitis B virus (HBV) infections. *Viruses.* 2010; 2: 1279–1305. [PubMed: 21994680]
12. Seeger C, Mason WS. Hepatitis B virus biology. *Microbiol Mol Biol Rev.* 2000; 64: 51–68.
13. Klug A, Caspar DLD. The Structure of Small Viruses. *Adv Virus Res.* 1961; 7: 225–325.
14. Conway JF, Cheng N, Zlotnick A, Wingfield PT, Stahl SJ, Steven AC. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature.* 1997.
15. Jeong J-K, Yoon G-S, Ryu W-S. Evidence that the 5'-End Cap Structure Is Essential for Encapsidation of Hepatitis B Virus Pregenomic RNA. *J Virol.* 2000; 74: 5502. [PubMed: 10823855]
16. Lee H-J, Lee J, Shin M-K, Ryu W-S. Polyadenylation Is Dispensable for Encapsidation and Reverse Transcription of Hepatitis B viral Pregenomic RNA. *Mol Cells.* 2008; 25: 545–552. [PubMed: 18460902]
17. Bartenschlager R, Schaller H. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. *EMBO J.* 1992; 11: 3413–3420. [PubMed: 1380455]
18. Bartenschlager R, Junker-Niepmann M, Schaller H. The P gene product of hepatitis B virus is required as a structural component for genomic RNA encapsidation. *J Virol.* 1990.
19. Junker-Niepmann M, Bartenschlager R, Schaller H. A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. *EMBO J.* 1990.
20. Selzer L, Zlotnick A. Assembly and Release of Hepatitis B Virus. *Cold Spring Harb Perspect Med.* 2015; 5
21. Beck J, Nassal M. Hepatitis B virus replication. *World J Gastroenterol.* 2007; 13: 48–64. [PubMed: 17206754]
22. Nassal M. Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* 2008.
23. Wang JC, Nickens DG, Lentz TB, Loeb DD, Zlotnick A. Encapsidated hepatitis B virus reverse transcriptase is poised on an ordered RNA lattice. *Proc Natl Acad Sci U S A.* 2014; 111: 11329–11334. [PubMed: 25034253]
24. Tsukuda S, Watashi K. Hepatitis B virus biology and life cycle. *Antiviral Res.* 2020; 182 104925 [PubMed: 32866519]
25. Wei L, Ploss A. Hepatitis B virus cccDNA is formed through distinct repair processes of each strand. *Nature Commun.* 2021; 12
26. Guo YH, Li YN, Zhao JR, Zhang J, Yan Z. HBc binds to the CpG islands of HBV cccDNA and promotes an epigenetic permissive state. *Epigenetics.* 2011; 6: 720–726. [PubMed: 21546797]
27. Rolfsson O, Middleton S, Manfield IW, White SJ, Fan B, Vaughan R, Ranson NA, Dykeman E, et al. Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *J Mol Biol.* 2016; 428: 4314–48.
28. Shakeel S, Dykeman EC, White SJ, Ora A, Cockburn JJB, Butcher SJ, Stockley PG, Twarock R. Genomic RNA folding mediates assembly of human parechovirus. *Nature Commun.* 2017; 8 5 [PubMed: 28232749]
29. Chandler-Bostock R, Mata CP, Bingham RJ, Dykeman EC, Meng B, Tuthill TJ, Rowlands DJ, Ranson NA, et al. Assembly of infectious enteroviruses depends on multiple, conserved genomic RNA-coat protein contacts. *PLoS Pathog.* 2020; 16
30. Tetter S, Terasaka N, Steinauer A, Bingham RJ, Clark S, Scott AJP, Patel N, Leibundgut M, et al. Evolution of a virus-like architecture and packaging mechanism in a repurposed bacterial protein. *Science (80.).* 2021; 372: 1220–1224.
31. Patel N, White SJ, Thompson RF, Bingham R, WeiR EU, Maskell DP, Zlotnick A, Dykeman EC, et al. HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral

- core protein that promote nucleocapsid assembly. *Nature Microbiol.* 2017; 2 17098 [PubMed: 28628133]
32. Borodavka A, Tuma R, Stockley PG. A two-stage mechanism of viral RNA compaction revealed by single molecule fluorescence. *RNA Biol.* 2013; 10: 481–489. [PubMed: 23422316]
  33. Twarock R, Stockley PG. RNA-Mediated Virus Assembly: Mechanisms and Consequences for Viral Evolution and Therapy. *Annu Rev Biophys.* 2019.
  34. Bunka DH, Lane SW, Lane CL, Dykeman EC, Ford RJ, Barker AM, Twarock R, Phillips SE, et al. Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: implications for the assembly of a T=1 capsid. *J Mol Biol.* 2011; 413: 51–65. [PubMed: 21839093]
  35. Dykeman EC, Stockley PG, Twarock R. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proc Natl Acad Sci U S A.* 2014; 111: 5361–5366. [PubMed: 24706827]
  36. Patel N, Clark S, Wei R EU, Mata CP, Bohon J, Farquhar ER, Maskell DP, Ranson NA, et al. In vitro functional analysis of gRNA sites regulating assembly of hepatitis B virus. *Commun Biol.* 2021; 41 (4) 1–12.
  37. Chandler-Bostock R, Bingham RJ, Clark S, Scott AJP, Wroblewski E, Barker A, White SJ, Dykeman EC, et al. RNA X-ray footprinting reveals the consequences of an in vivo acquired determinant of viral infectivity. *BioRxiv.* 2021. 2021.08.10.455819
  38. Porterfield JZ, Dhason MS, Loeb DD, Nassal M, Stray SJ, Zlotnick A. Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J Virol.* 2010; 84: 7174–7184. [PubMed: 20427522]
  39. Porterfield JZ, Zlotnick A. A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology.* 2010; 407: 281–288. [PubMed: 20850162]
  40. Connelly CM, Abulwerdi FA, Schneekloth JS. Discovery of RNA binding small molecules using small molecule microarrays. *Methods Mol Biol.* 2017.
  41. Bradner JE, McPherson OM, Mazitschek R, Barnes-Seeman D, Shen JP, Dhaliwal J, Stevenson KE, Duffner JL, et al. A Robust Small-Molecule Microarray Platform for Screening Cell Lysates. *Chem Biol.* 2006.
  42. Biacore, Guides A. Kinetics and affinity measurements with Biacore™ systems. 2008.
  43. Karlsson R, Falt A. Experimental design for kinetic analysis of protein-protein interactions with surface plasmon resonance biosensors. *J Immunol Methods.* 1997.
  44. Fatehi F, Bingham RJ, Dykeman EC, Patel N, Stockley PG, Twarock R. An intracellular model of hepatitis b viral infection: An in silico platform for comparing therapeutic strategies. *Viruses.* 2021; 13
  45. Endres D, Zlotnick A. Model-based analysis of assembly kinetics for virus capsids or other spherical polymers. *Biophys J.* 2002; 83: 1217. [PubMed: 12124301]
  46. Utrecht C, Versluis C, Watts NR, Roos WH, Wuite GJ, Wingfield PT, Steven AC, Heck AJ. High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc Natl Acad Sci U S A.* 2008; 105: 9216–9220. [PubMed: 18587050]
  47. Pollard TD. A guide to simple and informative binding assays. *Mol Bio Cell.* 2017; 21: 4061–4067. DOI: 10.1091/Mbc.E10-08-0683
  48. Fatehi F, Bingham RJ, Stockley PG, Twarock R. An age-structured model of hepatitis B viral infection highlights the potential of different therapeutic strategies. *Sci Rep.* 2022; 12: 1–12. [PubMed: 34992227]
  49. Parsons ID, Stockley PG. Quantitation of the Escherichia coli methionine repressor - Operator interaction by surface plasmon resonance is not affected by the presence of a dextran matrix. *Anal Biochem.* 1997; 254: 82–87. [PubMed: 9398349]
  50. Parsons ID, Persson B, Mekhafia A, Blackburn GM, Stockley PG. Probing the molecular mechanism of action of co-repressor in the E.coli methionine repressor-operator complex using surface plasmon resonance (SPR). *Nucleic Acids Res.* 1995; 23: 211–216. [PubMed: 7862523]

51. Routh A, Domitrovic T, Johnson JE. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proc Natl Acad Sci U S A*. 2012; 109: 1907–1912. [PubMed: 22308402]
52. 2022. [www.chembridge.com](http://www.chembridge.com)
53. 2022. [www.chemdiv.com](http://www.chemdiv.com)
54. Yang L, Liu F, Tong X, Hoffmann D, Zuo J, Lu M. Treatment of Chronic Hepatitis B Virus Infection Using Small Molecule Modulators of Nucleocapsid Assembly: Recent Advances and Perspectives. *ACS Infect Dis*. 2019.
55. Kim H, Ko C, Lee JY, Kim M. Current progress in the development of hepatitis b virus capsid assembly modulators: Chemical structure, mode-of-action and efficacy. *Molecules*. 2021; 26
56. Fatehi F, Bingham RJ, Dykeman EC, Stockley PG, Twarock R. An age-structured model of hepatitis B viral infection highlights the potential of different therapeutic strategies. *ArXiv*. 2108.01982. 2021.
57. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*. 1977; 81: 2340–2361.
58. Stray SJ, Bourne CR, Punna S, Lewis WG, Finn MG, Zlotnick A. A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. *Proc Natl Acad Sci U S A*. 2005.
59. Stray SJ, Ceres P, Zlotnick A. Zinc ions trigger conformational change and oligomerization of hepatitis B virus capsid protein. *Biochemistry*. 2004; 43: 9989–9998. [PubMed: 15287726]
60. Stray SJ, Zlotnick A. BAY 41–4109 has multiple effects on Hepatitis B virus capsid assembly. *J Mol Recognit*. 2006.
61. Deres K, Schröder CH, Paessens A, Goldmann S, Hacker HJ, Weber O, Krämer T, Niewöhner U, et al. Inhibition of hepatitis B virus replication by drug-induced depletion of nucleocapsids. *Science*. 2003; 299: 893–896. [PubMed: 12574631]
62. Campagna MR, Liu F, Mao R, Mills C, Cai D, Guo F, Zhao X, Ye H, et al. Sulfamoylbenzamide Derivatives Inhibit the Assembly of Hepatitis B Virus Nucleocapsids. *J Virol*. 2013; 87 6931 [PubMed: 23576513]
63. Zoulim F, Yogarathnam JZ, Vandebossche JJ, Lenz O, Talloen W, Vistuer C, Moscalu I, Streinu-Cercel A, et al. Safety, tolerability, pharmacokinetics and antiviral activity of JNJ-56136379, a novel HBV capsid assembly modulator, in non-cirrhotic, treatment-naive subjects with chronic hepatitis B. *Hepatology*. 2017.
64. Zhang H, Zhu X, Chen H, Li X, Hu Y, Wu M, Li C, Luo L, et al. Safety, pharmacokinetics and anti-viral efficacy of novel core protein allosteric modifier GLS4 in patients with chronic hepatitis B: Interim results from a 48 weeks phase 2a study. *Hepatology*. 2018.
65. Mani N, Cole AG, Phelps JR, Ardzinski A, Cobarrubias KD, Cuconati A, Dorsey BD, Evangelista E, et al. Preclinical profile of AB-423, an inhibitor of hepatitis B virus pregenomic RNA encapsidation. *Antimicrob Agents Chemother*. 2018.
66. Yuen MF, Gane EJ, Kim DJ, Weilert F, Yuen Chan HL, Lalezari J, Hwang SG, Nguyen T, et al. Antiviral Activity, Safety, and Pharmacokinetics of Capsid Assembly Modulator NVR 3–778 in Patients with Chronic HBV Infection. *Gastroenterology*. 2019; 156: 1392–1403. e7 [PubMed: 30625297]
67. Cihlar T, Fordyce M. Current status and prospects of HIV treatment. *Curr Opin Virol*. 2016; 18: 50–56. [PubMed: 27023283]
68. Ratni H, Scalco RS, Stephan AH. Risdiplam, the First Approved Small Molecule Splicing Modifier Drug as a Blueprint for Future Transformative Medicines. *ACS Med Chem Lett*. 2021.
69. Roberts TC, Langer R, Wood MJA. Advances in oligonucleotide drug delivery. *Nature Rev Drug Discov*. 2020; 19: 673–694. [PubMed: 32782413]
70. Ingemarsdotter CK, Zeng J, Long Z, Lever AML, Kenyon JC. An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *Retrovirology*. 2018; 15
71. Bell NM, L'Hernault A, Murat P, Richards JE, Lever AM, Balasubramanian S. Targeting RNA-protein interactions within the human immunodeficiency virus type 1 lifecycle. *Biochemistry*. 2013; 52: 9269–9274. [PubMed: 24358934]



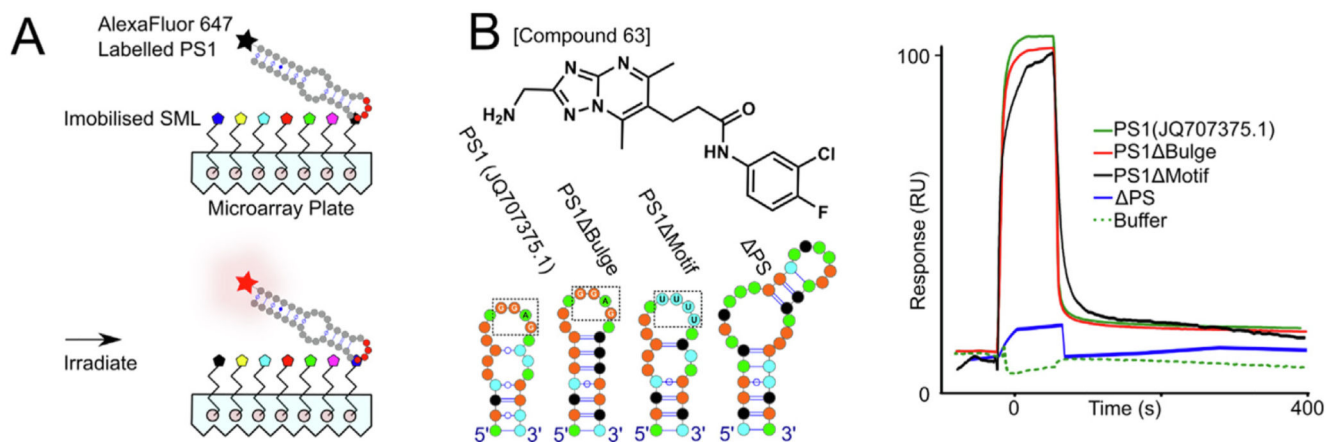
72. Brown RS, Anastakis DG, Hafner M, Kielian M. Multiple capsid protein binding sites mediate selective packaging of the alphavirus genomic RNA. *Nature Commun.* 2020; 11



**Figure 1. Components of the in vitro HBV NCP assembly system.**

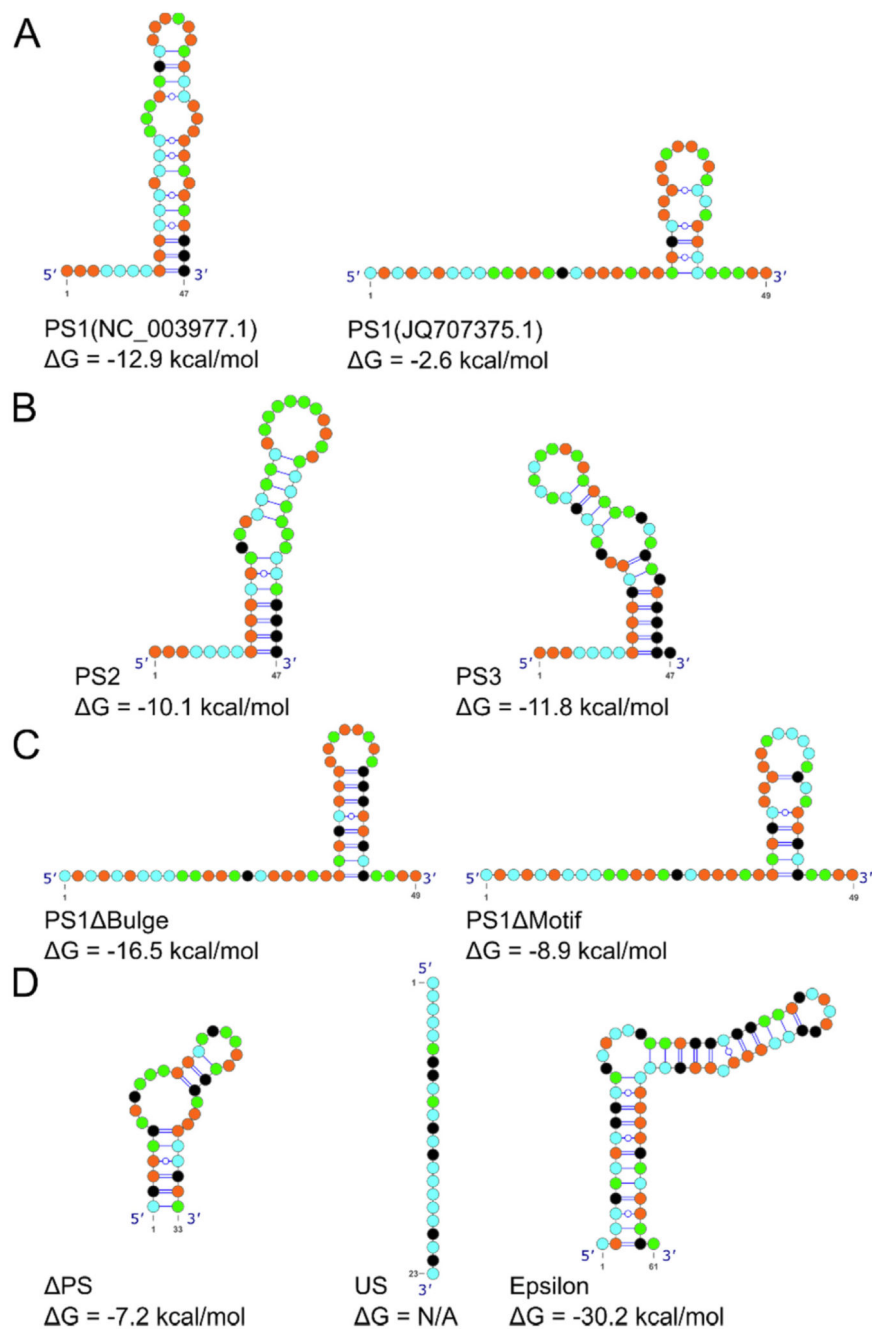
(A) *Left to right*: PS1, PS2 and PS3 (NC003977.1 strain) as identified from previous SELEX studies<sup>31</sup> presenting the Cp-recognition motif, –RGAG– within the apical loop. (Note nucleotides, here & throughout, are shown as coloured circles - green = A; orange = G; black = C; & cyan = U. Watson-Crick base pairs are indicated as lines, interrupted by circles for G-U pairs. (B) *Left*: negatively-stained, electron micrograph of a  $T=4$  NCP reassembled with the gRNA transcript used here. (C) *Left to right*: Front-half of the cryo-EM reconstruction of the reassembled NCP from (B) showing the N-terminal globular portions of the four quasi-equivalent monomers; A in yellow; B in green; C in blue & D in red (PDB ID: 7ABL).<sup>36</sup> The dashed magenta rectangle indicates the asymmetric unit composed of 2 Cp dimers shown in orthogonal views to the *right*. (D) Genetic map of the gRNA of HBV strain JQ707375.1, showing the locations (nucleotide numbers in blue

italics) of its ORFs, together with the RNA PSs (PS1, PS2 and PS3, highlighted as red lollipops), and the  $\epsilon$  and  $\psi$  (blue lollipops) stem-loops, implicated in polymerase binding and gRNA compaction.<sup>36</sup>



**Figure 2. Identification of PS1-binding compounds and their PS affinities.**

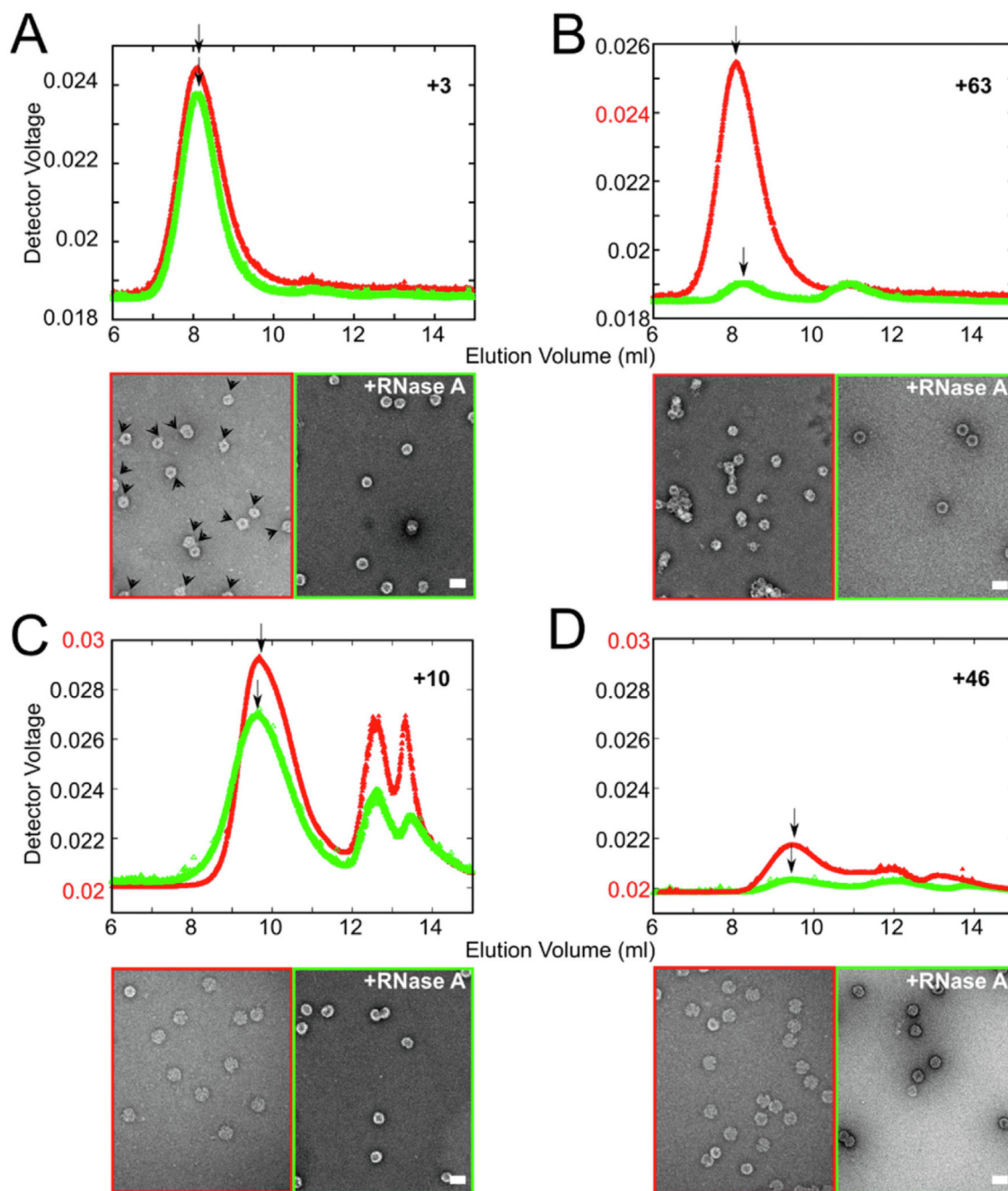
(A) Cartoon of AlexaFluor 647-labelled (black star) PS1 RNA oligonucleotide binding to immobilized ligands in the SMM. The fluorescent signals identifying potential binders is shown in Sup Figure 3. (B) *Left, top*, Structure of Compound #63, together with the PS RNA oligonucleotides used to assess binding specificity (*Left, bottom*) - PS1; PS1 Bulge; PS1 Motif & PS stem loops. (Note nucleotides, here & throughout, are shown as coloured circles - green = A; orange = G; black = C; & cyan = U. Watson-Crick base pairs are indicated as lines, interrupted by circles for G-U pairs. Mfold folding energies of these RNA stem-loops are shown in Figure 3). *Right*, SPR traces for Compound #63 (40 nM) binding to immobilized PS variants: PS1 (green); PS1 Bulge (red); PS1 Motif (black), and the unstructured (US) control (blue), Figure 3. Buffer alone SPR trace against PS1 (dashed green) is also shown. Compounds were washed over immobilized RNA oligonucleotides in a buffer containing 20 mM HEPES (pH 7.5), 250 mM NaCl, 2% (v/v) DMSO and 0.1% (v/v) Tween20, at a rate of 30  $\mu$ L/min for 1 min at 37  $^{\circ}$ C, with the dissociation of compounds monitored for 5 min.



**Figure 3. Sequences and secondary structures of RNA oligonucleotides used.**

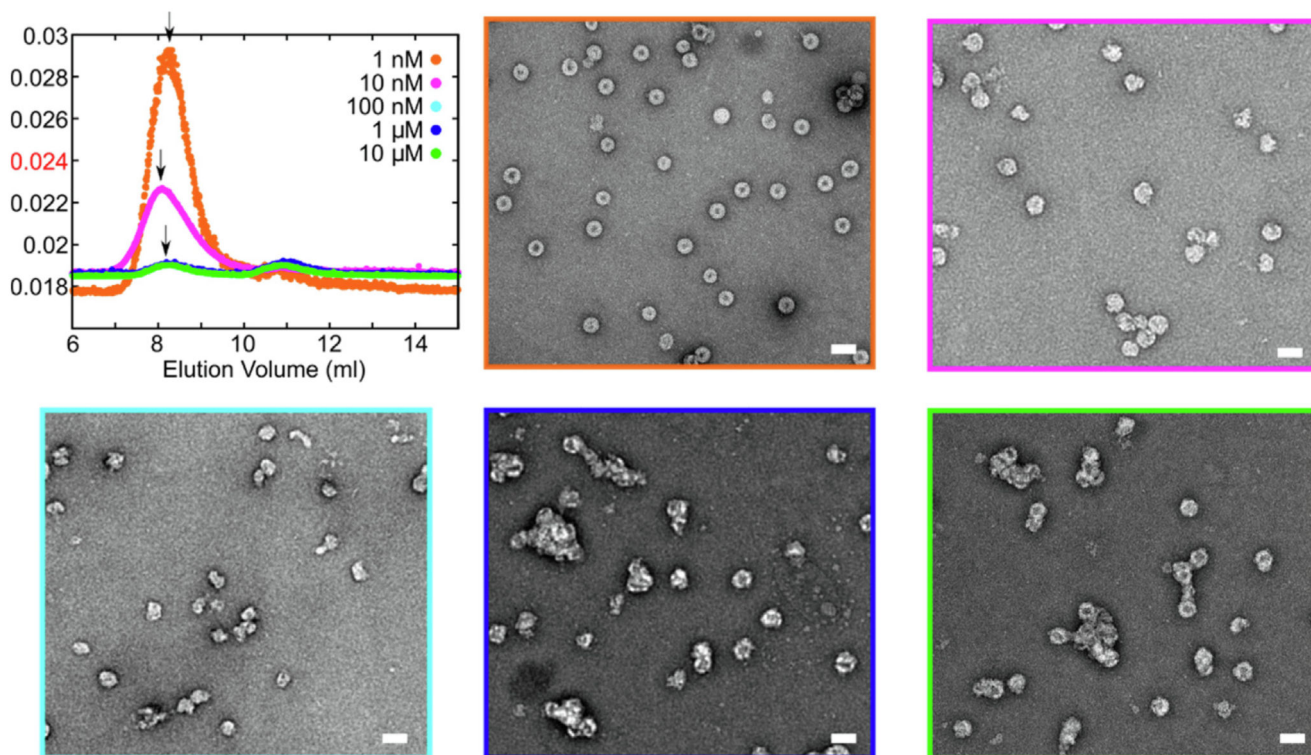
(A) PS1 (NC003977.1 and JQ707375.1) oligonucleotides. G:C clamps were used to induce the formation of the bulged stem-loop structure in the sequences from the NC\_003977.1 strain, presenting the Cp-recognition motif -RGAG- within the apical loop. (B) Similar stem-loops encompassing the PS2 and PS3 sites from strain NC\_003977.1. G:C. (C) PS Bulge and PS Motif variants of the homologous PS1 stem-loop in strain JQ707375.1.<sup>36</sup> (D) Control RNA oligonucleotides: PS (an RNA stem-loop from unrelated picornavirus);

US, an unstructured 23-mer; and epsilon ( $\epsilon$ ), the stem-loop at the 5' end of the HBV gRNA. Free folding energies for all RNA oligomers are shown below each fold.



**Figure 4. Inhibition of NCP reassembly by PS-binding compounds.**

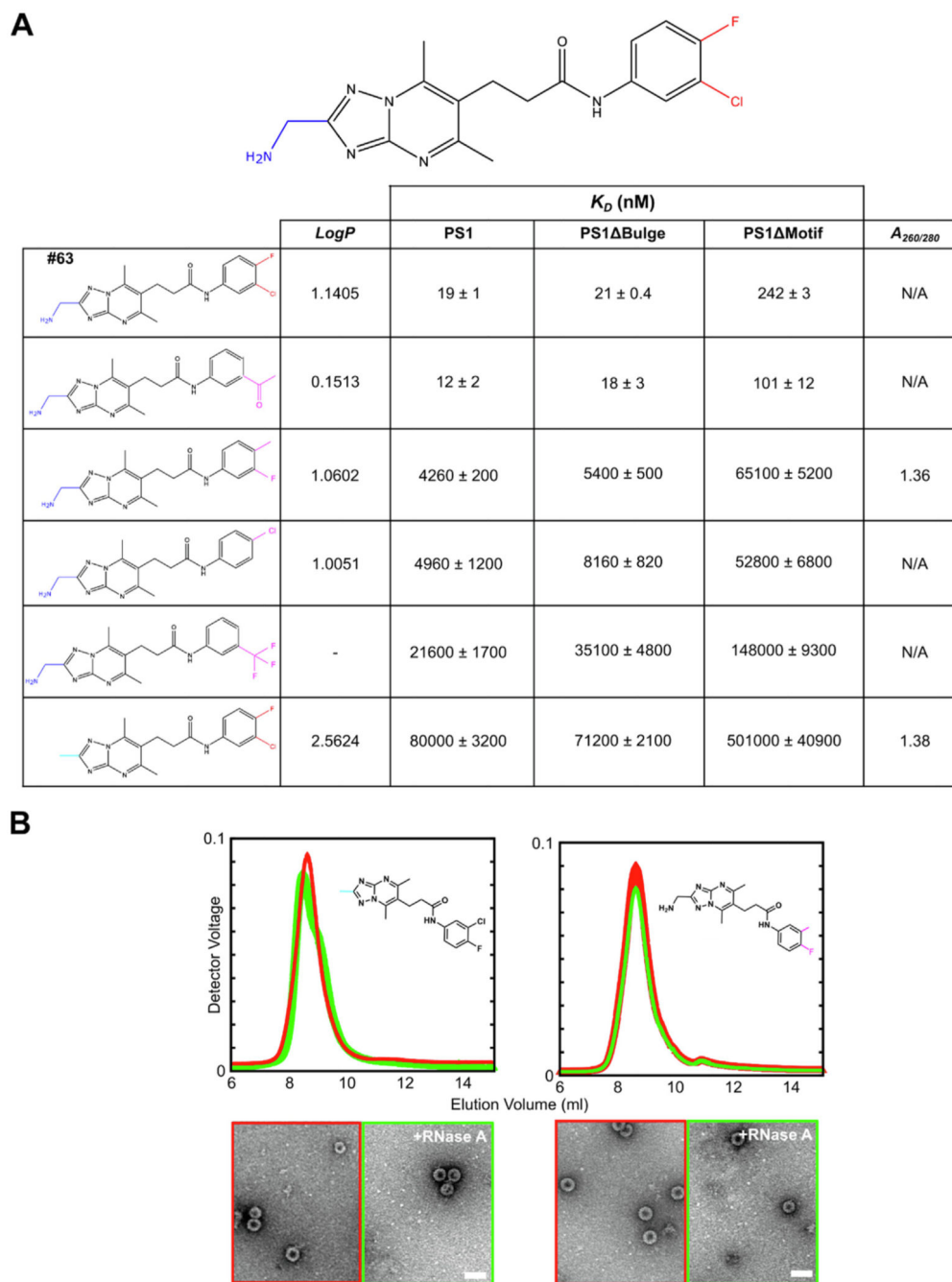
*In vitro* reassembly reactions were set up using a liquid-handling robot to titrate Cp dimer (final concentration = 1.2  $\mu\text{M}$ ) into heat-annealed gRNA (1 nM) in a 96-well plate.<sup>36</sup> Products were concentrated and analyzed, before (red) and after (green) RNase treatment by gel filtration (top traces) and nsEM (lower micrographs). Gel filtration light-scattering traces are shown for (A) Compound #3; (B) Compound #63; (C) Compound #10; & (D) Compound #46, each at a final concentration of 10  $\mu\text{M}$ .  $R_h$  values were recorded (Table 2) at the positions indicated by arrows. Scale bars, here and throughout = 50 nm.



**Figure 5. Concentration dependence of assembly inhibition by Compound #63.**

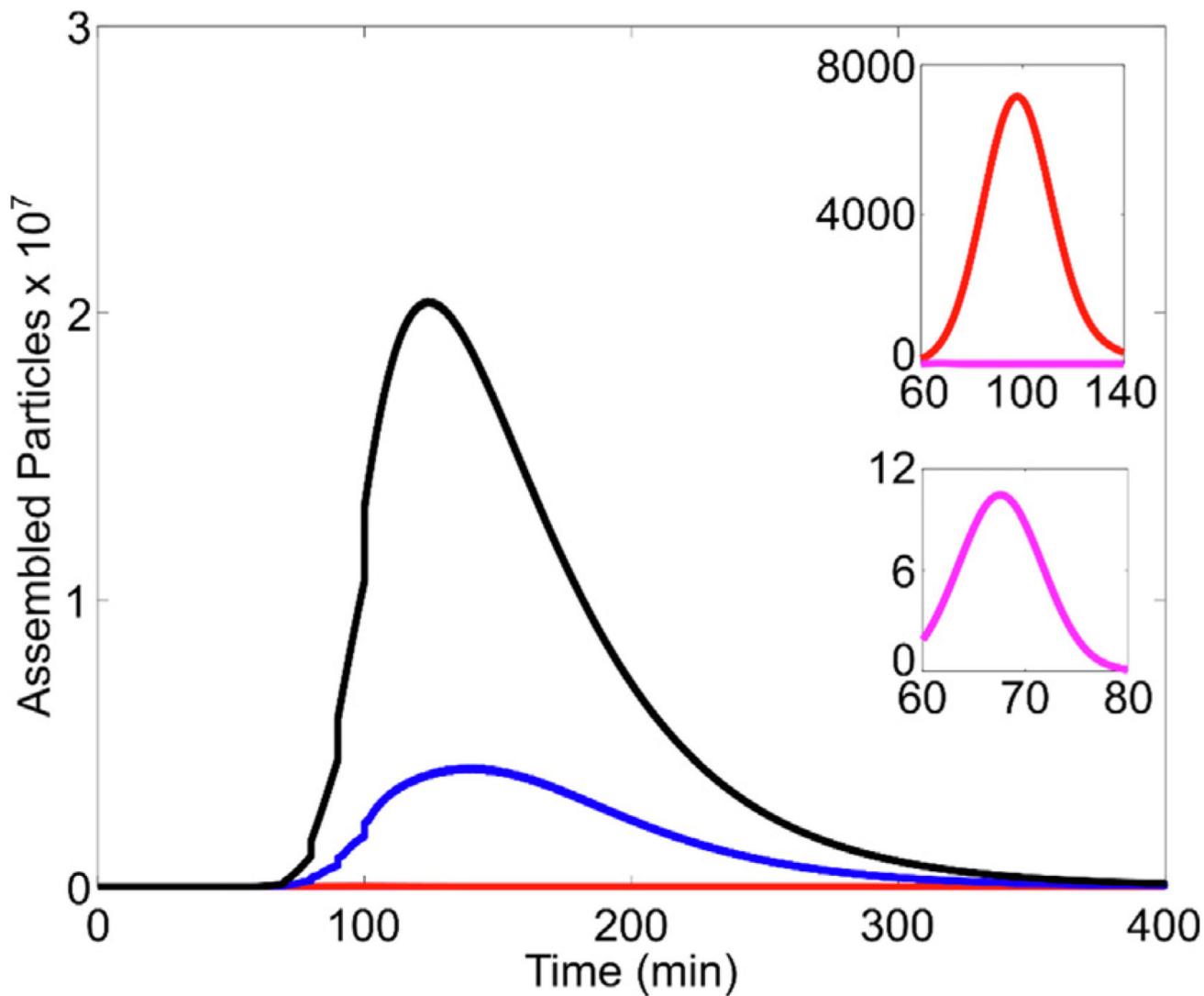
(A) LS traces of the NCP reassemblies at differing Compound #63 concentrations (details as in Figure 4): 1 nM (orange), 10 nM (magenta), 100 nM (cyan), 1  $\mu$ M (blue) and 10  $\mu$ M (green) prior to RNase treatment. Colour-coded nsEMs are shown. Nuclease treatment yields particulate products similar to reassembly in the absence of gRNA (Sup Figure 7).





**Figure 6. SAR analysis of Compound #63.**

(A) Compound #63 variants at the pendent amino and halogen functional groups (blue and red, top) were varied (cyan and magenta, respectively). SPR was used to determine their affinities ( $K_D$ ) for PS1 and its two variants, lacking the bulge in the stem (PS1 Bulge) or the Cp-recognition motif in the loop (PS1 Motif). Where available, the logP value for each compound is shown, right. (B) Two compounds (structures inset) with low affinities for PS1 were used in NCP reassemblies, at a final concentration of 10  $\mu$ M (details as in Figure 4). The  $A_{260/280}$  ratios and ns-EMs of the NCPs formed are shown in (A), right.



**Figure 7. Modelled NCP assembly in the presence of Compound #63.**

The numbers of fully-formed particles at 160 min post assembly initiation, calculated from a Gillespie simulation, are plotted against time for differing concentrations (0 (black), 1 (blue), 5 (red) & 10 (magenta) nM, respectively) of Compound #63.

**Table 1**  
**PS affinities & LogP values of test compounds.**

Compound	K <sub>D</sub> /nM			LogP
	PS1(NC_003977.1)	PS1 (JQ707375.1)	PS3	
3	211,000 ± 1179	224,000 ± 455	99,100 ± 591	4.0467
10	11,800 ± 284	11,400 ± 170	8530 ± 797	3
46	10,100 ± 119	10,200 ± 225	32,700 ± 227	1.32
63	12 ± 0.3	19 ± 0.5	15 ± 0.01	1.1405

Left, Binding affinities (nM) and their associated standard error of the mean of PS binding compounds characterised in this study, given to 3 significant figures, for RNA oligonucleotides encompassing PS1 from different strains of HBV (Left to right, NC\_003977.1, JQ707375.1) and PS3. Right, log P values for each PS binding compounds. Full table available in Supplementary Information (Sup Table 2).

**Table 2**  
**Effects of ligands on gRNA-mediated NCP reassembly in vitro.**

Compound	NCP Peak		
	$A_{260/280}$	RNA yield	$R_h$ /nm
N/A	1.43	92%	18.2
3	1.42	67%	18.0
10	1.23	63%	15.2
46	0.76	33%	14.7
63	0.98	16%	17.2

Left to Right  $A_{260/280}$  ratios, packaged RNA yield, and  $R_h$  values for reassembled materials resulting from the interaction between Cp and gRNA, in the absence and presence of Compounds #3, 10, 46 and 63. Values are shown after treatment with RNase A.

**Table 3**  
**Concentration-dependence of NCP assembly Inhibition by Compound #63.**

Compound #63/nM	NCP Peak		
	$A_{260/280}$	RNA yield	$R_h$ /nm
1	1.49	90%	17.7
10	1.46	46%	18.3
100	1.03	15%	16.4
1000	0.86	8%	18.1
10,000	0.93	15%	18.6

The yields and appearance of the NCPs produced at differing ligand concentrations (1, 10, 100, 1000 and 10,000 nM Compound #63) are listed above. Values are shown after nuclease treatment.

**Table 4**  
**Computational modelling of assembly inhibition by PS-binding ligands.**

Compound	Concentration		
	1 nM	3 nM	30 nM
3	$7.1 \times 10^{10}$	$6.9 \times 10^{10}$	$5.1 \times 10^{10}$
10	$6.2 \times 10^{10}$	$4.7 \times 10^{10}$	$1.0 \times 10^{10}$
46	$6.6 \times 10^{10}$	$5.8 \times 10^{10}$	$2.1 \times 10^{10}$
63	$1.4 \times 10^{10}$	$6.0 \times 10^8$	0.0