

Supporting Information

Transient RNA interactions leave a covalent imprint on a viral capsid protein

Zahra Harati Taji^{1,2,#}, Pavlo Bielytskyi^{1,2,#}, Mikhail Shein^{1,2}, Marc-Antoine Sani^{3,*}, Stefan Seitz^{4,5*}, Anne K. Schütz^{1,2,*}

¹Bavarian NMR Center, Department of Chemistry, Technical University of Munich, 85748, Garching, Germany, anne.schuetz@tum.de

²Institute of Structural Biology, Helmholtz Zentrum München, 85764, Neuherberg, Germany

³ School of Chemistry, Bio21 Institute, University of Melbourne, Melbourne, VIC 3010, Melbourne, Australia, msani@unimelb.edu.au

⁴Department of Infectious Diseases, Molecular Virology, University of Heidelberg, 69120 Heidelberg, Germany, stefan.seitz@med.uni-heidelberg.de

⁵Division of Virus-Associated Carcinogenesis (F170), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany

* Corresponding authors: msani@unimelb.edu.au, stefan.seitz@med.uni-heidelberg.de, anne.schuetz@tum.de

Table of contents

Materials and Methods

Production of HBV capsids
Capsid disassembly and core protein dimer purification
Preparation of RNA-free capsids
Preparation of DNA-containing capsids
Sample preparation for DNP
Cysteine labelling
Transmission Electron Microscopy
Solid-state NMR spectroscopy
Solution-state NMR spectroscopy
DNP-enhanced NMR
REDOR experiments and analysis
Core protein alignment and phylogenetic analysis

Supporting Figures

Figure S1. Characterization of DNP enhancement.
Figure S2. Control experiment for bleaching effects of radical.
Figure S3. Full spectrum of ^{31}P -filtered DARR experiment recorded on RNA-filled capsids.
Figure S4. Nucleic acid content of capsids assessed via UV absorbance.
Figure S5. Full ^{15}N -REDOR curves recorded on RNA-filled capsids under DNP.
Figure S6. Directly pulsed PDSD spectra recorded of various capsids and on lipid membranes.
Figure S7. Cysteine 48 signals from NCA and DARR spectra recorded on various capsids.
Figure S8. Size exclusion chromatograms of disassembled capsids.
Figure S9. Time series of disulfide linkage within capsid dimers.
Figure S10. Characterization of dsDNA-filled capsids.
Figure S11. Phylogenetic analysis of hepadnaviral core proteins.

Supporting Tables

Table S1. NMR experiments to probe CTD state in different capsids and redox environments.
Table S2. Experimental parameters MAS ssNMR experiments at cryogenic temperature.
Table S3. Experimental parameters for MAS ssNMR experiments at ambient temperature.

Supporting References

Materials and Methods

Production of HBV capsids

Wildtype and mutant Cp183 and Cp149 capsids were expressed in *Escherichia coli* BL21(DE3) cells using a pRSF1b-based vector containing the codon-optimized sequences for Cp183 genotype D and Cp163. Point mutations and the stop codon at position 150 to obtain Cp149 were introduced using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich). The purification was based on established protocols¹. For uniformly ¹⁵N, ¹³C labelled capsids, cells were grown in M9 minimal medium containing ¹⁵NH₄Cl and ¹³C glucose as the only nitrogen and carbon sources; first at 37 °C until OD600 of 0.8, then at 20 °C overnight after induction with IPTG. The cells were harvested and lysed by sonication in a TRIS-based buffer supplemented with 5 mM EDTA, 10 mM MgCl₂, protease inhibitor (SERVA, Heidelberg), Igepal CA-650 (0.2 % final concentration) and DNase I. The lysate was spun down at 18,000xg for 20 minutes and the supernatant was filtered and loaded onto a 10% to 60% sucrose gradient. The gradients were spun using an ultracentrifuge at 134,000xg for 15 hours and delayered. SDS-PAGE was used to identify fractions containing capsids. These fractions were pooled and diluted 1 in 2 with HBS-300 (150 mM HEPES pH 7.5, 300 mM NaCl). The protein was then precipitated by step-wise addition of (NH₄)₂SO₄ to 60 % saturation, spun down at 30,000 xg for 30 minutes, resuspended in HBS-300 and dialyzed overnight at 4 °C to the final buffer used for all solid-state NMR experiments (50 mM HEPES pH 7.5, 150 mM NaCl, 5 mM EDTA) unless otherwise specified. Solution-state NMR measurements were carried out in HBS-300 buffer supplemented with 10 % D₂O in 3 mm NMR tubes.

Capsid disassembly and core protein dimer purification

Purified Cp183 capsids were disassembled into dimers as previously reported² and purified on a Superdex 200 (Cytiva, Uppsala) column. The fractions containing Cp183 dimers were combined and concentrated using a centrifugal filter unit with a 30 kDa cut-off. Solution-state NMR measurements were conducted in disassembly buffer (50 mM HEPES pH 7.5, 1.5 M GnHCl, 0.5 M LiCl, with and without 2 mM TCEP) with 10 % D₂O in 3 mm NMR tubes.

Preparation of RNA-free capsids

Purified Cp183 capsids were “emptied” of nucleic acids using either reducing conditions or oxidizing conditions adapted from the protocols reported by Porterfield *et al.*² or Strods *et al.*³, respectively.

For oxidized RNA-free capsids, purified Cp183 capsids were solubilized in 8 M urea with 2 mM DTT at 42 °C and placed on a Superdex S75 column equilibrated with 0.1 M Na₂CO₃ pH 11.6, 2 mM DTT. The fractions containing the capsid protein were then dialyzed in an alkaline solution (0.1 M Na₂HPO₄/Na₃PO₄ pH 11.8, 650 mM NaCl) overnight at 37 °C with two buffer exchanges. The protein was then dialyzed in restoration buffer (0.1 M Na₂HPO₄/Na₃PO₄ pH 7.5, 650 mM NaCl) for two hours at room temperature (RT) followed by overnight at 4 °C, with two buffer exchanges. The capsids were dialyzed into solid-state NMR buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 5 mM EDTA; same as above but including 250 mM NaCl) overnight at 4 °C. Solution-state measurements were carried out in restoration buffer supplemented with 10 % D₂O in 3 mm NMR tubes.

For reduced RNA-free capsids, purified Cp183 capsids were disassembled and purified as per the protocol for dimer production. After size-exclusion chromatography, the dimers were reassembled by overnight dialysis with re-assembly buffer (50 mM HEPES, 250 mM NaCl, 2 mM TCEP). The capsids were briefly dialyzed over 4 hours into solid-state NMR buffer (50 mM HEPES pH 7.5, 250 mM NaCl, 5 mM EDTA; same as above but including 250 mM NaCl) to remove excess TCEP.

Preparation of dsDNA-containing capsids

A circular dsDNA of 2686 base pairs, the *Escherichia coli* plasmid pUC18, was prepared by growing DH10β cells transformed with the plasmid. The plasmid was purified from the harvested cells using the PureLink™ HiPure Expi Plasmid Gigaprep Kit (Invitrogen, Waltham).

The core protein dimer was purified as described above and concentrated using a centrifugal filter unit with a 30 kDa cut-off. The protein was added to twofold molar excess of purified pUC18 (two plasmids per 120 core dimers, equivalent to one T4 capsid), with the final buffer conditions reaching 16.7 mM HEPES pH 7.5, 0.5 M GnHCl, 167 mM LiCl to allow capsid re-assembly. The dsDNA-containing capsids were then separated from unencapsidated dsDNA using a 10% to 60% sucrose gradient. The gradients were spun using an ultracentrifuge at 134,000xg for 15 h and layered. SDS-PAGE was used to identify fractions containing capsids, which were then pooled and dialyzed into solid-state NMR buffer (50 mM HEPES

pH 7.5, 250 mM NaCl, 5 mM EDTA; same as above but including 250 mM NaCl). The dsDNA was further analysed using Native Agarose Gel Electrophoresis. The samples were loaded onto a 1% agarose gel using Tris-acetate-EDTA as the running buffer. The nucleic acids were stained using DNA Stain Clear G (SERVA, Heidelberg) and visualized using UV illumination.

Sample preparation for DNP

The uniformly ^{13}C , ^{15}N labelled sample used for the DNP-enhanced experiments was dialysed into a degassed, deuterated buffer with $\text{D}_2\text{O}:\text{H}_2\text{O}$ (90:10, v:v) and prepared without additional cryoprotectant^{4 5}. The protein solution with 5 mM of AMUPol radical (CortecNet, Voisins-Le-Bretonneux, France) was sedimented into ZrO_2 3.2 mm thick-walled MAS rotor by ultracentrifugation (134,000 g, 24 h, 4°C) using commercially available filling tools (Giotto Biotech, Sesto Fiorentino, Italy). The filled rotor was pre-spun at a MAS rate of 5 kHz at 4 °C for 1 h to ensure homogeneous distribution of a protein sediment against rotor walls prior to insertion of rotor into the cold low-temperature MAS probe.

Cysteine labelling

Selective $^{13}\text{CH}_3\text{-S}$ -labeling of cysteine side chains was performed using ^{13}C -methyl-methanethiosulfonate (^{13}C -MMTS)⁶. Purified Cp183 capsids were dialysed overnight into degassed solid-state NMR buffer with either 5 mM TCEP or without TCEP, as indicated in Figure 5b. The TCEP was subsequently removed with PD-10 desalting columns (GE Healthcare) according to the manufacturer's protocol using degassed solid-state NMR buffer. A stock solution of 100 mM MMTS in DMSO was added in ten-fold molar excess relative to Cp183 protein monomers times the number of cysteine residues in the construct and left to react overnight at 4 °C. The unreacted MMTS was removed with PD-10 desalting columns.

Transmission Electron Microscopy

The transmission electron microscopy (TEM) samples were prepared on 3 mm continuous carbon film supported copper grids (mesh 300, Plano) by glow discharge sputtering. The sample (5 μL) was applied to the grid and incubated at RT for 5 minutes. The sample was then removed while applying 20 μL of HEPES wash (50 mM HEPES pH 7.5). Uranyl acetate solution (5 μL , 2 % w/v) was applied to the grid and incubated for 30 seconds. The staining solution was then removed and the grid was left to air-dry. Prepared grids were kept stored at

RT until imaging. Images were obtained on a JEOL JEM-1400 Plus transmission electron microscope at 120 kV using a charge-coupled device camera (JEOL CCD Ruby, 8 Mpix), at 60,000x magnification (0.275 nm/pix). The underfocus was set to 500 nm.

Solid-state NMR spectroscopy

The RT NMR experiments were performed at magnetic field of 17.6 T (750 MHz ^1H Larmor frequency) on a Bruker AVANCE III NMR spectrometer equipped with a 1.9-mm triple-resonance ($^1\text{H}/^{13}\text{C}/^{15}\text{N}$) MAS probe. For each capsid type (Cp183 and point mutants thereof, Cp149), 9 mg of ^{13}C , ^{15}N -labelled capsids (as assessed by measuring the absorption value at $\lambda = 280$ nm) were sedimented into a 1.9 mm ZrO_2 MAS rotor (134,000 g, 15 h, 4 °C). Experiments were conducted at a MAS frequency of 16650 Hz and sample temperature of 14 °C as determined by the frequency of water resonance⁶. Typical ^1H , ^{13}C and ^{15}N 90 ° pulse lengths were 1.5, 5.0 and 4.5 μs , respectively. All spectra were referenced to sodium trimethylsilylpropanesulfonate (DSS) added directly into the MAS rotor. Further experimental NMR parameters are listed in Table S3. The dsDNA capsids were filled into a 1.3 mm ZrO_2 MAS rotor and experiments were conducted at a MAS frequency of 21000 Hz at a 18.8 T (800 MHz ^1H Larmor frequency) Bruker AVANCE III HD NMR spectrometer.

The spectra were processed with Bruker TopSpin 3.5 and analysed with CcpNmr Analysis⁷. Prior to the Fourier transformation the time domain signals in the direct and indirect dimensions were zero-filled to twice the number of acquired points and apodized with a 60° shifted squared sine bell window function. The peaks in 2D spectra were assigned based on the reported chemical shifts⁸ (BMRB entry 27317).

Solution-state NMR spectroscopy

The experiments were performed at a magnetic field of 22.3 T (950 MHz ^1H Larmor frequency) at a Bruker AVANCE III NMR spectrometer equipped with triple-resonance cryogenic probe at a temperature of 293 K. The typical ^1H and ^{15}N 90 ° pulses were 8.5-12.7 and 37.5 μs , respectively.

DNP-enhanced NMR

$^1\text{H}/^{13}\text{C}/^{15}\text{N}$ and $^1\text{H}/^{31}\text{P}/^{13}\text{C}$ mode

The NMR experiments at cryogenic temperatures were performed at a static magnetic field of 18.8 T (800 MHz ^1H Larmor frequency) on a Bruker AVANCE III NMR spectrometer equipped with 3.2-mm triple-resonance low-temperature MAS probe, operating either in

$^1\text{H}/^{13}\text{C}/^{15}\text{N}$ or $^1\text{H}/^{31}\text{P}/^{13}\text{C}$ mode. The microwaves were generated by 527 GHz gyrotron (Bruker) with 110 mA beam current that provided the best ^1H enhancement for AMUPol, as was determined empirically. Experiments were conducted at a MAS frequency of 9470 Hz (unless further mentioned) and a sample temperatures of 108 K (mw on) and 104 K (mw off). Typical ^1H , ^{13}C , ^{15}N and ^{31}P 90° pulse lengths were 2.8, 5.5, 7.0 and 5.0 μs , respectively. The DNP enhancement factor was estimated as the ratio between signal intensities with (mw on) and without (mw off) microwaves. Hence, enhancement ε^{DNP} is reported as $\varepsilon_{\text{on}}/\varepsilon_{\text{off}}$. The spectra were referenced to adamantane as an external standard. Detailed NMR parameters are listed in Table S2.

$^1\text{H}/^{31}\text{P}/^{15}\text{N}$ mode

MAS NMR experiments in $^1\text{H}/^{31}\text{P}/^{15}\text{N}$ mode were conducted on a 9.4 T (400 MHz ^1H Larmor frequency) wide-bore Bruker Avance-III HD DNP NMR Spectrometer (Wissembourg, France) equipped with a 3.2 mm low-temperature triple-resonance MAS probe and a 263.33 GHz gyrotron source. The temperature of all experiments was maintained at *ca.* 110 K, the spinning frequency was set to 8 or 9 kHz.

^{15}N CPMAS experiments were performed with 98 kHz proton excitation pulse followed by 2.75 ms Hartmann-Hahn contact with a 30% RAMP and a $1.3 \times T_{\text{B,on}}$ recycle delay; 32-128 scans and 800 complex points with a spectral window of 24 kHz were acquired under 98 kHz SPINAL decoupling. FIDs were zero-filled to 4096 points and 100 Hz line broadening was applied.

^{31}P CPMAS experiments were performed with 98 kHz proton excitation pulse followed by 2.5 ms Hartmann-Hahn contact with a 30 % RAMP and a $1.3 \times T_{\text{B,on}}$ recycle delay; 16-64 scans and 4096 complex points with a spectral window of 125 kHz were acquired under 98 kHz SPINAL decoupling. FIDs were zero-filled to 16384 points and an 80 Hz line broadening was applied.

REDOR experiments and analysis

$^{15}\text{N}\{^{31}\text{P}\}$ -REDOR experiments were performed under the CPMAS conditions described above, using 9 μs and 13.5 μs 180° pulses for ^{31}P and ^{15}N , respectively. Several rotor-synchronized dephasing times were acquired with and without the ^{31}P dephasing pulses. The J-decoupled REDOR experiments⁹ were performed under similar conditions but with a 1 % truncated 4 ms gaussian pulse for ^{15}N selective recoupling.

To obtain the REDOR curve, the S/S_0 ratio, with S and S_0 being the intensities of dephased and full-echo signal, respectively, is plotted as a function of the dephasing time. The time-dependent REDOR signal of a heteronuclear spin pair is expressed as:

$$\frac{S}{S_0} = \int_0^\pi \int_0^{2\pi} d\alpha \sin(\beta) d\beta \cos[2\sqrt{2\lambda_D} \sin(2\beta) \sin(\alpha')]$$

where α and β are the azimuthal and polar angles between the inter-nuclear vector and the rotor spinning axis; $\alpha' = \alpha + 2\pi\phi$ for a π pulse centered at ϕ fraction of a rotor period; $\lambda_D = DN_c T_r$ where N_c is the number of rotor periods at MAS of $1/T_r$ and D is the dipolar coupling constant $D = (\mu_0 \hbar \gamma_I \gamma_S) / 8\pi^2 r^3$ at the internuclear distance r .

The BS-REDOR fitting program provides the most probable distribution of internuclear distances consistent with the REDOR data¹⁰. It is obtained by finding the optimal configuration q of Lagrange multipliers $\vec{\lambda}_q$ for points S/S_0 , with the goodness of a fit χ_q^2 given by:

$$\chi_q^2 = \sum_j \frac{1}{\sigma_j^2} [(S/S_0)_{j_q} - (S/S_0)_j]^2$$

where σ is rms noise¹⁰.

The REDOR dephasing curves were simulated with a distance resolution of 0.1 Å and 180 discrete angles per 2π . The fits were optimized using a simulated annealing approach with 100-500 iterations.

Core protein alignment and phylogenetic analysis

The alignment of C protein sequences of *hepadna*- and *nackednaviruses* was built on a previously published profile-based alignment that accounted for predicted protein secondary structure (Data S4 in Lauber et al.¹¹). Additional C protein sequences of novel *hepadnaviruses* discovered since 2017 were added secondarily. Alignment positions appearing as insertions relative to the C protein of HBV genotype D were removed. The phylogenetic tree was inferred in MEGA7¹² with the Maximum Likelihood method using an LG+G+I substitution model and 500 bootstrap replicates.

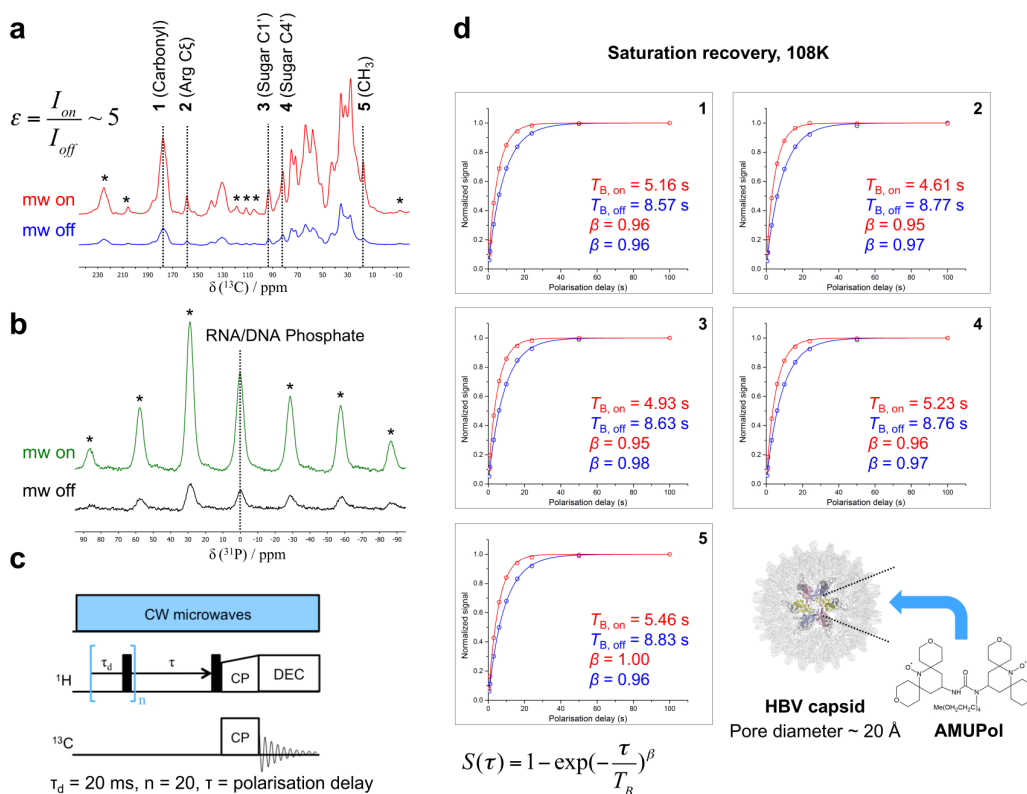


Figure S1 (a) DNP-enhanced ^1H - ^{13}C and (b) ^1H - ^{31}P CP-MAS spectra of nucleic-acid filled capsids assembled from Cp183, with $^{13}\text{C}^{15}\text{N}$ labelling, obtained at 108 K, $B_0 = 18.8 \text{ T}$ and MAS frequency 9.4 kHz with (mw on) and without (mw off) microwave radiation. The DNP enhancement factor $\varepsilon = 5$ was calculated as a ratio between signal intensities with (I_{on}) and without (I_{off}) microwave radiation. Spinning sidebands are highlighted with asterisks. (c) Saturation recovery pulse sequence; black rectangles represent 90° pulses. (d) ^1H - ^{13}C CP-MAS build-up curves obtained with and without microwave radiation (red and blue, respectively), normalized to steady-state signal intensity corresponding to the longest delay τ . Fitting was performed with a stretched exponential, where $S(\tau)$ is the intensity of the normalized signal at time τ ; T_B is the effective build-up time with or without microwaves ($T_{B, on}$ and $T_{B, off}$, respectively); β is the stretching parameter. The error bars were omitted due to insignificance at the presented scale. Faster polarisation build-up in the presence of microwave radiation indicates relayed DNP transfer mediated by spin diffusion¹³. Similar build-up rates close to monoexponential (with $\beta \sim 1$) for the signals corresponding to amino acid residues (inserts 1, 2, 5) and sugar moieties of the encapsidated nucleic acids (inserts 3 and 4) indicate homogeneous distribution of polarizing agent throughout the sample volume. The penetration of AMUPol into the capsid interior most likely occurs through surface large quasi-sixfold pores with a diameter up to $\sim 20 \text{ \AA}$ (highlighted with colour)¹⁴.

Standard ssNMR buffer, 14°C
 DNP buffer, 14°C

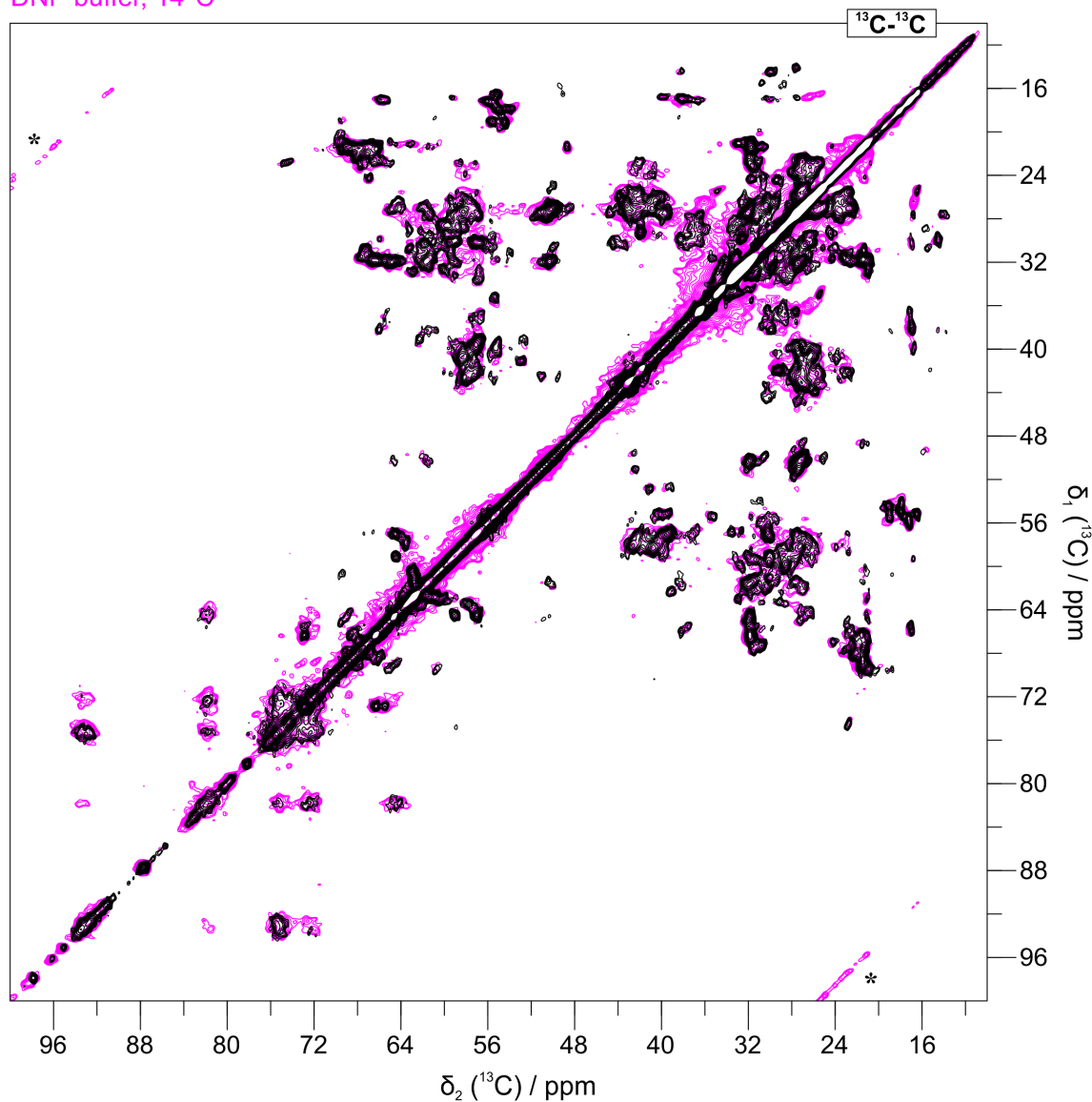


Figure S2 2D ^{13}C - ^{13}C DARR spectra with 25 ms mixing time of capsids assembled from Cp183 with $^{13}\text{C}^{15}\text{N}$ labelling in standard solid-state NMR buffer (150 mM NaCl, 50 mM HEPES, 5mM EDTA, pH=7.5) recorded at $B_0 = 17.6$ T, MAS frequency 16.5 kHz and temperature 14°C (black), compared to the sample used for the DNP-enhanced NMR experiments (5mM AMUPol, $\text{D}_2\text{O}:\text{H}_2\text{O}$ 90:10 v:v, 150 mM NaCl, 50 mM HEPES, 5mM EDTA, pH=7.5) recorded at $B_0 = 18.8$ T, MAS frequency 15.0 kHz and temperature 14°C (magenta). Spinning sidebands are highlighted with asterisks. The presence of polarizing agent in the DNP sample does not cause visible bleaching of the NMR signals.

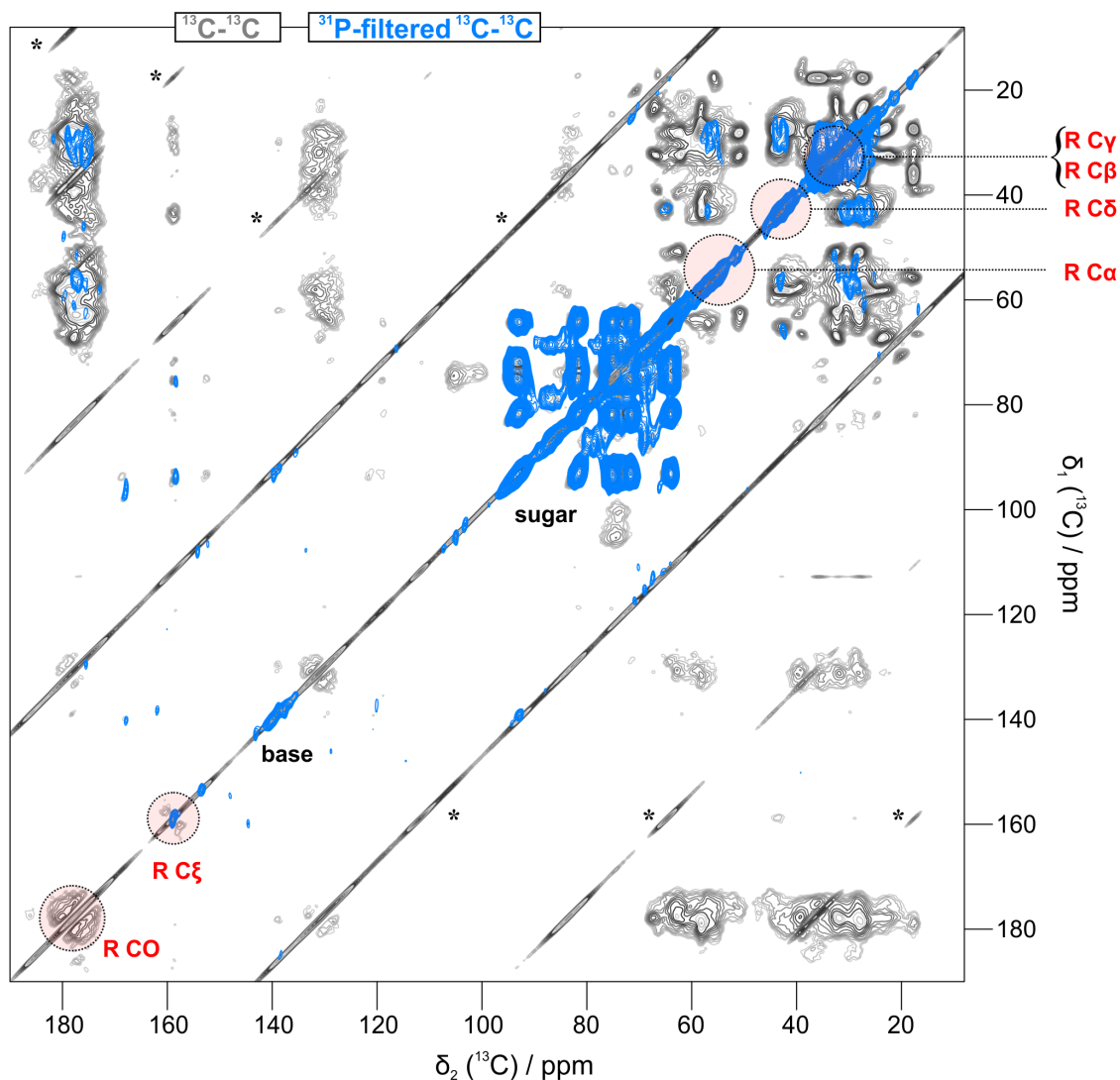


Figure S3 Comparison of DNP-enhanced ^{13}C - ^{13}C DARR (grey) and ^{31}P -filtered ^{13}C - ^{13}C DARR (light blue) acquired on capsids assembled from Cp183 $^{13}\text{C}/^{15}\text{N}$ with $500\ \mu\text{s}$ ^1H - ^1H spin diffusion time and 50 ms DARR mixing (light blue). The spectra were collected at 108 K, $B_0 = 18.8\ \text{T}$ and MAS frequency 9330 Hz with microwave radiation. Spinning sidebands are highlighted with asterisks. Asymmetric nature of the ^{31}P -filtered ^{13}C - ^{13}C DARR spectrum with respect to diagonal points to magnetization transfer originating from nucleic acid phosphate groups directed towards side chains of the residues located in the close vicinity. The correlation pattern can be rationalized with signals from the ribose moiety of RNA and arginine spin systems (R, highlighted in red), suggesting that the interaction between capsid and nucleic acids is realized via arginine-rich motifs (ARMs) of the C-terminal domain alone.

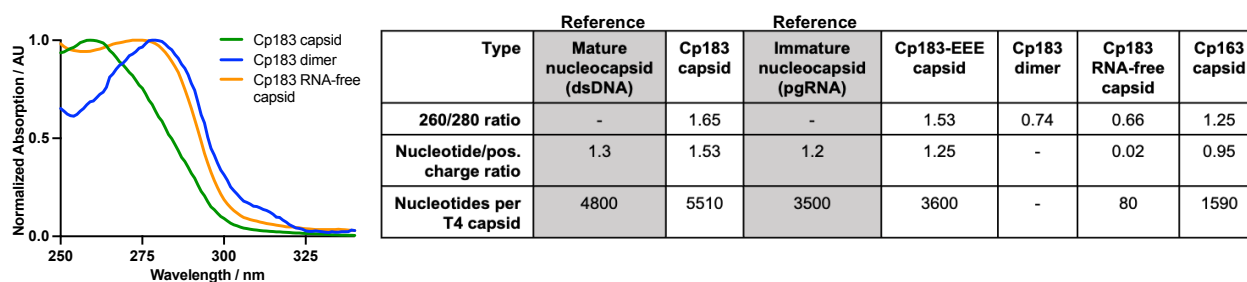


Figure S4 UV absorption spectra were measured of RNA-containing capsids assembled from Cp183, Cp183-EEE, Cp163 and of Cp183 dimers and RNA-free capsids. The nucleotide content of each capsid type was assessed as previously reported¹⁵ with light scattering adjustments applied to all capsid measurements. The number of charges in the CTD are: +15 for Cp183, +12 for Cp183-EEE and +7 for Cp163. The Cp183 and Cp183-EEE capsids mimic mature and immature nucleocapsids, respectively, according to phosphorylation state. The properties of nucleocapsids found *in vivo* are provided as references in the grey columns. The number of positive charges in mature dsDNA-containing and immature pgRNA-containing nucleocapsids is considered to be +15 and +12, respectively, due to differences in phosphorylation; the number of nucleotides takes into account the polyadenylation of RNA and the partial synthesis of the second DNA strand¹⁵. UV spectra were measured on an Implen Nanophotometer N60 using the corresponding sample buffer to subtract the background.

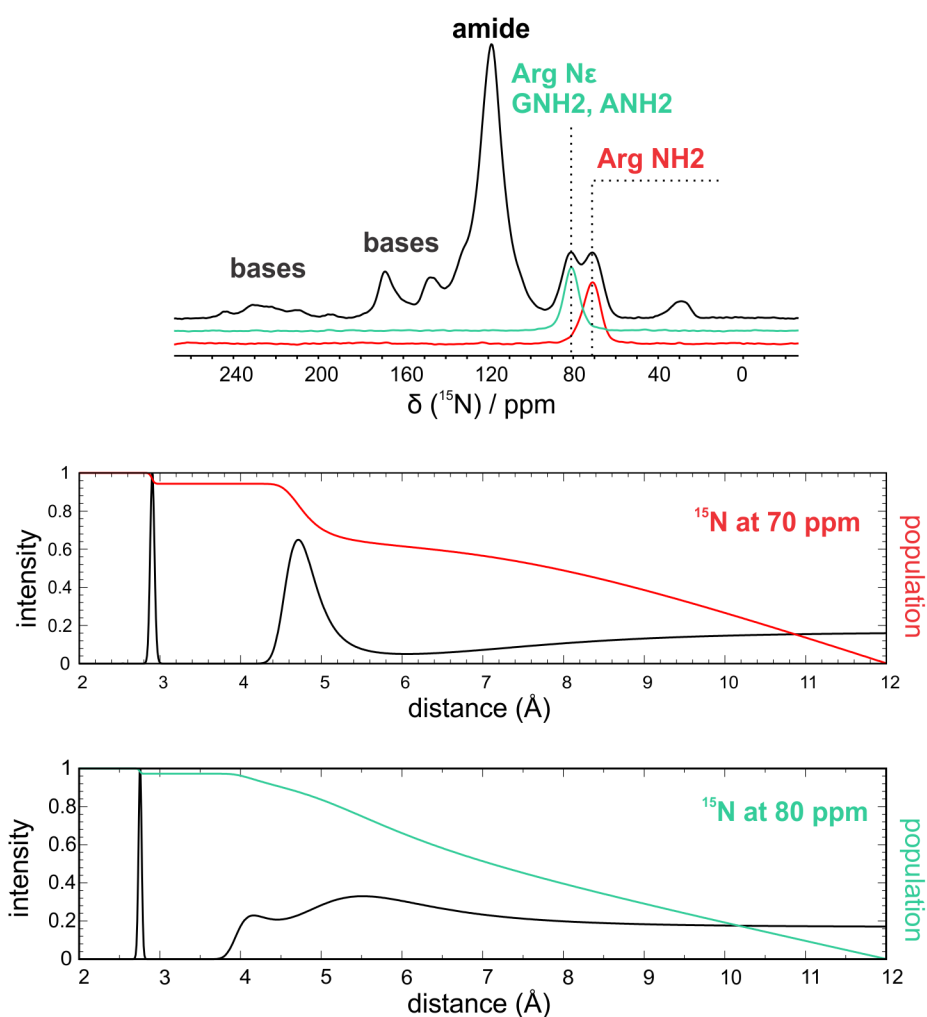


Figure S5 Top: 1D DNP-enhanced ^1H - ^{15}N CP of RNA-filled capsids assembled from Cp183 capsid with broadbanded excitation (black) and 1D slices of $^{15}\text{N}\{^{31}\text{P}\}$ DNP-enhanced J-decoupled REDOR with a 4 ms gaussian pulse for ^{15}N selective excitation at 70 and 80 ppm (red, green) performed at 8 kHz spinning speed and a stator temperature of 111 K. Bottom: BS-REDOR reconstruction from 70 ppm and 80 ppm signals of $^{15}\text{N}\{^{31}\text{P}\}$ J-decoupled REDOR. The distance distribution (black lines) and respective populations (coloured lines) are presented. The bulk of the dephased populations is accounted for by a continuous ensemble starting from ~ 6 Å, due to ^{15}N spins not engaged in the distinct ^{31}P network of an arginine-phosphate salt bridge. The sharp intensity peak at a distance of ~ 2.7 Å corresponds to a minor population (less than 4%) and is observed for both reconstructions yet has no equivalent in ^{15}N - ^{31}P distances expected for phosphate-arginine interactions. It an artefact from BS-REDOR analysis that is likely due to an incomplete dephasing plateauing at a *ca.* S/S_0 ratio of 0.7.

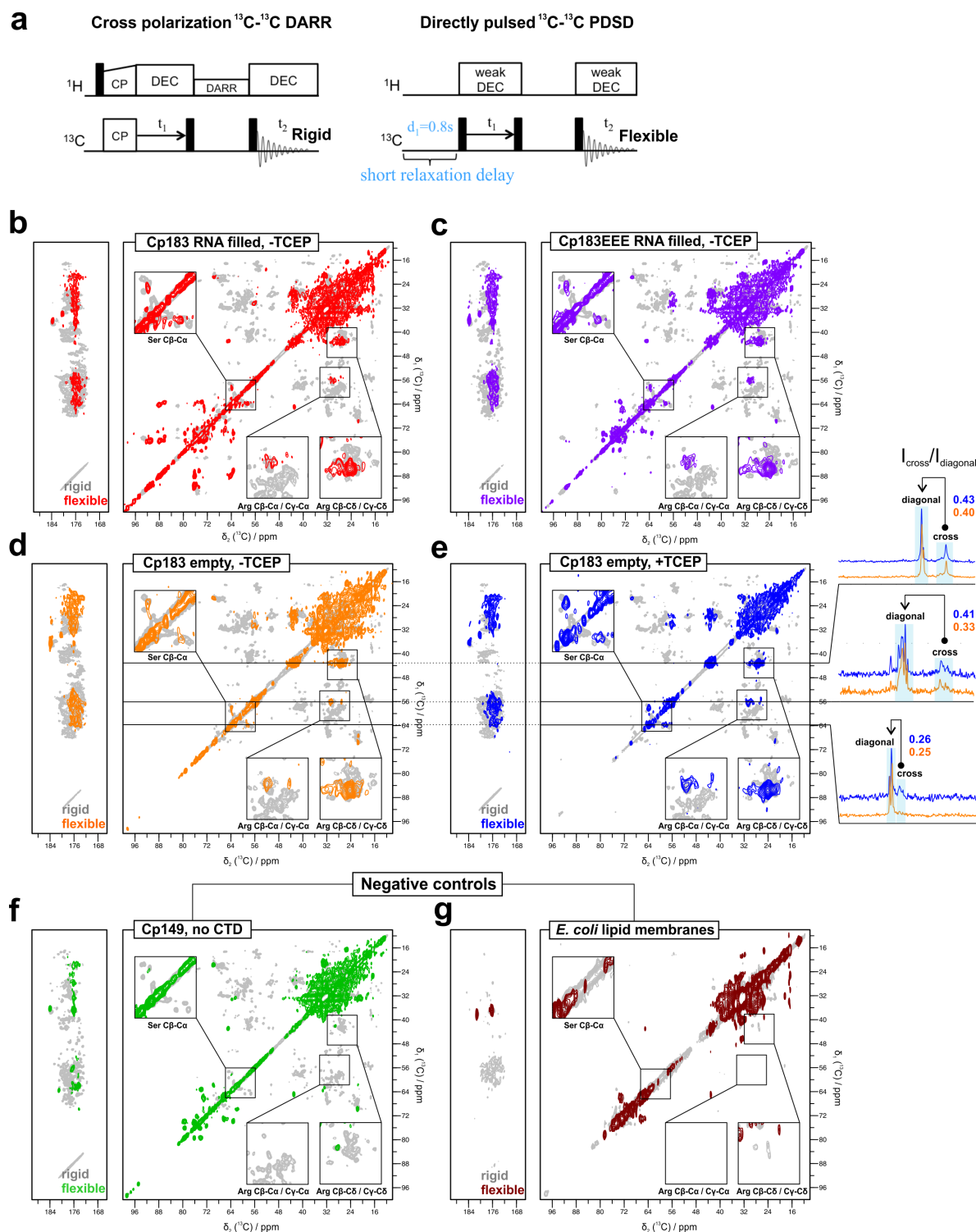


Figure S6 (a) NMR pulse sequences used to differentiate protein segments based on their flexibility. The CP-based DARR with high power heteronuclear decoupling (left) establishes ^{13}C - ^{13}C correlations within rigid domains, while the directly pulsed (dp-)PDSD with short relaxation delay and weak decoupling (right) highlights residues belonging to flexible regions. (b-f) Overlays of dp-PDSD spectra of different capsid types, acquired at ambient temperature with 0.8 s relaxation delay and 200 ms mixing time, highlighting arginine and serine spin systems belonging to flexible segments (coloured contours), with respective CP-based DARR spectra featuring

residues from rigid domains (grey contours). Since the magnetization transfer during the mixing period in both CP-DARR and dp-PDSD schemes is mediated by dipolar couplings, the abundance of signals from the CTD arginine and serine residues in dp-PDSD spectra points towards an intermediate dynamic regime under which such couplings are not fully averaged out. The dynamical changes of the CTD between different capsid types are either minor or below the sensitivity level of our experimental setup, as was tentatively estimated from the ratios of signal intensities between cross and diagonal peaks. Increased mobility would be reflected in more intense diagonal peaks and, in contrast, less pronounced cross peaks, leading to a smaller ratio. **(g)** Reference spectra of *E. coli* lipid membranes, traces of which are present in RNA-filled capsid samples, feature no signals that overlap with arginine and serine cross peaks in the capsid samples.

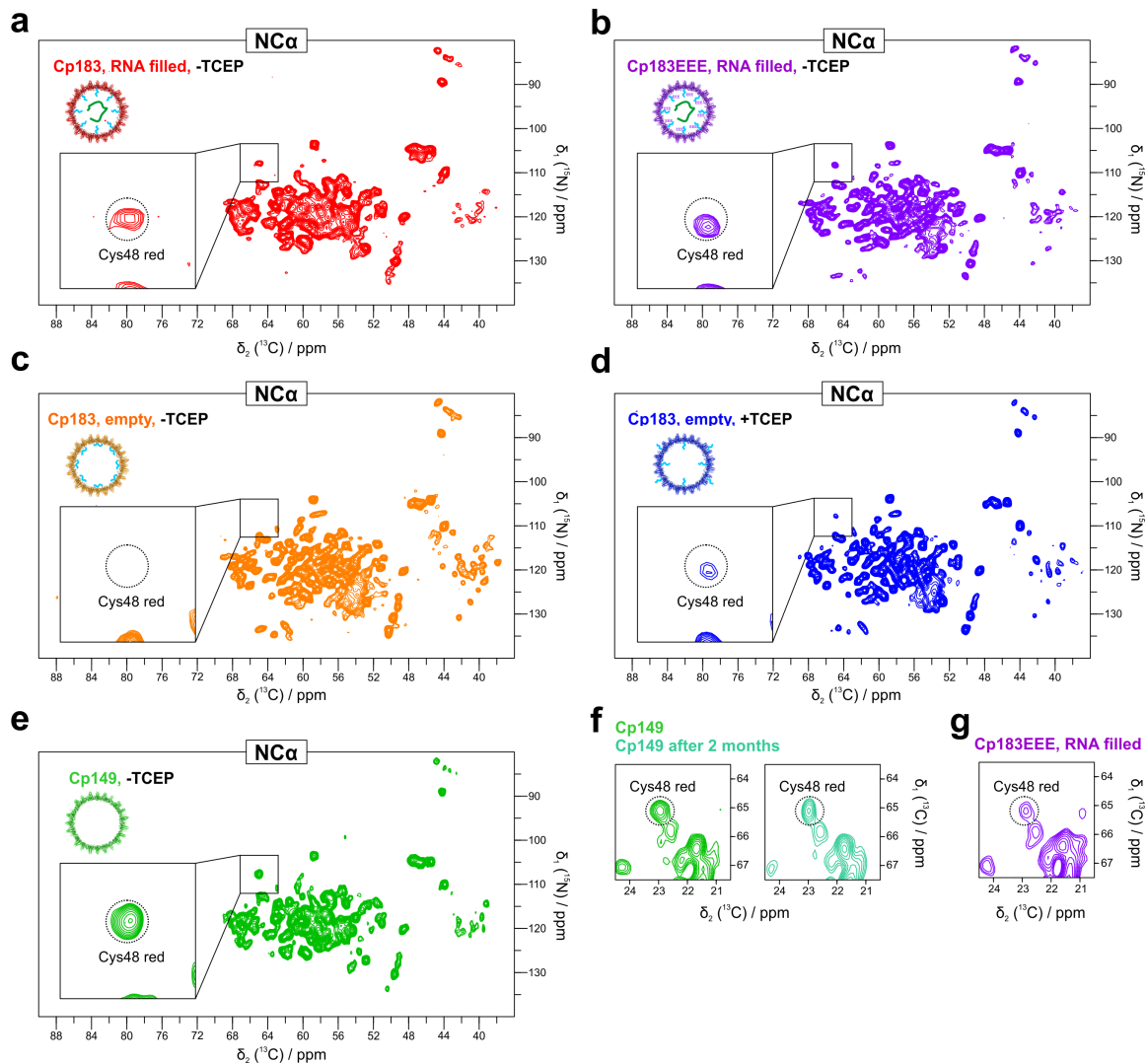


Figure S7 (a-e) 2D NMR spectra of different capsid types. The characteristic signal from Cys48 in reduced state is present in all spectra except capsids assembled from Cp183, emptied in an oxidizing environment (c). **(f)** Zoom on ^{13}C - ^{13}C DARR spectra of freshly prepared capsids assembled from Cp149 (left) and stored for 2 months at RT (right). Virtually identical spectral regions belonging to Cys48 in reduced state rule out the hypothetical existence of an intradimer Cys48-Cys48 disulphide bridge in the absence of the CTD. **(g)** Zoom on Cys48 Ca - $\text{C}\beta$ region from DARR spectra of capsids assembled from phosphorylation-mimic mutant Cp183EEE.

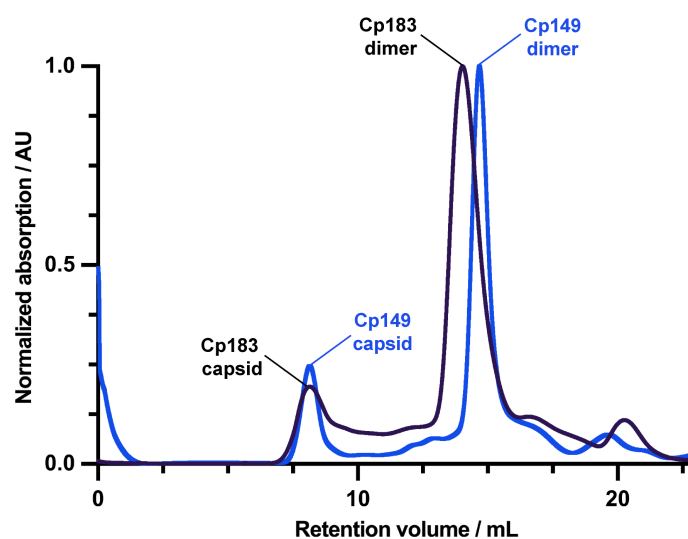


Figure S8 Size exclusion chromatograms obtained after disassembly of capsids following previously reported protocols for Cp183² and Cp149¹⁶⁻¹⁷. After treatment of Cp183 and Cp149 capsids with guanidine hydrochloride and urea, respectively, the samples were loaded on a Superdex 200 column to separate C protein dimers from residual capsids. The Cp149 dimers elute later than Cp183 dimers, consistent with their lower molecular weight (MW) of 34 kDa compared to 42 kDa. In addition, size exclusion chromatography (SEC) separates proteins based on their hydrodynamic radius (R_h). Intrinsically disordered proteins are less compact and therefore have a larger R_h and hence larger apparent MW compared to globular proteins¹⁸. The same holds for the intrinsically disordered C-terminal region in Cp183. As the Cp183 dimers are concentrated prior to solution-state NMR experiments, it is possible for the dimers to form transient oligomeric species, which could rationalize why the assembly domain is not visible in the solution-state spectrum in Figure 6b.

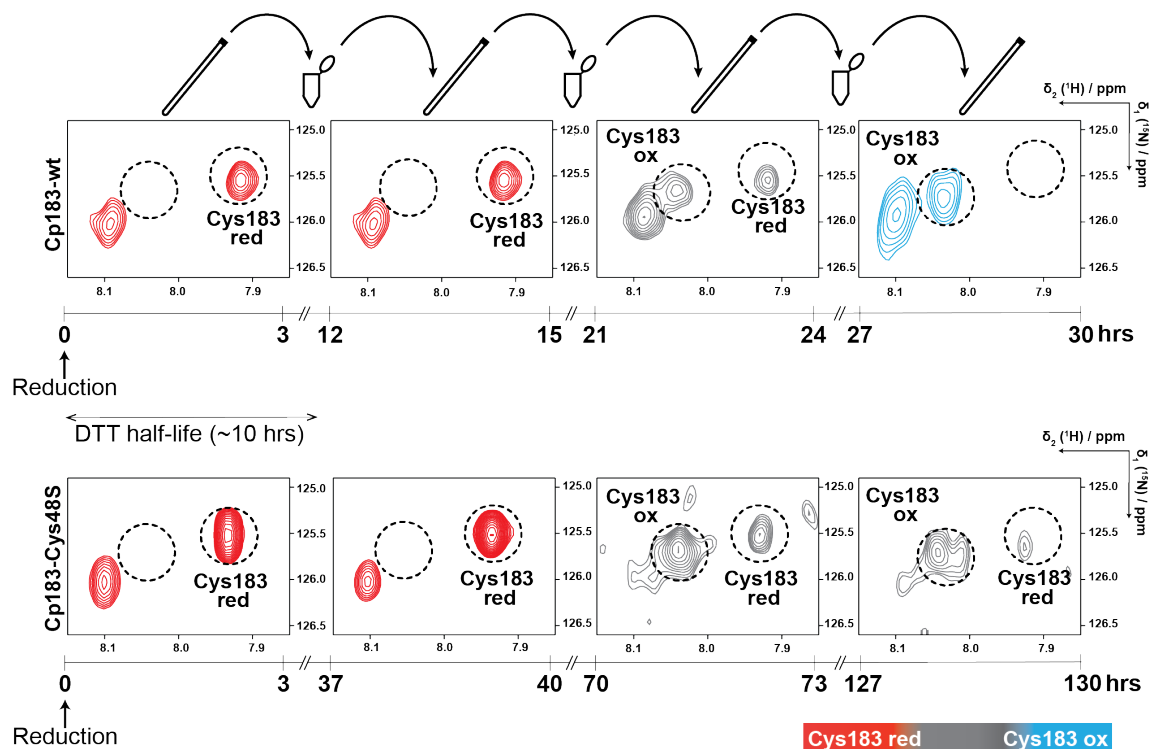


Figure S9 Dimers were produced by disassembly of wild-type and Cys48S-mutant Cp183 capsids. To ensure the fully reduced state, as confirmed by NMR, 2 mM DTT was added to both samples at the beginning of the experiment series. In between experiments, the samples were left partially covered at room temperature (open microcentrifuge tube sealed with holey Parafilm®). The oxidation state of Cys183 was monitored periodically. During the NMR measurement, the sample was kept at 293 K in a closed 3 mm NMR tube, sealed with parafilm. After an initial lag phase of at least 15 hours, the oxidation of wild-type dimers proceeded rapidly, from about half to completion within 3 hours. The initial lag phase is consistent with the half-life of DTT at 293 K at pH 7.5 of 10 hours¹⁹. In contrast, the oxidation of Cys48S-mutant dimers occurred over several days, much longer than the lag phase due to DTT depletion. A decrease in the intensity of all protein signals was noted in the spectra of Cys48S-mutant but not wild-type dimers over time. This could be caused by partial oligomerization due to inter-molecular disulfide linkages. The last two spectra of Cys48S-mutant dimers in gray were plotted closer to the noise level. The oxidation was reversible upon addition of more reducing agent (shown in Figure 6c). In conclusion, the Cys183-Cys48 disulfide linkage forms more rapidly than Cys183-Cys183.

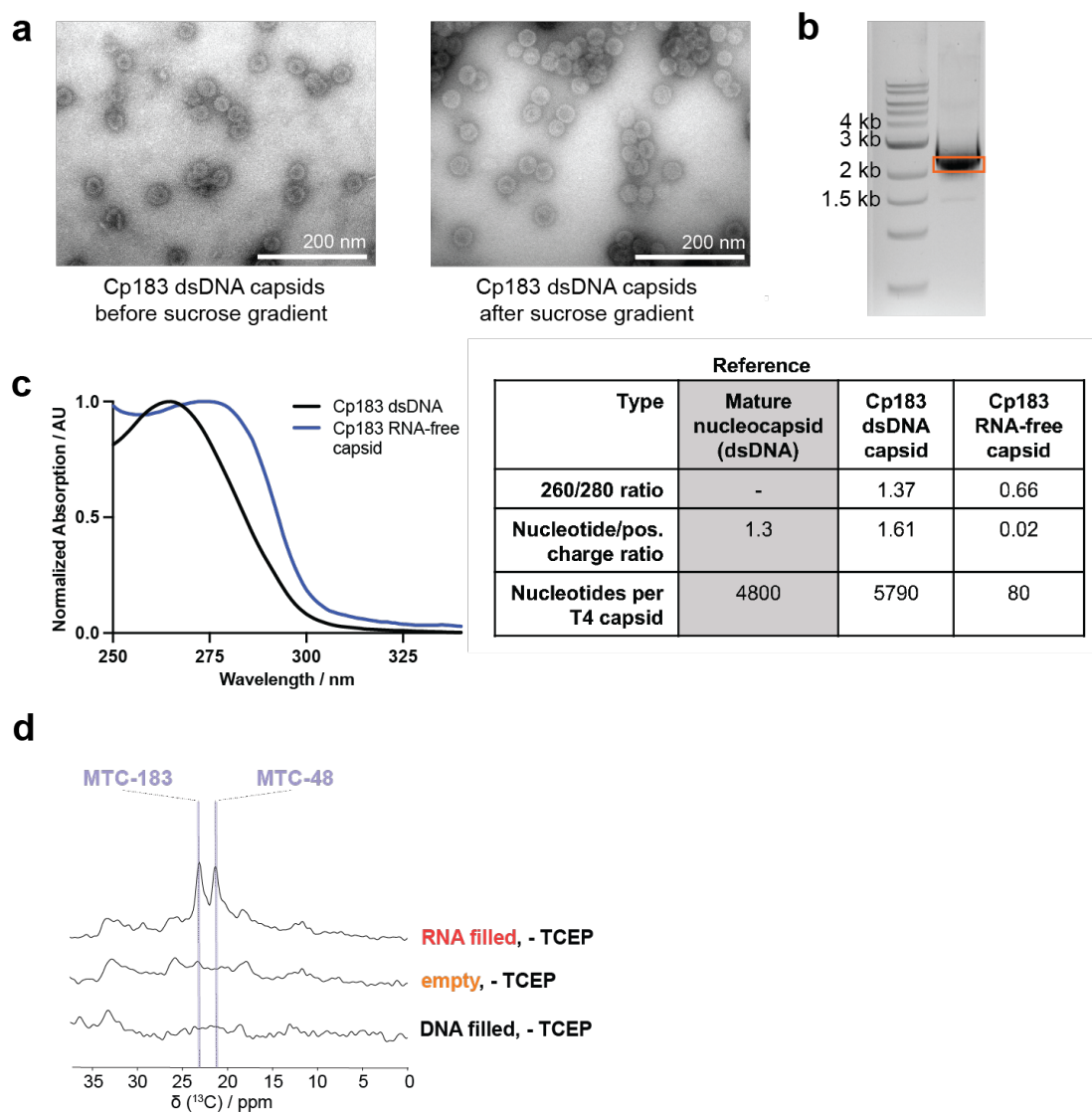


Figure S10 Cp183 dimers were assembled into capsids in the presence of a circular double-stranded (ds) bacterial plasmid pUC18 of 2686 base pairs. **(a)** Electron micrograph of capsids following DNA encapsidation, before and after sucrose gradient purification. **(b)** Native agarose gel electrophoresis of pUC18 after purification. The plasmid runs below the expected value of 2.7 kb, indicating the presence of a supercoiled plasmid. **(c)** UV absorption spectra of empty versus dsDNA-containing capsids indicate the incorporation of nucleic acids. The nucleotide content of the dsDNA-containing capsids was calculated using the extinction coefficient for DNA¹⁵. **(d)** Directly pulsed ¹³C spectra with proton decoupling of RNA-filled, empty and DNA-filled capsids labelled with MMTS under non-reducing conditions.

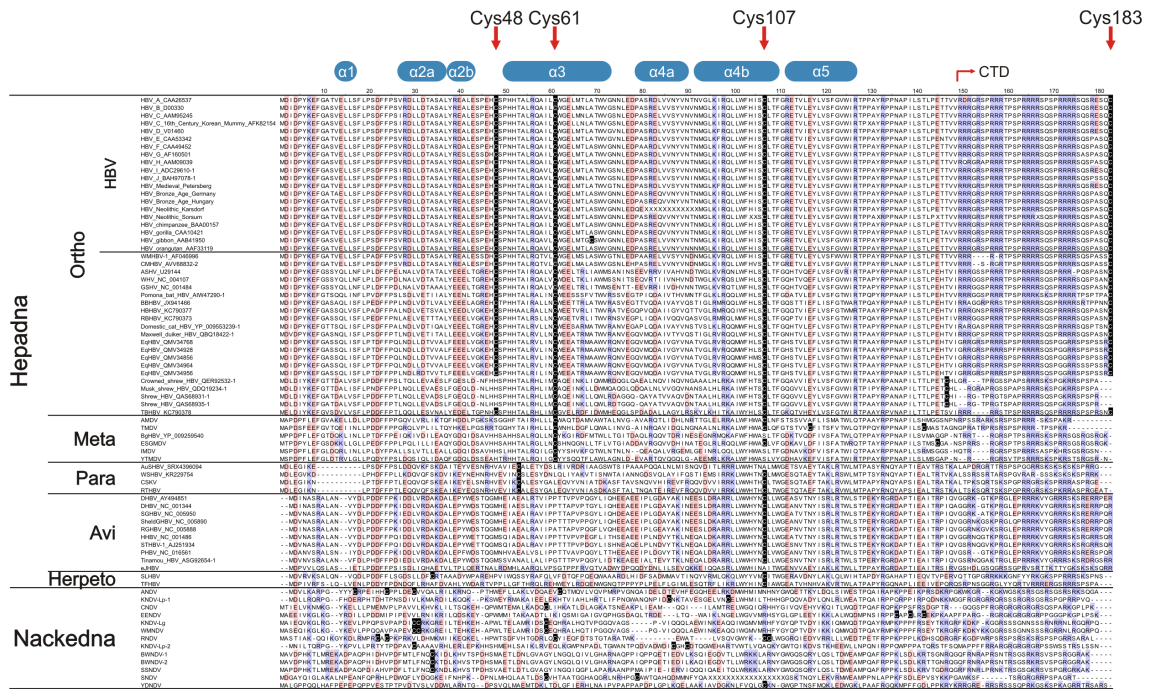


Figure S11 Alignment of C protein sequences of representative *hepadna*- and *nackednavirus* species. Alignment positions that appeared as insertions relative to HBV genotype D were cropped. Such insertions occur in the longer C protein sequences of *herpes*- and *avihepadnaviruses* in the non-helical parts, are highly diverse and thus, do not contain a meaningful phylogenetic signal. The location of the α -helices as determined in the crystal structure of HBV C protein¹⁴ is depicted on top with blue boxes. Cysteine residues are highlighted by black boxes. Cys107 is common to *hepadnaviruses* and demarcates them from the sister family of *nackednaviruses*. Cys61 is shared between mammalian ortho- and piscine *metahepadnaviruses*. Cys48 and Cys183 appear as a pair unique to the mammalian isolates. Both residues have been replaced in the viral isolates from shrews, which in turn have a shared cysteine corresponding to residue 147 in HBV (Cys147*). Fish *parahepadnaviruses* display a conserved cysteine at alignment position 53 (Cys53*).

Supporting tables

Capsid type		RNA-filled capsid		Empty capsid		Isolated C protein dimer	
Environment		oxidizing	reducing	oxidizing	reducing	oxidizing	reducing
Cryogenic MAS ssNMR	Dipolar (DARR, NhhC, hPhCC)	CTD engaged with RNA		-	-	-	-
Ambient temp. MAS ssNMR	Dipolar (DARR, NCA)	CTD not detected	CTD not detected	CTD not detected	CTD not detected	-	-
	Directly pulsed (PDSD)	CTD random coil	CTD random coil	CTD random coil	CTD random coil	-	-
	Scalar transfer HC-INEPT- (TOBSY)	CTD not detected	CTD not detected	CTD not detected	CTD not detected	-	-
Ambient temp. Solution NMR	Scalar (NH-HSQC)	undetected	undetected	undetected	reduced	CTD disordered	CTD disordered
Disulphide formation	Cys48 state	S-H	S-H	S-S	S-H	undetected	undetected
	Cys183 state	undetected	undetected	undetected	S-H	S-S	S-H
Interpretation							
CTD state		bound (engaged with RNA)	bound (engaged with RNA)	locked (disulphide formed)	free (no disulphide)	locked (disulphide formed)	free (no disulphide)
Figures with spectra		2b; 3b,d; 4d,e; 5b,c	5b	4f; 5b,e; 6a,c	4g; 5b,d; 6a,c	6b,c	6b,c

Table S1. Summary of NMR experiments that probe the state of CTD as a function of capsid type and redox environment.

Experiment	DARR	NhhC	hPhhC-C
B ₀ (T)	18.8		
MAS (Hz)	9470	9470	9330
90° pulse (μs)	2.8(¹ H)/5.5(¹³ C)	2.8(¹ H)/5.5(¹³ C)/7.0(¹⁵ N)	2.8(¹ H)/7.5(¹³ C)/5.0(³¹ P)
Transfer I	¹ H- ¹³ C CP	¹ H- ¹⁵ N CP	¹ H- ³¹ P CP
Field (kHz)	45(¹³ C)/90(¹ H)	36(¹⁵ N)/77(¹ H)	52(³¹ P)/85(¹ H)
Shape	Ramp 50-100	Ramp 50-100	Ramp 70-100
Contact time (ms)	2	1.5	2
Carrier (ppm)	100 (¹³ C)	120 (¹⁵ N)	0 (³¹ P)
Transfer II	DARR	¹⁵ N- ¹ H CP	³¹ P- ¹ H CP
Field (Hz)	9470	36(¹⁵ N)/77(¹ H)	52(³¹ P)/85(¹ H)
Shape	-	Ramp 50-100	Ramp 70-100
Time (ms)	50, 500	1.5	2
Carrier (ppm)	100 (¹³ C)	120 (¹⁵ N)	0 (³¹ P)
Transfer III		¹ H- ¹ H spin diff	¹ H- ¹ H spin diff
Field (Hz)		-	-
Shape		-	-
Time (ms)		0.5	0.5
Carrier (ppm)			
Transfer IV		¹ H- ¹³ C CP	¹ H- ¹³ C CP
Field (Hz)		31(¹³ C)/90(¹ H)	31(¹³ C)/90(¹ H)
Shape		Ramp 70-100	Ramp 70-100
Time (ms)		1	0.4
Carrier (ppm)		100 (¹³ C)	100 (¹³ C)
Transfer V			DARR
Field (Hz)			9330
Shape			-
Time (ms)			50
Carrier (ppm)			100 (¹³ C)
¹ H-decoupling during acq.	SPINAL64	SPINAL64	SPINAL64
Field (kHz)	90	90	90
t ₁ points	800	80	320
SW t ₁ (kHz)	53.3	16.2	56
Acq. time t ₁ (ms)	7.5	2.5	2.9
t ₂ points	1184	1800	1184
SW t ₂ (kHz)	59.5	81.5	59.5
Acq. Time t ₂ (ms)	9.9	11	9.9
Relaxation delay (s)	4	4	4
Number of scans	16	64	384

Table S2. Experimental parameters MAS ssNMR experiments under DNP at cryogenic temperature.

Experiment	DARR	dpPDS	Nca
B ₀ (T)	17.6		
MAS (Hz)	16650	16650	16650
90° pulse (μs)	1.45(¹ H)/5.0(¹³ C)	1.45(¹ H)/5.0(¹³ C)/7.0(¹⁵ N)	1.45(¹ H)/5.0(¹³ C)/4.5(¹⁵ N)
Transfer I	¹ H- ¹³ C CP	-	¹ H- ¹⁵ N CP
Field (kHz)	50(¹³ C)/75(¹ H)		58(¹⁵ N)/88(¹ H)
Shape	Ramp 70-100		Ramp 70-100
Contact time (ms)	1		1
Carrier (ppm)	100 (¹³ C)	100 (¹³ C)	100 (¹³ C)
Transfer II	DARR	PDS	N-Ca
Field (Hz)	16650 (¹ H)	-	57(¹⁵ N)/30(¹³ C)
Shape	-	-	OC-DCP*
Time (ms)	25, 500	200	4.2
Carrier (ppm)	100 (¹³ C)	100 (¹³ C)	55 (¹³ C) 122 (¹⁵ N)
¹ H-decoupling during acq.	SPINAL64	SPINAL64	SPINAL64
Field (kHz)	90	90	90
t ₁ points	1200-1400	500	200
SW t ₁ (kHz)	50	50	8
Acq. time t ₁ (ms)	14	14	12.5
t ₂ points	3072	3072	3072
SW t ₂ (kHz)	57	57	57
Acq. Time t ₂ (ms)	27	27	27
Relaxation delay (s)	2.5	0.8	2.5
Number of scans	48-64	160	128-256

*Optimal control double CP ²⁰

Table S3. Experimental parameters for MAS ssNMR experiments at ambient temperature.

Supporting references

1. Wingfield, P. T.; Stahl, S. J.; Williams, R. W.; Steven, A. C., Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and in vitro capsid assembly. *Biochemistry* **1995**, *34* (15), 4919-32.
2. Porterfield, J. Z.; Dhason, M. S.; Loeb, D. D.; Nassal, M.; Stray, S. J.; Zlotnick, A., Full-length hepatitis B virus core protein packages viral and heterologous RNA with similarly high levels of cooperativity. *J Virol* **2010**, *84* (14), 7174-84.
3. Strods, A.; Ose, V.; Bogans, J.; Cielens, I.; Kalnins, G.; Radovica, I.; Kazaks, A.; Pumpens, P.; Renhofa, R., Preparation by alkaline treatment and detailed characterisation of empty hepatitis B virus core particles for vaccine and gene therapy applications. *Sci Rep* **2015**, *5*, 11639.
4. Ravera, E.; Corzilius, B.; Michaelis, V. K.; Luchinat, C.; Griffin, R. G.; Bertini, I., DNP-enhanced MAS NMR of bovine serum albumin sediments and solutions. *J Phys Chem B* **2014**, *118* (11), 2957-65.
5. Ravera, E.; Corzilius, B.; Michaelis, V. K.; Rosa, C.; Griffin, R. G.; Luchinat, C.; Bertini, I., Dynamic nuclear polarization of sedimented solutes. *J Am Chem Soc* **2013**, *135* (5), 1641-4.
6. Bockmann, A.; Gardienet, C.; Verel, R.; Hunkeler, A.; Loquet, A.; Pintacuda, G.; Emsley, L.; Meier, B. H.; Lesage, A., Characterization of different water pools in solid-state NMR protein samples. *J Biomol NMR* **2009**, *45* (3), 319-27.
7. Stevens, T. J.; Fogh, R. H.; Boucher, W.; Higman, V. A.; Eisenmenger, F.; Bardiaux, B.; van Rossum, B. J.; Oschkinat, H.; Laue, E. D., A software framework for analysing solid-state MAS NMR data. *J Biomol NMR* **2011**, *51* (4), 437-47.
8. Lecoq, L.; Wang, S.; Wiegand, T.; Bressanelli, S.; Nassal, M.; Meier, B. H.; Bockmann, A., Solid-state [^{13}C - ^{15}N] NMR resonance assignment of hepatitis B virus core protein. *Biomol NMR Assign* **2018**, *12* (1), 205-214.
9. Jaroniec, C. P.; Tounge, B. A.; Rienstra, C. M.; Herzfeld, J.; Griffin, R. G., Measurement of C-13-N-15 distances in uniformly C-13 labeled biomolecules: J-decoupled REDOR. *J Am Chem Soc* **1999**, *121* (43), 10237-10238.
10. Gehman, J. D.; Separovic, F.; Lu, K.; Mehta, A. K., Boltzmann statistics rotational-echo double-resonance analysis. *J Phys Chem B* **2007**, *111* (27), 7802-11.
11. Lauber, C.; Seitz, S.; Mattei, S.; Suh, A.; Beck, J.; Herstein, J.; Borold, J.; Salzburger, W.; Kaderali, L.; Briggs, J. A. G.; Bartenschlager, R., Deciphering the Origin and Evolution of Hepatitis B Viruses by Means of a Family of Non-enveloped Fish Viruses. *Cell Host Microbe* **2017**, *22* (3), 387-399 e6.
12. Kumar, S.; Stecher, G.; Tamura, K., MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **2016**, *33* (7), 1870-4.
13. Pinon, A. C.; Schlagnitweit, J.; Berruyer, P.; Rossini, A. J.; Lelli, M.; Socie, E.; Tang, M.; Pham, T.; Lesage, A.; Schantz, S.; Emsley, L., Measuring Nano- to Microstructures from Relayed Dynamic Nuclear Polarization NMR. *J Phys Chem* **2017**, *121* (29), 15993-16005.
14. Wynne, S. A.; Crowther, R. A.; Leslie, A. G., The crystal structure of the human hepatitis B virus capsid. *Mol Cell* **1999**, *3* (6), 771-80.
15. Porterfield, J. Z.; Zlotnick, A., A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. *Virology* **2010**, *407* (2), 281-8.
16. Zlotnick, A.; Cheng, N.; Conway, J. F.; Booy, F. P.; Steven, A. C.; Stahl, S. J.; Wingfield, P. T., Dimorphism of hepatitis B virus capsids is strongly influenced by the C-terminus of the capsid protein. *Biochemistry* **1996**, *35* (23), 7412-21.

17. Zhao, Z.; Wang, J. C.; Segura, C. P.; Hadden-Perilla, J. A.; Zlotnick, A., The Integrity of the Intradimer Interface of the Hepatitis B Virus Capsid Protein Dimer Regulates Capsid Self-Assembly. *Acs Chem Biol* **2020**, *15* (12), 3124-3132.
18. Uversky, V. N., What does it mean to be natively unfolded? *Eur J Biochem* **2002**, *269* (1), 2-12.
19. Stevens, R.; Stevens, L.; Price, N. C., The Stabilities of Various Thiol Compounds Used in Protein Purifications. *Biochem Educ* **1983**, *11* (2), 70-70.
20. Tosner, Z.; Vosegaard, T.; Kehlet, C.; Khaneja, N.; Glaser, S. J.; Nielsen, N. C., Optimal control in NMR spectroscopy: numerical implementation in SIMPSON. *J Magn Reson* **2009**, *197* (2), 120-34.