Making the invisible visible: fast magic-angle-spinning NMR reveals the evasive hepatitis B virus capsid functional C-terminal domain

Morgane Callon[1#*], Alexander A. Malär[1#], Lauriane Lecoq[2], Marie Dujardin[2], Marie-Laure Fogeron[2], Shishan Wang[2], Maarten Schledorn[1], Thomas Bauer[1], Michael Nassal[3], Anja Böckmann[2], Beat H. Meier[1*]

[1] Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland

[2] Molecular Microbiology and Structural Biochemistry (MMSB) UMR 5086 CNRS/Université de Lyon, Labex Ecofect, 7 passage du Vercors, 69367 Lyon, France

[3] University Hospital Freiburg, Dept. of Medicine II / Molecular Biology, Medical Center, University of Freiburg, Germany

[#] These authors contributed equally

*Corresponding authors morgane.callon@phys.chem.ethz.ch, beme@ethz.ch

Abstract

Experimentally determined protein structures often feature missing domains. One example is the C terminal domain (CTD) of the hepatitis B virus capsid protein, a functionally central part of this assembly, crucial in regulated nucleic-acid interactions, cellular trafficking, nuclear import, particle assembly and maturation. However, its structure remained elusive to all current techniques, including NMR. Here we show that the recently developed proton-detected fast magic-angle-spinning solid-state NMR at >100 kHz MAS is a game changer that allows to detect this domain and unveil its structural and dynamic behavior. We describe the experimental framework used and compare the domain's behavior in different capsid states. The developed approaches extend solid-state NMR observations to residues characterized by large-amplitude motion on the microsecond timescale, and shall allow to shed light on other flexible protein domains still lacking their structural and dynamic characterization.

Introduction

It regularly happens, be it in X-ray crystallography, cryo-electron microscopy (EM) or nuclear magnetic resonance (NMR) spectroscopy, that parts or even entire domains of a protein remain invisible or poorly defined in the acquired data. We here do not refer to "invisible" transiently populated high-energy structures which are invisible due to low population, but to conformations of the protein in its ground state. The absence of well-defined electron density for these domains in X-ray and cryo-EM signals is due to positional disorder, be it static or dynamic. At physiological temperatures, the disorder is often dynamic and can be detected and characterized by NMR, in particular if the dynamics is fast on the NMR timescale [1]. For example, intrinsically disordered regions (IDR), can be analyzed both in solution and solid-state NMR using similar experiments. This has indeed been used in the past to selectively observe rigid versus particularly mobile regions in solid-state NMR [2–4]. Some intermediate dynamical regimes do however interfere with the observation of the corresponding resonances. This happens in solution-state NMR, where the slowing of the overall tumbling for larger proteins causes rapid transverse relaxation and line broadening beyond detection [5,6] (Figure 1a). Solid-state NMR in contrast does specifically target larger proteins or assemblies, in crystalline form or as sediments, that show no overall tumbling. Still, line broadening can be induced by internal motions, in particular on intermediate time scales [4], at a correlation time comparable with the inverse magic-angle-spinning (MAS) rotation frequency (τ_c about 1 μs), which leads to incoherent, i.e. relaxation-induced, line broadening [7–11]. This is shown in Figure 1b, where incoherent line broadening is calculated for an isolated ^{1}H-^{15}N two-spin system. It can be seen that for motions at a time scale similar to the MAS frequency, substantial line broadening can occur, notably at slower MAS frequencies. However, for fast spinning, e.g. at 100 kHz, predicted linewidths do not exceed 200 Hz using a simple model (Figure 1b). Thus, while the loss of resonance lines in solid-state NMR has sometimes been attributed to broadening beyond the detection limit of the NMR resonances [4,12,13], these simulations are a first indication that this may not be the case for MAS frequencies of 100 kHz and above, even for the most unfavorable motional regimes.

An alternative cause for disappearing resonances lies in inefficient heteronuclear polarization transfer either by (i) cross polarization (CP) through the dipolar couplings (Figure 1c-d) [14–17] or, if the motion is fast and close to isotropic with averaged dipolar interaction, through (ii) transfer *via* the isotropic J-couplings as in Insensitive Nuclei Enhancement by Polarization Transfer (INEPT) (Figure 1c-d) [18–24]. INEPT relies on long transverse relaxation times (T_2'), which are increased through faster spinning, deuteration, and absence of intermediate dynamics.

As an alternative, homonuclear 2D exchange spectroscopy (EXSY) [25] *via* spin diffusion [26–30] or nuclear Overhauser effect (NOE) can be used to observe proton homonuclear polarization transfer (Figure 1d). While spin diffusion is most efficient in rigid samples, the NOE relies on the presence of molecular motion; while the two mechanisms have a different dependence on spinning frequency and temperature, in practice, it is often difficult to experimentally distinguish the two effects due to the limited range of spinning frequencies and temperatures that can be probed in a protein sample.

We here aim to recover the lost resonances of the functional C-terminal domain (CTD) of the hepatitis B virus (HBV) capsid, which almost completely escaped detection by virtually all structural methods used until now, despite its central role in many important aspects of virion formation and maturation [31]. For this, we need to resolve the question whether, in the fast-spinning experiments we employ, the disappearance of resonances is due to line broadening beyond detection or to inefficient polarization transfer due to largely averaged dipolar couplings, and use this information to derive experiments able to detect the resonances. The HBV capsid is formed by 240 copies of the core protein (Cp) [32]. In immature capsids, packaged pregenomic RNA (pgRNA) is reverse-transcribed into the viral DNA inside the capsid [33] by the viral polymerase. Mature nucleocapsids can take two different pathways; they are either directed towards the nucleus to amplify the viral DNA pool, or to envelopment and release as new infectious virions. The Cp contains an N-terminal assembly domain (NTD) of 140 residues, a 10-residue linker, and the highly basic 34-residue CTD (Figure 1e) required for encapsidation of the pgRNA amongst other activities. The NTD is sufficient for self-assembly at high enough concentrations, and has been characterized via cryo-electron microscopy [34,35], X-ray crystallography [36], and solid-state NMR [37,38,38]. The CTD has been described as a flexible, disordered domain [39]; it shows four arginine-rich repeats (Figure 1e) commonly found in RNA- and DNA-binding proteins [40] and interacts with the pgRNA during capsid assembly. Therefore, the phosphorylated CTD is not moving freely, but engages in interactions, likely of electrostatic nature between the arginine side chains and the pgRNA phosphate groups. It might provide a rationale for the only transient exposure of the CTD [41–43] for its interaction with cellular proteins, for instance for nuclear import processes [35]. For this, parts of the CTD must, at least transiently, be exposed to the exterior of mature capsids [44–47]. CTD phosphorylation is known to regulate the capsid's interactions with different partners during the viral life cycle [41,48]. In sample preparations mimicking different capsid forms, the CTD remained invisible in solution-state (INEPT), as well as solid-state NMR (CP) experiments, and spectra from capsids formed from the full-length protein Cp183 show virtually identical spectra to those of capsids formed from the assembly domain (Cp149) alone [49,50].
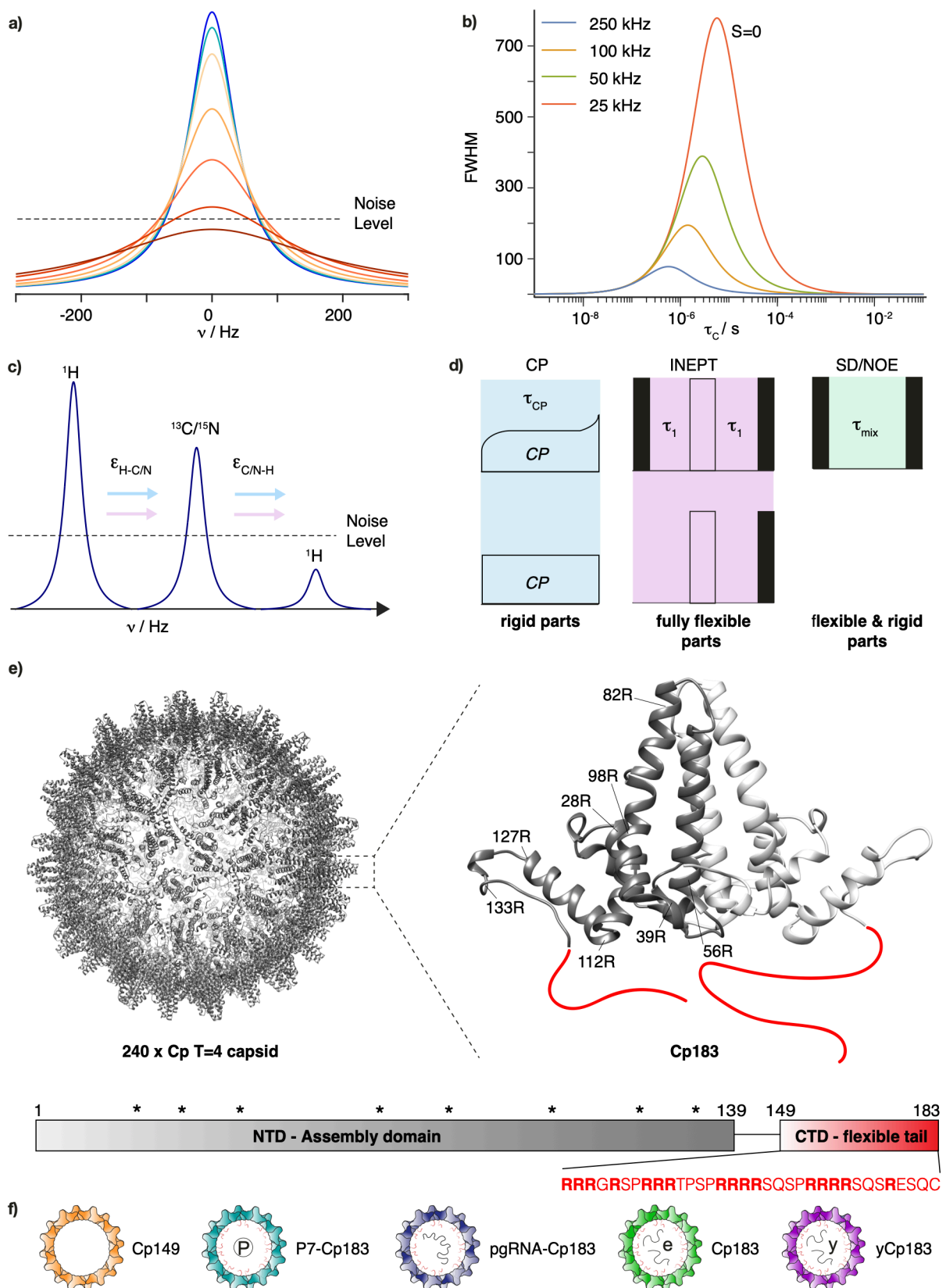
a)

ν / Hz

Noise Level

b)

FWHM

250 kHz
100 kHz
50 kHz
25 kHz

S=0

τ_C / s

c)

^{1}H

^{13}C/^{15}N

$\varepsilon_{H-C/N}$ $\varepsilon_{C/N-H}$

Noise Level

^{1}H

ν / Hz

d)

CP	INEPT	SD/NOE
τ_{CP}	τ_1 τ_1	τ_{mix}
CP		
CP		
rigid parts	**fully flexible parts**	**flexible & rigid parts**

e)

240 x Cp T=4 capsid

82R
98R
28R
127R
133R
112R   39R   56R

Cp183

1   *   *   *   *   *   *   *   * 139   149   183

NTD - Assembly domain   **CTD - flexible tail**

RRRGR**SP**RRR**TPSP**RRRR**SQSP**RRRR**SQS**R**ESQC

f)

Cp149   ⓟ P7-Cp183   pgRNA-Cp183   ⓔ Cp183   ⓨ yCp183

FIGURE 1 a) Gaussian NMR line shape with increasing full width at half maximum (FWHM), illustrating how line broadening beyond detection leads to peak disappearance in NMR spectra. **b)** Incoherent proton FWHM as function of the motional correlation time of the NH vector (τ_C) simulated for an isolated ^{1}H-^{15}N two-spin system with the parameters of an amide group in

a protein and an order parameter of S=0 (strongest broadening) at MAS frequencies between 25 and 250 kHz. Microsecond time scales of the underlying dynamics can lead to incoherent line broadening, but can be reduced at faster MAS frequencies. **c)** Scheme highlighting how NMR signal can be lost due to inefficient transfers. Polarization-transfer efficiencies (ε) can become too low when dipolar couplings are averaged by motion or when INEPT transfers are inefficient due to fast T_2' relaxation times, which lead to loss of signal. **d)** Summary of different polarization-transfer schemes available in solid-state NMR and their best efficiencies for different motional regimes. In contrast to heteronuclear CP and INEPT transfers, spin diffusion (SD)/NOE transfer is homonuclear and mediated by the network of strong ^{1}H-^{1}H dipolar couplings. **e)** Representation of the hepatitis B virus capsid 3D structure (PDB: 1qgt [36]), with a zoom on the dimeric subunit structure, and its domain organization with the amino-acid sequence of the CTD in red. The X-ray structure was determined on the isolated NTD. The stars show positions of the arginine residues in the NTD. **f)** Schematic representation of the different capsid forms investigated: truncated Cp149 for reference of capsids devoid of the CTD; empty hyperphosphorylated P7-Cp183; pg-RNA containing reassembled capsids; *E. coli* RNA containing capsids; and yeast-RNA containing reassembled capsids.

Here, we are able for the first time to detect the NMR signals from the CTD using EXSY-type proton magnetization transfers at fast MAS frequencies. This allowed us to measure its NMR relaxation properties, and revealed that poor heteronuclear polarization transfer in both CP and INEPT rather than line broadening beyond detection causes the loss of the signals. We find that the CTD's chemical shifts center around the random-coil values. Surprisingly, the relaxation times measured on CTDs of empty, phosphorylated or nucleic-acid filled capsids are very similar indicating that the phosphorylated CTD is not moving freely, but engages in interactions, likely of electrostatic nature, between the arginine side chains and the phosphate groups to form a network at the capsid inner lining. The capsid's CTDs must form dynamically similar to Coulomb interaction-driven condensates, either in interaction with the viral pgRNA, or with other CTDs in their phosphorylated form.

Results

The CTD is visible in CP-NMR spectra only at low temperatures

Cp183, when compared to Cp149 capsids, should show three times the number of peaks at the typical arginine (Arg) resonance positions (Figure 1e). However, 2D hNH CP spectra of Cp149 (taken from [37]) and Cp183 (taken from [50]) virtually overlay (Figure S1a), with comparable peak intensities for both Cp149 and Cp183. This is as well the case for the ^{13}C-detected DARR spectra [49], and for the 2D hCH INEPT spectra (Figure S1b) recorded at 17.5 kHz, where highly flexible residues moving stochastically on fast timescales should be observable. Arg residues from the CTD thus remain invisible in these spectra.

To confirm the hypothesis that this absence is due to dynamics, we first used a brute-force approach to slow down motions, and recorded 2D NC spectra of Cp183 at low temperatures. Figure S2 shows an overlay of the spectra recorded at 4 °C, -30 °C and -168 °C. Indeed, an increase of the intensities of Arg peaks by a factor of roughly 3 is observed at the subfreezing

temperatures. This clearly points to signal loss at room temperature through motional effects, which is recovered when freezing out dynamics. Since the broad lines observed at low temperatures impede further detailed characterization [51,52] and obviously also defeat any dynamics measurements, we aimed in the following to devise a strategy to observe the CTD at temperatures close to physiological conditions.

A strategy to measure CTD dynamic parameters at room temperature

As mentioned above, peaks disappear either through broadening beyond detection, or poor polarization transfers, both caused by dynamical behavior. A first approach in this context is to find a way to assess the actual spectral linewidths, as well as the relaxation times important for the different polarization-transfer schemes employed. Due to the absence of resonances of the CTD in heteronuclear correlation spectra, we decided to directly investigate the proton spectra. To reduce spectral crowding, we used selective ^{1}H,^{13}C,^{15}N isotope labeling of arginine residues (Arg), which do not suffer from scrambling in *E. coli* [53], while using per-deuteration of all other residues. In addition, the samples were exchanged to a D_2O-based buffer (for resulting labelling schemes, see Figure S3). Thus, only resonances from Arg are observed in the protein ^{1}H spectra. Five different capsid samples were investigated: capsids containing pgRNA (Cp183-pgRNA), *E.coli* RNA (Cp183) or yeast RNA (yCp183); phosphorylated Cp at all seven phosphorylation sites (P7-Cp183 [48]); and truncated Cp149 as a reference sample to identify the contributions from the 8 Arg residues in the NTD (see Figure 1f and Figure S3).

While ^{1}H 1D spectra showed signal for all samples, they are not sufficiently resolved to identify individual signals. We thus recorded EXchange SpectroscopY (EXSY) spectra [25] (for pulse sequence see Figure S4a) which include a longitudinal ^{1}H-^{1}H polarization-transfer step to identify resonances connected by the polarization transfer. The resulting spectra (Figures 2a-c, full view in Figure S5) indeed show strong Arg intra-residue ^{1}H-^{1}H cross peaks for the Cp183 samples, and only weak ones, and at different chemical shifts, for Cp149. This can also be clearly seen in the first traces of the 2D spectra (Figure 2d), where the Arg resonances are about three times higher for Cp183 when compared to Cp149. The traces extracted at 1.67 ppm (Figure 2.e) show that the strong cross peaks, only present in Cp183, localize at the random-coil chemical shifts (4.34 ppm for H_α, 3.21 ppm for $H_{\delta2/3}$), and do not coincide with the weaker Arg resonances from the NTD (Figure 2e top). Together, this demonstrates that the cross peaks observed in the EXSY spectra mainly originate from CTD resonances, which are thus unambiguously detected for the first time at room temperature.
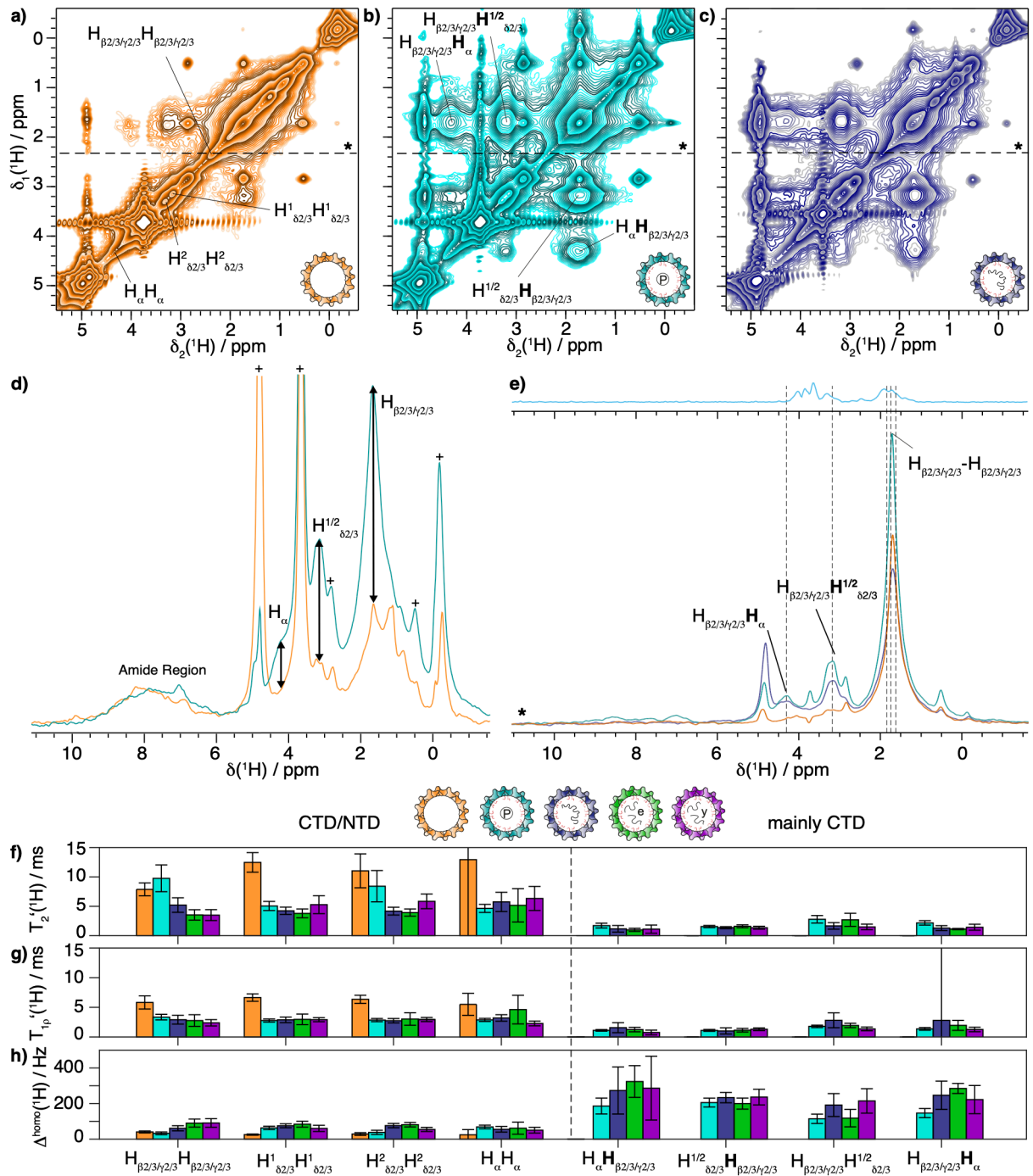
FIGURE 2 ^{1}H-^{1}H spectroscopy on selectively Arginine-labelled samples based on ^{1}H-^{1}H homonuclear mixing by spin diffusion or NOE effects. EXSY spectra using 50 ms mixing time of **a)** Cp149, **b)** P7-Cp183 and **c)** Cp183-pgRNA recorded at 100 kHz MAS and 850 MHz proton frequency. **d)** Overlay of the first FIDs of EXSY spectra of Cp149 and P7-Cp183 recorded with a mixing time of 10 ms. The peaks labelled with a cross correspond to buffer signals. **e)** Overlay of extracted traces from EXSY spectra at δ_1(^{1}H) = 1.67 ppm with the projection of a standard hCH CP based experiment recorded on P7-Cp183 (Figure S8b) on top. The dashed lines indicate random-coil chemical shift positions of Arg [60]. **f)** T_2'(^{1}H) and **g)** $T_{1\rho}$(^{1}H) values recorded from the EXSY spectra for all five capsid samples (Figure 1f). **h)** T_2'(^{1}H) corresponding homogeneous line width Δ^{homo}(^{1}H). In the figure the proton of the cross peak where the relaxation takes place, is labelled in bold.

That the use of the dipolar ^{1}H-^{1}H homonuclear transfer step was successful in creating cross peaks can be explained by higher-order ^{1}H-^{1}H dipolar couplings still present [28,54–56,56–59],

despite the fast MAS and the molecular motion of the CTD. Alternatively, NOE effects can be present in the CTD due to its high flexibility, and can add to, or replace, spin diffusion. Thus, the slower polarization-transfer buildup curves observed for Cp149 compared to Cp183 cross peaks are either attributed to their lower proton abundance in the NTD (only isolated Arg are labelled there, whereas in the CTD they cluster together) or to reduced NOE effects (Figure S6). It clearly follows from the above observations that line broadening beyond detection is not causing the loss of the resonances in the CP spectra.

As cross peaks exclusively from the CTD can be detected in EXSY spectra, we could measure and compare dynamic parameters of the CTD in the different capsid forms. We determined both proton transverse $T_2'(^1\text{H})$ and rotating-frame $T_{1\rho}(^1\text{H})$ relaxation times (pulse sequence in Figure S4b and c respectively). The results for all five capsid samples are shown in Figure 2f-h (relaxation decay curves in Figure S7) and are summarized in Table 1. The cross peaks and diagonal peaks are treated separately, as the major contribution to the cross peaks is from the CTD ("mainly CTD"), while the diagonal peaks contain both NTD and CTD resonances in the full-length samples (NTD/CTD). The diagonal peak decay curves would thus be multiexponential but at the given SNR this cannot be fitted. One can see though, that the relaxation times of Cp149 (diagonal peaks) are longer by a factor of 3 longer than in Cp183 (cross peaks).

TABLE 1 Averaged relaxation parameters obtained from the ^{1}H-^{1}H-2D relaxation measurements of the five samples presented in Figure 2f-h. Averages have been taken over diagonal (NTD/CTD) and off-diagonal peaks (mainly CTD) separately. Values given in parenthesis correspond to the standard deviation, Δ denotes the linewidth.

	$T_{2,\text{diag}}'\,(^1\text{H})$	$T_{2,\text{off-diag}}'\,(^1\text{H})$	$\Delta_{\text{diag}}^{\text{homo}}\,(^1\text{H})$	$\Delta_{\text{off-diag}}^{\text{homo}}\,(^1\text{H})$	$T_{1\rho,\text{diag}}\,(^1\text{H})$	$T_{1\rho,\text{off-diag}}\,(^1\text{H})$
Cp149	11(2) ms	-	29(5) Hz	-	6.1(5) ms	-
P7-Cp183	7(2) ms	2.0(5) ms	55(13) Hz	159(40) Hz	3.0(3) ms	1.3(3) ms
Cp183-pgRNA	4.8(7) ms	1.4(2) ms	66(9) Hz	227(40) Hz	3.0 (2) ms	2.1(9) ms
Cp183	4.1(7) ms	1.6(8) ms	78(13) Hz	199(100) Hz	3.3(9) ms	1.6(4) ms
yCp183	5(1) ms	1.3(2) ms	64(13) Hz	245(38) Hz	2.6(3) ms	1.2(3) ms

Further analysis of the rotating-frame-relaxation parameters (lower panel in Figure 2f) reveals a decrease by about a factor of 2 or 3 in $T_{1\rho}$ between Cp149 (diagonal peaks) and Cp183 (off-diagonal peaks to select CTD), which clearly reveals the presence of significant motion in the CTD. Indeed, short $T_{1\rho}(^1H)$ are associated with high mobility of the protein on the microsecond timescale and thus stochastic averaging of the dipolar interactions, which provides a rational for magnetization loss during CP periods. The relationship to the order parameter and correlation time is given by equations (1) and (2) in the Material and Methods part. We compare the observed $T_{1\rho}$ relaxation times between 1.3 and 2.1 ms (Table 1), with

the simple ^{1}H-^{15}N dipole-dipole $T_{1\rho}$ (^1H) 2 spin model that predicts a $T_{1\rho}$ (^1H) of 0.2 μs where we have neglected the effect of chemical-shift anisotropy and proton-proton interactions [61]. It can be concluded that the correlation time is around $\tau \approx 1.4$ μs within a factor of ten and the order parameter S is below 0.9, in order to explain the measured relaxation-rate constants by the simple model.

The Arg proton chemical shifts clearly cluster around the typical random-coil shifts; this is a further indication for the presence of considerable dynamics in the CTD. It should however be realized that the related motion is not necessarily isotropic and does not fully average the dipolar interactions, but rather leads to a significant reduction thereof [62].

Interestingly, the CTD relaxation times are highly similar for all full-length samples, whether containing RNA or not. This provides evidence that in absence of RNA, when phosphorylated CTD shows dynamics comparable to the RNA-containing capsids. This clearly indicates that the CTD, when phosphorylated, is not as flexible as an IDP, but must instead engage in interactions, likely amongst itself, within the capsid.
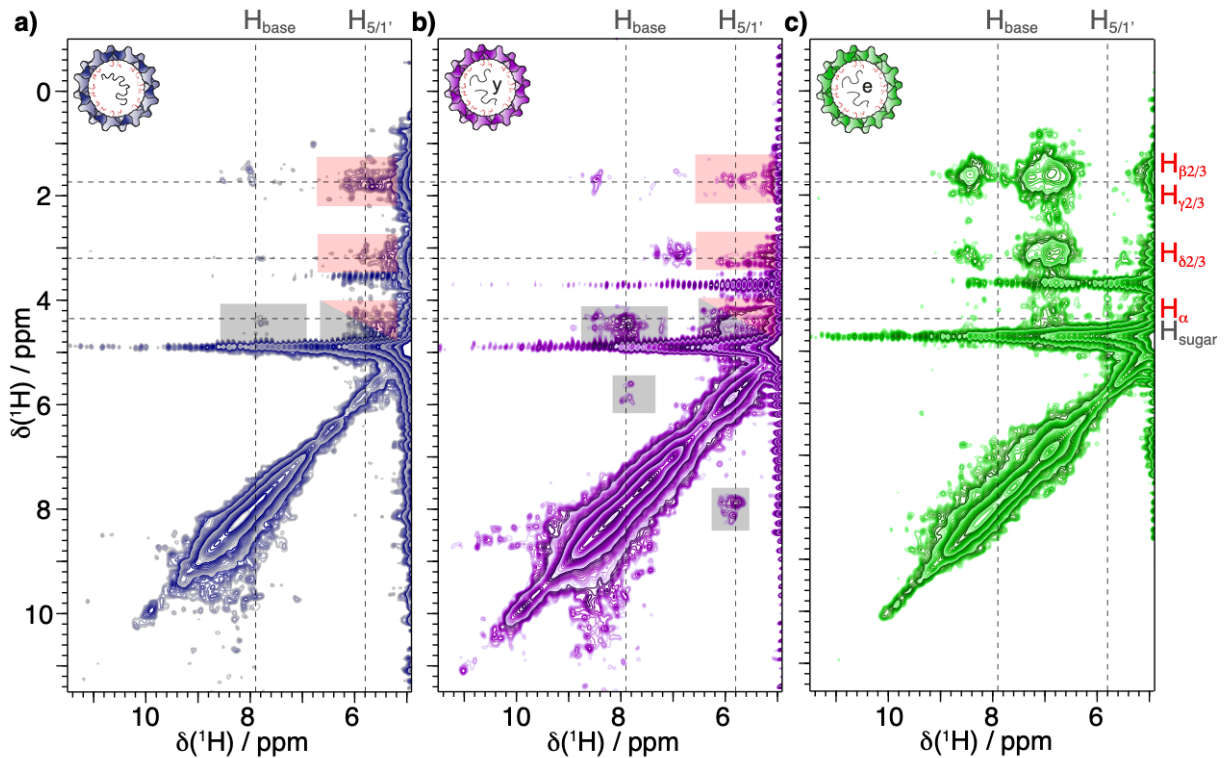

Figure 3. RNA-CTD interaction is detected in EXSY spectra. EXSY spectra using 50 ms ^{1}H-^{1}H homonuclear mixing time of **a)** Cp183-pgRNA, **b)** yCp183 and **c)** Cp183 samples recorded at 100 kHz MAS and 850 MHz proton frequency. The RNA proton chemical shifts are indicated by vertical dashed lines while the horizontal ones indicate the positions of the cross peaks observed in the 2D EXSY spectra (Figure 2a-c). The red-shaded boxes highlight protein-RNA interactions, the grey-shaded boxes RNA-RNA interactions. We attribute the presence of stronger amide proton-sidechain proton cross peaks in Cp183 to a lower level of deuteration at exchangeable positions.

Finally, the EXSY spectra of the Cp183 samples containing RNA reveal cross peaks between the Arg residues and RNA resonances (Figure 3) indicating that the Arg residues are in contact with the nucleic acids, such that the RNA-protein couplings are strong enough for spin diffusion or NOE transfer to be active. On the RNA side, the visible resonances can be assigned to the base and sugar moiety protons (red boxes in Figure 3a-b). These cross peaks are weaker than the intra-residue ones (grey boxes in Figure 3a-b) due to the greater internuclear distances. No RNA resonances are observed in the *E. coli*-assembled Cp183 spectrum (Figure 3c) since the RNA is deuterated (see Figure S3).

Revisiting heteronuclear 2D experiments

Heteronuclear spectroscopy is a cornerstone of NMR, since the resolution and dispersion added by a ^{13}C or ^{15}N dimension is of fundamental importance to resolve all signals; this is underlined by the important efforts which have been invested from the very beginning of protein NMR into developing ^{13}C and ^{15}N isotope-labelling schemes. In solids, this is even more the case, since the poorer resolution makes ^{1}H homonuclear spectroscopy in fully protonated systems demanding. With the information collected on relaxation properties measured in the homonuclear correlation spectra on the selectively labelled Cp samples, we set out to design heteronuclear experiments adapted to the short proton $T_{1\rho}$ and T_2' relaxation times as determined for the CTD. We first evaluated whether 2D CP-based hCH spectra using short CP contact times (250 and 300 μs $\tau_{CP,a}$ and $\tau_{CP,b}$, pulse sequence in Figure S4d) would reveal signals of the CTD. This is not the case (Figure S8), which confirms that the averaging of the dipolar couplings and not the short $T_{1\rho}$ interferes with polarization transfer. We thus turned to use, instead of dipolar couplings, J-couplings by implementing INEPT, which we optimized for the short proton T_2' relaxation times (average 1.8 ms) using numerical simulations (Figure S9) [19,24,63,64]. While solid-state NMR usually employs refocused INEPT (rINEPT), we turned, in order to minimize relaxation losses, to the shorter non-refocused INEPT (referred simply to as INEPT in the following). Delay times of around 1.2 ms should be optimal to observe the CTD residues. To avoid problems with water suppression, we used samples in D_2O.

Figure 4a, b and c show extracts of the INEPT hCH spectra of the capsids (pulse sequence in Figure S4f). New resonances, not present in the NTD spectrum (orange) appear in both full-length capsid spectra around random-coil positions [60] (red circles in Figure 4a and Figure S10) which we assign to the CTD, since their proton chemical shifts coincide with the ones in the EXSY (see Figure S11 for 1D overlay). Spectra of Cp183 and yCp183 are shown in Figure S12.

Intriguingly, an additional Arg spin system is observed in the INEPT spectra of empty (Cp149 and P7-Cp183) capsids compared to RNA-filled ones, which we connected using a TOCSY experiment [65] (Figure S13). We assigned this spin system to R133, which is the only NTD Arg residue not observed in CP experiments [37,66] (Figure S8). The strong upfield shift observed for the R133 resonances could be caused by the ring-current effect [24,67,68] of the W125 it is facing in the last hairpin of the NTD (Figure S14). The absence of R133 from INEPT spectra of the capsids containing RNA indicates a modification in the dynamic regime of this hairpin upon RNA binding. Note that a peak is present at the R133 position in the $H\delta C\delta$ region of the INEPT spectrum of Cp183pgRNA. This δ residue is the most flexible one and its dynamics is less affected by the binding to RNA.
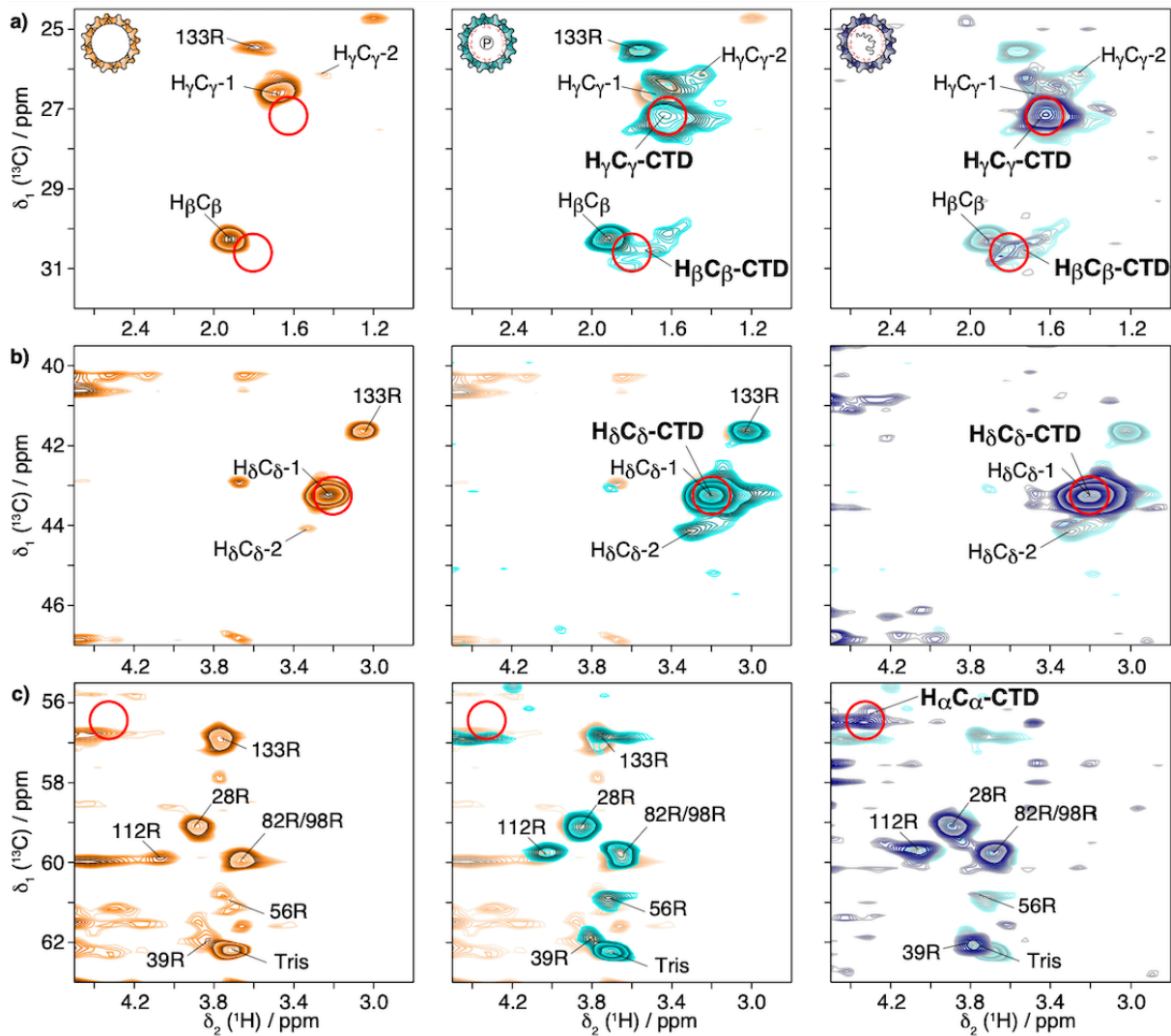

FIGURE 4 2D hCH INEPT-based spectra with $\tau_1 = \tau_2 = 1$ ms of Cp149 (orange), P7-Cp183 (cyan) and Cp183pgRNA (steel blue) in D_2O buffer recorded at 100 kHz and 850 MHz proton frequency. Zooms of the **a)** $H_{\beta2/3}C_\beta / H_{\gamma2/3}C_\gamma$, **b)** $H_{\delta2/3}C_\delta$ and **c)** $H_\alpha C_\alpha$ regions of 2D-hCH INEPT-based spectra. The NTD assignment is transferred from [37]. For sake of simplicity $H_{\gamma2/3}C_\gamma$, $H_{\beta2/3}C_\beta$ and $H_{\delta2/3}C_\delta$ correlations from the CTD are labelled $H_\gamma C_\gamma$, $H_\beta C_\beta$ and $H_\delta C_\delta$. The random-coil chemical-shift positions are shown with red circles [60]. Spectra are overlaid with the spectrum from the previous panel for better comparison.

Optimized INEPT experiments with short mixing times thus successfully display CTD resonances. It confirms the above observation based on the relaxation times that CTDs from various phosphorylated and RNA-filled capsids are characterized by similar motional regimes, determined by interaction of the CTDs with RNA, or amongst themselves. Interestingly, the last hairpin of the NTD shows increased dynamic behavior in empty capsids, while the presence of RNA restricts its mobility.

Since we have shown that spin diffusion/NOE can efficiently be used as homonuclear polarization transfer, we added such a step to the INEPT-based experiment (INEPT-EXSY, Figure S4g) The spectra in Figures 5a, b and c show that connections can be indeed established, including to the Hα frequency in Cp183 capsids, as shown by the additional red dashed lines in Figures 5b and c. These chemical shifts match with the ones in the EXSY spectra (Figure 2b and c) supporting their assignment to the CTD. These spectra allow for connecting the entire Arg spin systems, and would also allow for establishing further connections in more extensively labelled proteins, revealing partially ordered segments if present.
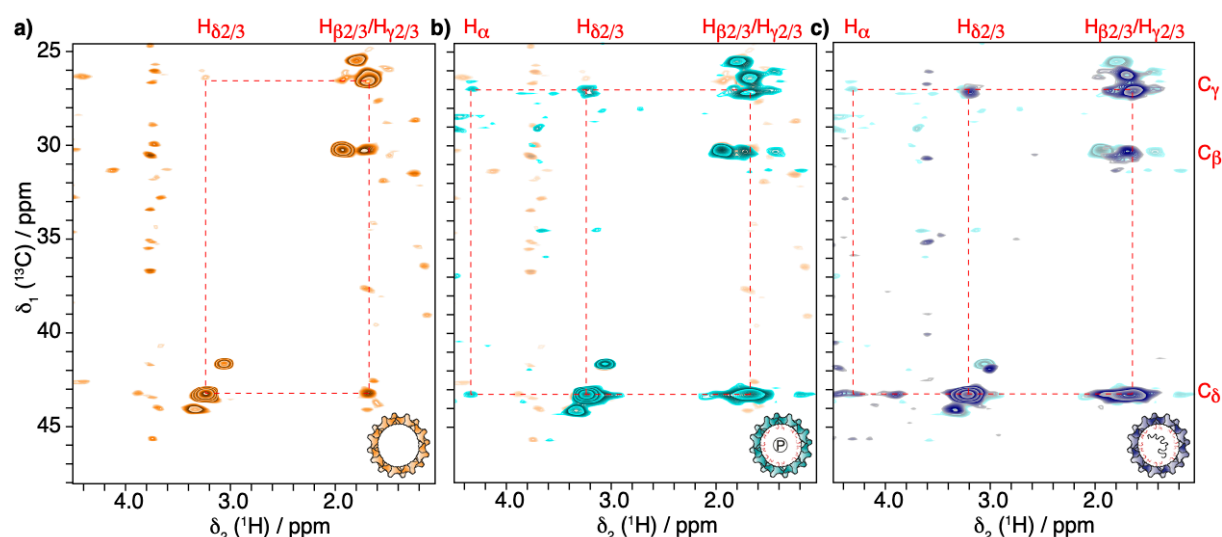

FIGURE 5 INEPT – EXSY NMR experiments. 2D hCH INEPT-based spectra with $\tau_1 = \tau_2 = 1$ ms followed by 50 ms ^{1}H-^{1}H mixing time of **a)** Cp149 **b)** P7-Cp183 and **c)** Cp183-pgRNA in D$_2$O buffer recorded at 100 kHz MAS and 850 MHz proton frequency. The Arg ^{1}H random coil chemical shift positions observed in the EXSY spectra in Figure 2 are shown in red, and correspond to the observed cross peaks here. Spectra are overlaid with the spectrum from the previous panel for better comparison.

Discussion

We have developed proton-detected fast magic-angle-spinning solid-state NMR experiments that allowed us to detect the formerly invisible CTD resonances in HBV capsids. This enabled us to characterize its structural and dynamic behavior and investigate dynamical differences upon CTD phosphorylation as well as the presence of different types of RNA. We showed that, despite its random coil conformation, the CTD does not behave like an intrinsically disordered domain, but shows hindered motion, including when it is phosphorylated. This may have implications for capsid function, as the localization of the CTD determines the interactions in which it can participate during maturation, viral particle assembly and nuclear import. CTD exposure is particularly important in the context of nuclear import, and has been previously been studied by trypsin digestion, which showed that the CTD of phosphorylated capsids can be partially digested at additional sites when compared to unphosphorylated ones [48]. This reaction takes place outside the capsid and requests the CTD to be exposed to the outside. In this context, a loop-out mechanism, where part of the CTD (not including the free end) escapes from the capsid, as well as an extrusion model where the entire CTD is moving outside the capsid, are currently discussed [48]. Our data clearly supports the loop-out mechanism, where only the N-terminal end of the flexible CTD is exposed to the outside of the capsid, since the extrusion model, where the entire CTD moves outside the capsid, would ask for higher motional freedom, and thus faster dynamics over what was observed. The higher extent of exposure in phosphorylated capsids is likely due to the more flexible Pro rich loop (residues 129 to 138), as shown by the increased flexibility of R133. This points to a decrease in flexibility of this proline-rich hairpin on heterologous interactions, i.e. RNA binding, and an increase in absence thereof. In any case, our data clearly show that looping out is not a highly dynamic process, since the CTDs in the phosphorylated state do not move freely like an IDP. The CTDs motion thus must be restricted in the phosphorylated capsid by homologous interactions, which can be intra and/or intermolecular. Intramolecular interactions could correspond to folding back of the arginine side chains onto the neighboring phosphorylated Ser- and Thr-bound PO_4 moieties of the same chain. This kind of turns would however hardly be compatible with the random-coil chemical shifts observed. We thus conclude that these interactions are more likely intermolecular, and form a network of Arg-P-Ser/P-Thr interactions at the inner capsid lining, and/or its lumen as shown in Figure 6, where schematic drawings are displayed of the electrostatic interactions between the CTD and the RNA (Figure 6a), and between the CTDs (Figure 6b). Thus, neither RNA-filled capsids nor phosphorylated capsids feature highly dynamic CTD extrusion. This picture might be different in mature, DNA-filled capsids, which are preferentially transported to the nucleus via CTD interactions with cellular

importins. Considering that nuclear import of empty immature (pgRNA-containing) capsids would not present any gain for virus replication, it can be hypothesized that these interactions are functional, and thus different in mature capsids, where the CTD might show higher mobility. Finally, our findings might also have implications in the context of liquid-liquid phase separation (LLPS). Given the disordered nature of the CTD and the interactions in which it participates, the lumen of the capsid is reminiscent of biomolecular assemblies that occur during LLPS, where often Coulomb-type interactions are formed between a disordered protein domain and RNA or phosphates [69,70]. The weak, transient, and multivalent interactions the capsid's CTD arginine-rich low-complexity stretches engage in, and which can be modulated by phosphorylation [71], are indeed a hallmark of LLPS condensates. The dynamic parameters which we measured here for the HBV capsid CTD might thus well apply in a more general manner to these kinds of condensates, and the derived methods might allow detection of LLPS condensate components in solid-state NMR spectra.
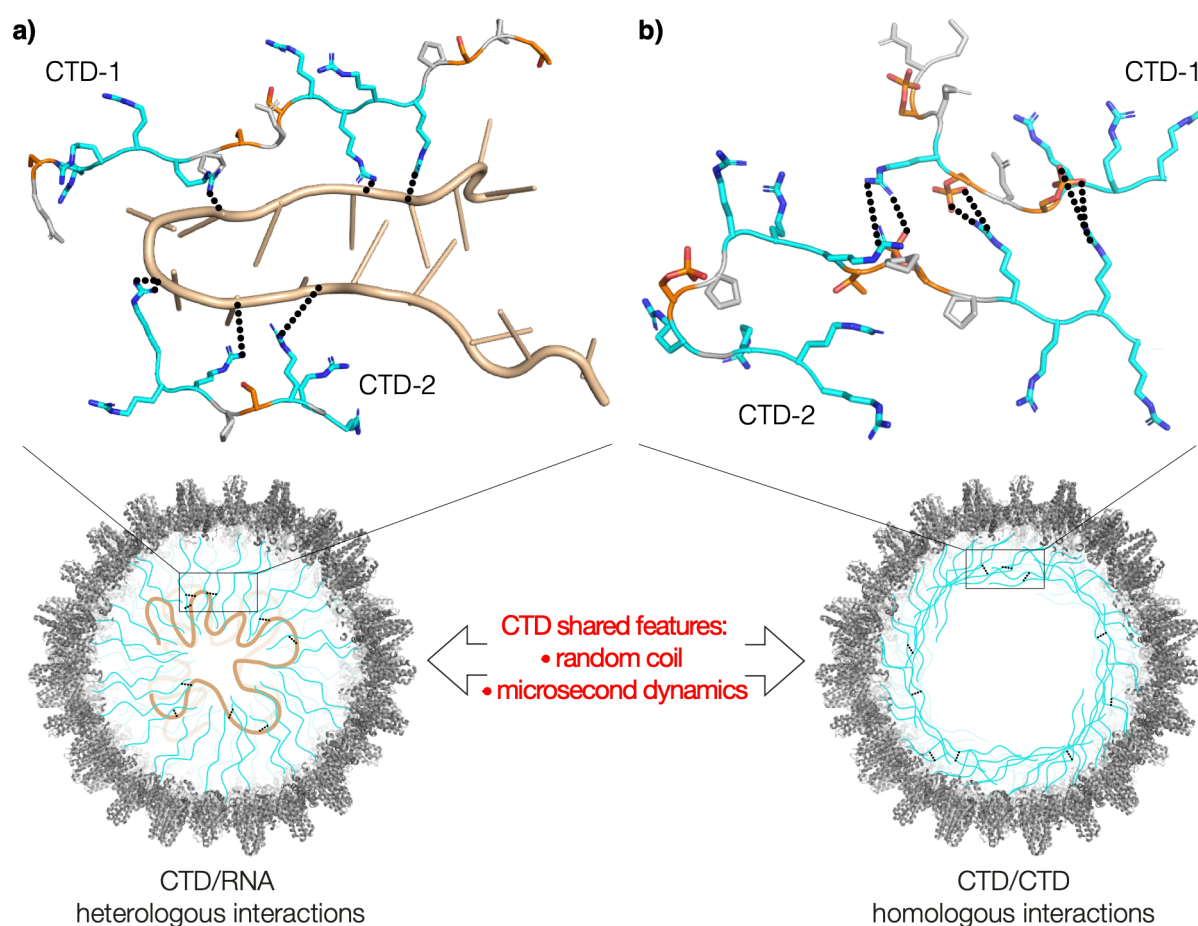

FIGURE 6 Schematic of the heterologous and homologous interactions where the CTD is involved in **a)** RNA-containing and **b)** phosphorylated capsids respectively. The arginine residues are colored in cyan (N in blue), the serine (phosphorylated in b) and threonine in orange (O in red). The RNA backbone is colored in pale brown. The intermolecular interactions are shown as dotted lines. The capsid shell is represented in grey.

Conclusion

We report for the first time the observation of the functional CTD of the HBV capsid protein, through the combined use of selective labelling schemes and 1D ^{1}H-MAS NMR experiments at fast spinning frequencies. The detection of the CTD in such an experiment is an immediate indication that dynamics do not broaden the lines beyond detection although indeed additional line broadening is observed. $T_{1\rho}$ measurements allow to estimate the correlation time of the motion as approximately 1 μs with an order parameter of $S < 0.9$.

The CTD resonance frequencies revealed that the arginine residues, and most probably all CTD residues, adopt a random coil conformation, and thus do not form specific and defined secondary structures in the context of RNA binding, nor upon phosphorylation. Importantly, we found, from comparing T_2' and $T_{1\rho}$ relaxation times, that CTD dynamic time scales are very similar in phosphorylated empty capsids and RNA-filled unphosphorylated ones.

Besides characterizing for the first time the structure and dynamics of the CTD of the HBV capsid, we present an experimental approach combining fast MAS with minimum-length magnetization-transfer periods which points to a way out of the frustrating situation where crucial protein domains in structural biology remain invisible. Since the physical origins of the evasiveness of these domains in general lies in dynamic behavior, and more precisely microsecond timescale motions, we would expect that the methodology presented herein, broadening the dynamic time scales accessible by NMR, is of use for the characterization of flexible domains in a wider range of proteins, including membrane proteins, fibrils, subviral particles and other nucleoprotein assemblies.

Acknowledgement

References

[1] A. K. Mittermaier, L. E. Kay, *Trends in Biochemical Sciences* **2009**, *34*, 601–611.

[2] I. Matlahov, P. C. A. van der Wel, *Methods* **2018**, *148*, 123–135.

[3] A. Siemer, A. Arnold, C. Ritter, T. Westfeld, M. Ernst, R. Riek, B. H. Meier, *Journal Of The American Chemical Society* **2006**, *128*, 13224–13228.

[4] A. B. Siemer, *Solid State Nuclear Magnetic Resonance* **2020**, *106*, 101643.

[5] S. F. Cousin, P. Kadeřávek, B. Haddou, C. Charlier, T. Marquardsen, J.-M. Tyburn, P.-A. Bovier, F. Engelke, W. Maas, G. Bodenhausen, P. Pelupessy, F. Ferrage, *Angewandte Chemie International Edition* **2016**, *55*, 9886–9889.

[6] I. C. Felli, B. Brutscher, *ChemPhysChem* **2009**, *10*, 1356–1368.

[7] U. Haeberlen, J. S. Waugh, *Physical Review* **1969**, *185*, 420–429.

[8] A. Jasinski, *Acta Physica Polonica* **1979**, *55*, 295–301.

[9] A. Krushelnitsky, T. Zinkevich, B. Reif, K. Saalwächter, *Journal of Magnetic Resonance* **2014**, *248*, 8–12.

[10] R. Kurbanov, T. Zinkevich, A. Krushelnitsky, *The Journal of Chemical Physics* **2011**, *135*, 184104.

[11] S. Penzel, A. Oss, M. L. Org, A. Samoson, A. Böckmann, M. Ernst, B. H. Meier, *Journal of Biomolecular NMR* **2019**, *73*, 19–29.

[12] I. Matlahov, P. C. A. van der Wel, *Methods* **2018**, *148*, 123–135.

[13] P. Rovó, C. A. Smith, D. Gauto, B. L. de Groot, P. Schanda, R. Linser, *Journal of the American Chemical Society* **2019**, *141*, 858–869.

[14] S. R. Hartmann, E. L. Hahn, *Physical Review* **1962**, *128*, 2042–2053.

[15] S. Hediger, B. H. Meier, R. R. Ernst, *Chemical Physics Letters* **1995**, *240*, 449–456.

[16] A. Pines, M. G. Gibby, J. S. Waugh, *The Journal of Chemical Physics* **1973**, *59*, 569–590.

[17] E. O. Stejskal, J. Schaefer, J. S. Waugh, *Journal of Magnetic Resonance (1969)* **1977**, *28*, 105–112.

[18] G. Bodenhausen, D. J. Ruben, *Chemical Physics Letters* **1980**, *69*, 185–189.

[19] C. Coelho, T. Azaı, G. Laurent, C. Bonhomme, P. M. Curie-paris, L. Chimie, D. Matie, T. Azaïs, L. Bonhomme-Coury, G. Laurent, C. Bonhomme, *Inorganic Chemistry* **2007**, *46*, 1379–1387.

[20] B. Elena, A. Lesage, S. Steuernagel, A. Böckmann, L. Emsley, *Journal of the American Chemical Society* **2005**, *127*, 17296–17302.

[21] T. Gopinath, S. E. D. Nelson, G. Veglia, *Journal of Magnetic Resonance* **2017**, *285*, 101–107.

[22] R. Linser, U. Fink, B. Reif, *Journal of Magnetic Resonance* **2008**, *193*, 89–93.

[23] S. Penzel, A. A. Smith, V. Agarwal, A. Hunkeler, M.-L. L. Org, A. Samoson, A. Böckmann, M. Ernst, B. H. Meier, *Journal of Biomolecular NMR* **2015**, *63*, 165–186.

[24] T. Wiegand, A. A. Malär, R. Cadalbert, M. Ernst, A. Böckmann, B. H. Meier, *Frontiers in Molecular Biosciences* **2020**, *7*, 1–11.

[25] J. Jeener, B. H. Meier, P. Bachmann, R. R. Ernst, *The Journal of Chemical Physics* **1979**, *71*, 4546–4553.

[26] V. Agarwal, *Journal of Magnetic Resonance* **2020**, *311*, 106661.

[27] A. Böckmann, M. Ernst, B. H. Meier, *Journal of Magnetic Resonance* **2015**, *253*, 71–79.

[28] M. Chávez, T. Wiegand, A. A. Malär, B. H. Meier, M. Ernst, *Magnetic Resonance* **2021**, *2*, 499–509.

[29] D. Suter, R. R. Ernst, *Physical Review B* **1982**, *25*, 6038–6041.

[30] M. Veshtort, R. G. Griffin, *The Journal of Chemical Physics* **2011**, *135*, 134509.

[31] M. Niklasch, P. Zimmermann, M. Nassal, *Biomedicines* **2021**, *9*, 1577.

[32] R. A. Crowther, N. A. Kiselev, B. Böttcher, J. A. Berriman, G. P. Borisova, V. Ose, P. Pumpens, *Cell* **1994**, *77*, 943–950.

[33] M. Nassal, *Virus Research* **2008**, *134*, 235–249.

[34] B. Böttcher, S. A. Wynne, R. A. Crowther, *Nature* **1997**, *386*, 88–91.

[35] A. Zlotnick, N. Cheng, S. J. Stahl, J. F. Conway, A. C. Steven, P. T. Wingfield, *Proceedings of the National Academy of Sciences* **1997**, *94*, 9556–9561.

[36] S. A. Wynne, R. A. Crowther, A. G. W. Leslie, *Molecular Cell* **1999**, *3*, 771–780.

[37] L. Lecoq, M. Schledorn, S. Wang, S. Smith-Penzel, A. A. Malär, M. Callon, M. Nassal, B. H. Meier, A. Böckmann, *Frontiers in Molecular Biosciences* **2019**, *6*, 1–10.

[38] L. Lecoq, S. Wang, T. Wiegand, S. Bressanelli, M. Nassal, B. H. Meier, A. Böckmann, *ChemPhysChem* **2018**, *19*, 1336–1340.

[39] J. Kim, J. Wu, *Biophysical Journal* **2014**, *107*, 1453–1461.

[40] W. H. Hudson, E. A. Ortlund, *Nature Reviews Molecular Cell Biology* **2014**, *15*, 749–760.

[41] H. de Rocquigny, V. Rat, F. Pastor, J. L. Darlix, C. Hourioux, P. Roingeard, *Viruses* **2020**, *12*, E738.

[42] R. C. Oliver, W. Potrzebowski, S. M. Najibi, M. N. Pedersen, L. Arleth, N. Mahmoudi, I. André, *ACS Nano* **2020**, *14*, 10226–10238.

[43] C. Seeger, W. S. Mason, *Virology* **2015**, *479–480*, 672–686.

[44] C. Chen, J. C.-Y. Wang, E. E. Pierson, D. Z. Keifer, M. Delaleau, L. Gallucci, C. Cazenave, M. Kann, M. F. Jarrold, A. Zlotnick, *PLOS Pathogens* **2016**, *12*, e1005802.

[45] M. Kann, B. Sodeik, A. Vlachou, W. H. Gerlich, A. Helenius, *Journal of Cell Biology* **1999**, *145*, 45–55.

[46] M. Kann, W. H. Gerlich, *Journal of Virology* **1994**, *68*, 7993–8000.

[47] B. Rabe, A. Vlachou, N. Pante, A. Helenius, M. Kann, *Proceedings of the National Academy of Sciences* **2003**, *100*, 9849–9854.

[48] J. Heger-Stevic, P. Zimmermann, L. Lecoq, B. Böttcher, M. Nassal, *PLOS Pathogens* **2018**, *14*, e1007488.

[49] L. Lecoq, S. Wang, M. Dujardin, P. Zimmermann, L. Schuster, M.-L. Fogeron, M. Briday, M. Schledorn, T. Wiegand, L. Cole, R. Montserret, S. Bressanelli, B. H. Meier, M. Nassal, A. Böckmann, *Proceedings of the National Academy of Sciences* **2021**, *118*, e2022464118.

[50] S. Wang, M. L. Fogeron, M. Schledorn, M. Dujardin, S. Penzel, D. Burdette, J. M. Berke, M. Nassal, L. Lecoq, B. H. Meier, A. Böckmann, *Frontiers in Molecular Biosciences* **2019**, *6*, 1–11.

[51] T. Bauer, C. Dotta, L. Balacescu, J. Gath, A. Hunkeler, A. Böckmann, B. H. Meier, *Journal of Biomolecular NMR* **2017**, *67*, 51–61.

[52] R. Gupta, H. Zhang, M. Lu, G. Hou, M. Caporini, M. Rosay, W. Maas, J. Struppe, J. Ahn, I. J. L. Byeon, H. Oschkinat, K. Jaudzems, E. Barbet-Massin, L. Emsley, G. Pintacuda, A. Lesage, A. M. Gronenborn, T. Polenova, *Journal of Physical Chemistry B* **2019**, *123*, 5048–5058.

[53] D. Lacabanne, B. H. Meier, A. Böckmann, *Journal of Biomolecular NMR* **2018**, *71*, 141–150.

[54] B. S. de Almeida, P. Moutzouri, G. Stevanato, L. Emsley, *The Journal of Chemical Physics* **2021**, *155*, 084201.

[55] A. A. Malär, S. Smith-Penzel, G.-M. Camenisch, T. Wiegand, A. Samoson, A. Böckmann, M. Ernst, B. H. Meier, *Physical Chemistry Chemical Physics* **2019**, *21*, 18850–18865.

[56] D. C.-D. Paepe, J. Stanek, K. Jaudzems, K. Tars, L. B. Andreas, G. Pintacuda, *Solid State Nuclear Magnetic Resonance* **2017**, *87*, 1–19.

[57] M. Schledorn, A. A. Malär, A. Torosyan, S. Penzel, D. Klose, A. Oss, M. L. Org, S. Wang, L. Lecoq, R. Cadalbert, A. Samoson, A. Böckmann, B. H. Meier, *ChemBioChem* **2020**, *21*, 2540–2548.

[58] U. Sternberg, R. Witter, I. Kuprov, J. M. Lamley, A. Oss, J. R. Lewandowski, A. Samoson, *Journal of Magnetic Resonance* **2018**, *291*, 32–39.

[59] K. Xue, R. S. ORCID: 0000-0001-9055-7897, C. Motz, S. Asami, V. Decker, S. Wegner, Z. Tošner, B. R. ORCID: 0000-0001-7368-7198, *The Journal of Physical Chemistry C* **2018**, *122*, 1–6.

[60] D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges, B. D. Sykes, *Journal of Biomolecular NMR* **1995**, *5*, 67–81.

[61] P. Rovó, R. Linser, *The journal of physical chemistry B* **2017**, *121*, 6117–6130.

[62] M. R. Jensen, P. R. L. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, P. Bernadó, M. Blackledge, *Structure* **2009**, *17*, 1169–1185.

[63] D. T. Pegg, D. M. Doddrell, W. M. Brooks, M. Robin Bendall, *Journal of Magnetic Resonance (1969)* **1981**, *44*, 32–40.

[64] O. W. Sørensen, R. R. Ernst, *Journal of Magnetic Resonance (1969)* **1983**, *51*, 477–489.

[65] M. Baldus, B. H. Meier, *Journal of Magnetic Resonance, Series A* **1996**, *121*, 65–69.

[66] L. Lecoq, S. Wang, T. Wiegand, S. Bressanelli, M. Nassal, B. H. Meier, A. Böckmann, *Biomolecular NMR Assignments* **2018**, *12*, 205–214.

[67] E. Carignani, S. Borsacchi, J. P. Bradley, S. P. Brown, M. Geppi, *The Journal of Physical Chemistry C* **2013**, *117*, 17731–17740.

[68] S. J. Perkins, K. Wüthrich, *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1979**, *576*, 409–423.

[69] A. C. Murthy, N. L. Fawzi, *Journal of Biological Chemistry* **2020**, *295*, 2375–2384.

[70] A. Savastano, A. Ibáñez de Opakua, M. Rankovic, M. Zweckstetter, *Nat Commun* **2020**, *11*, 6041.

[71] M. Niklasch, P. Zimmermann, M. Nassal, *Biomedicines* **2021**, *9*, 1577.