

# SUPPLEMENTARY INFORMATION FOR:

## Experimental Characterization of the Hepatitis B Virus Capsid Dynamics by Solid-State NMR

Alexander A. Malär<sup>1#</sup>, Morgane Callon<sup>1#</sup>, Albert A. Smith<sup>2</sup>, Shishan Wang<sup>3</sup>, Lauriane Lecoq<sup>3</sup>, Carolina Pérez-Segura<sup>4</sup>, Jodi A. Hadden-Perilla<sup>4</sup>, Anja Böckmann<sup>3\*</sup> and Beat H. Meier<sup>1\*</sup>

<sup>1</sup>Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland

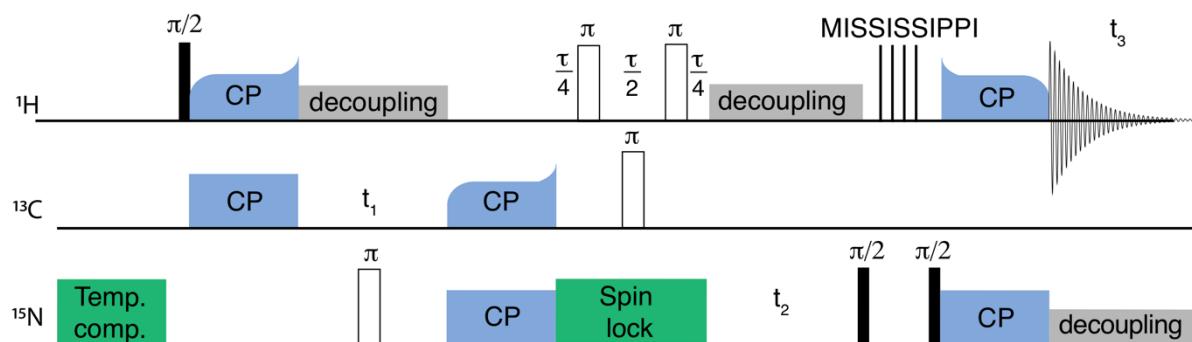
<sup>2</sup>Departement of Medical Physics and Biophysics, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany

<sup>3</sup>Molecular Microbiology and Structural Biochemistry (MMSB) UMR 5086 CNRS/Université de Lyon, Labex Ecofect, 7 passage du Vercors, 69367 Lyon, France

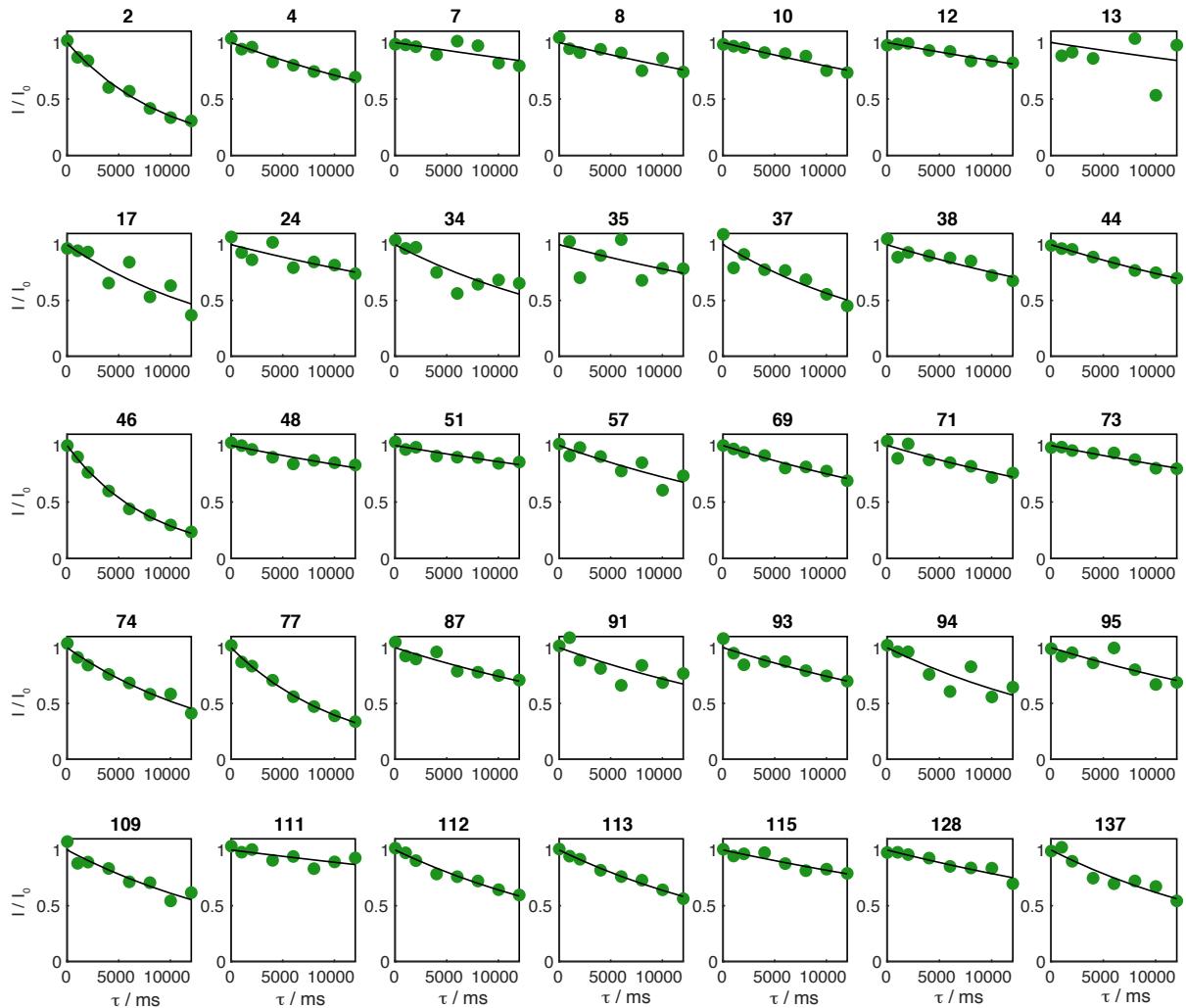
<sup>4</sup>Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, United States

# These authors contributed equally

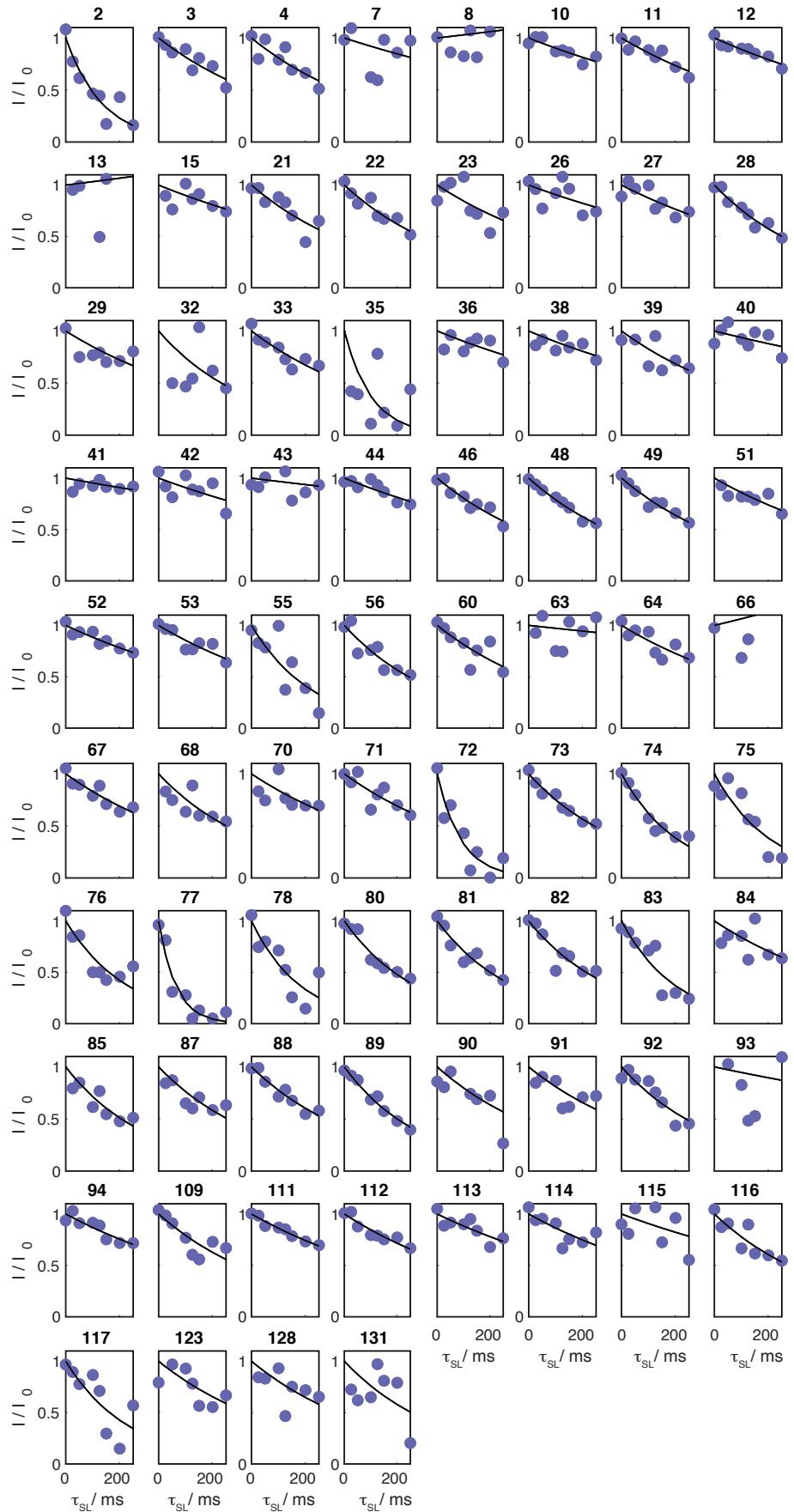
\*Corresponding Authors: Anja Böckmann, [a.bockmann@ibcp.fr](mailto:a.bockmann@ibcp.fr); Beat H. Meier, [beme@ethz.ch](mailto:beme@ethz.ch)



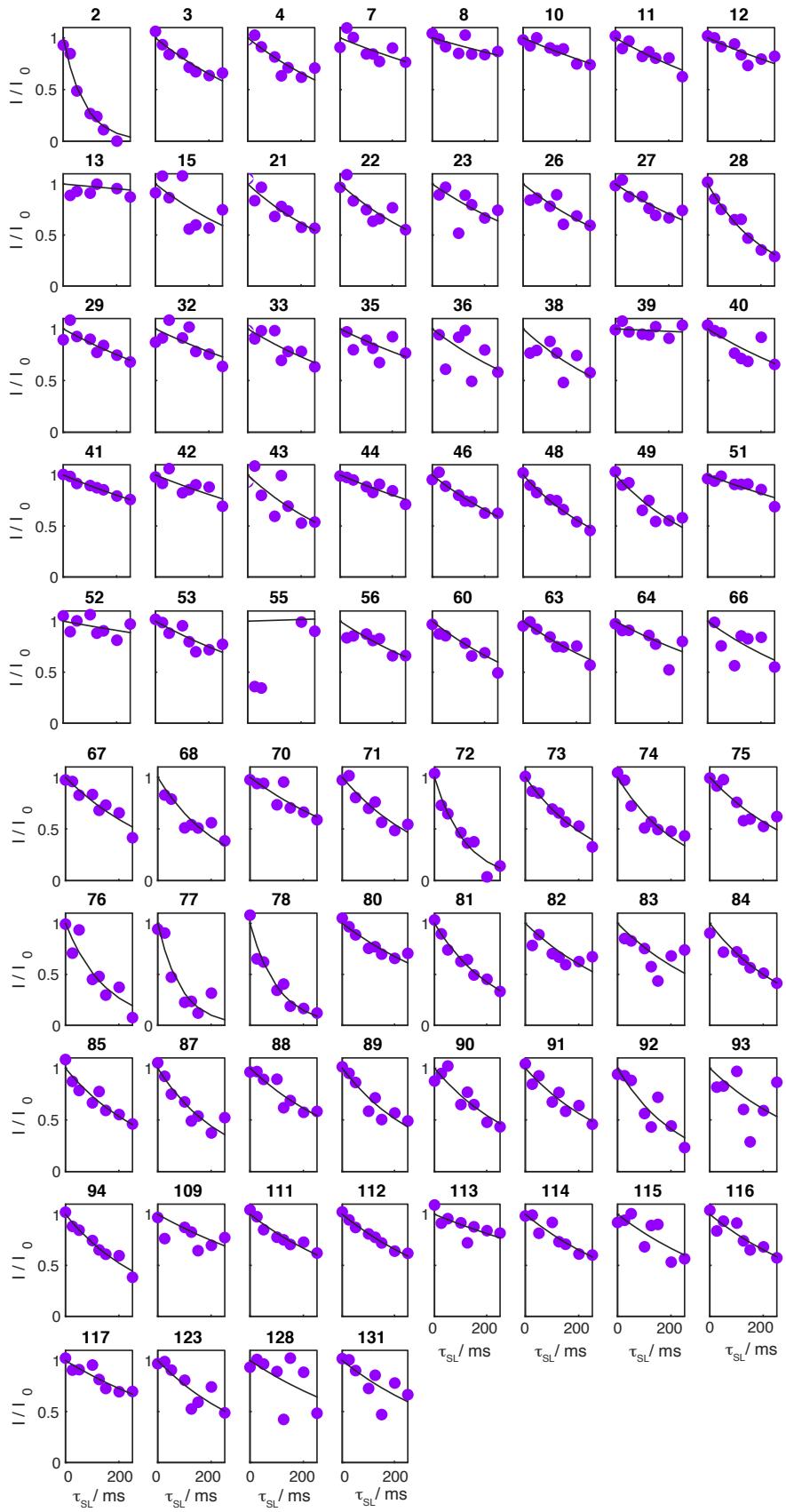
**Fig. S1:** hCANH-based pulse sequence for  $R_{1\rho}({}^{15}\text{N})$  measurements in three-dimensional spectra. Due to the long spin-lock times required for  ${}^{15}\text{N}$  relaxation measurements, a temperature compensation block is placed at the beginning of the sequence on the  ${}^{15}\text{N}$  channel, to ensure that for each spin-lock duration, the same total power is deposited in the sample. The  $\pi$  pulses on  ${}^1\text{H}$  applied at  $\frac{1}{4}$  and  $\frac{3}{4}$  of the  ${}^{15}\text{N}$  spin-lock duration time are used to refocus a possible effect of  ${}^{15}\text{N}$  CSA / H-N dipolar cross-correlated relaxation.



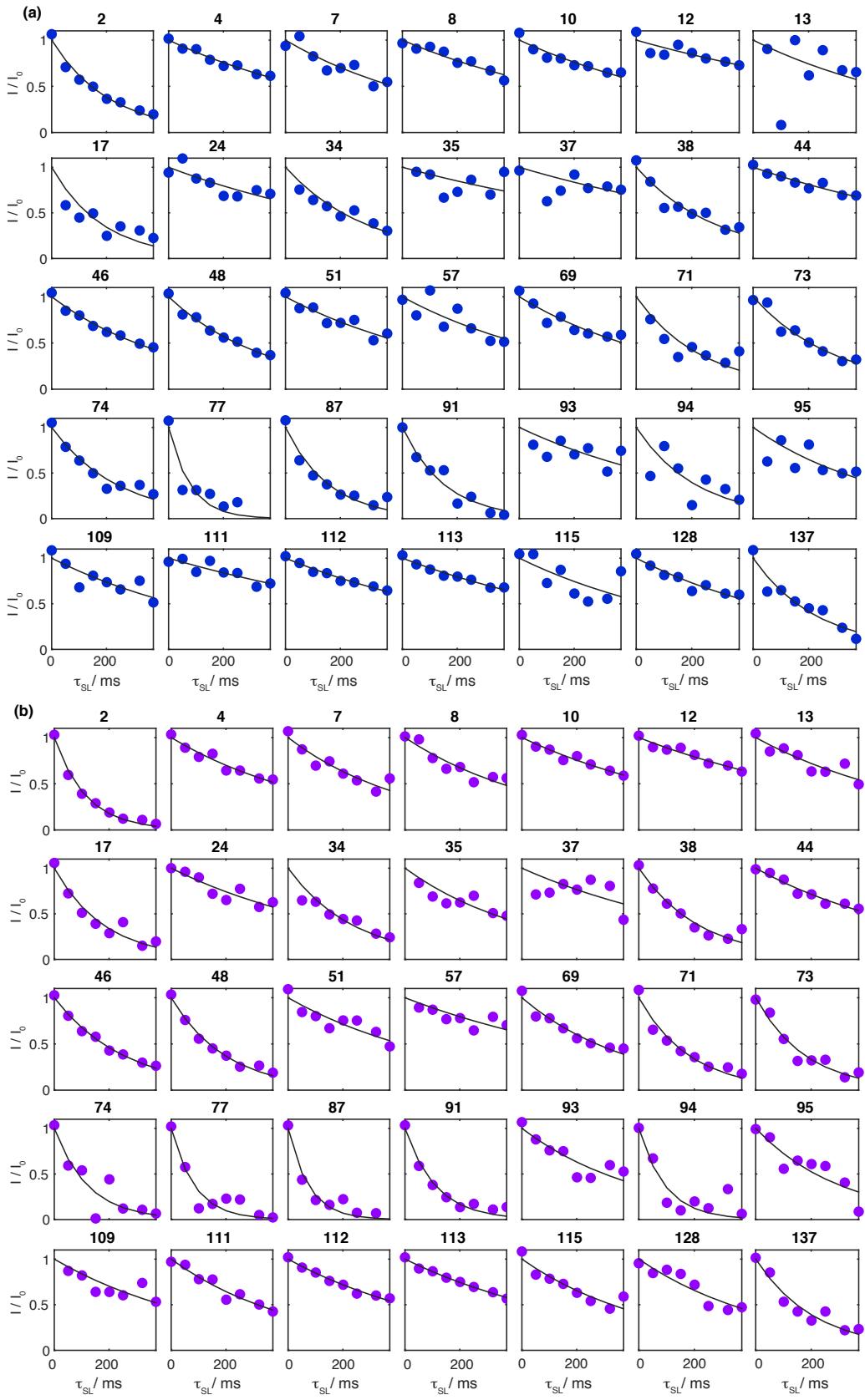
**Fig. S2:** Site-specific  $T_1(^{15}\text{N})$  relaxation traces, extracted from 2D hNH spectra recorded at 20.0 T external magnetic field, 110 kHz MAS (0.7mm rotor) with corresponding mono-exponential fits.



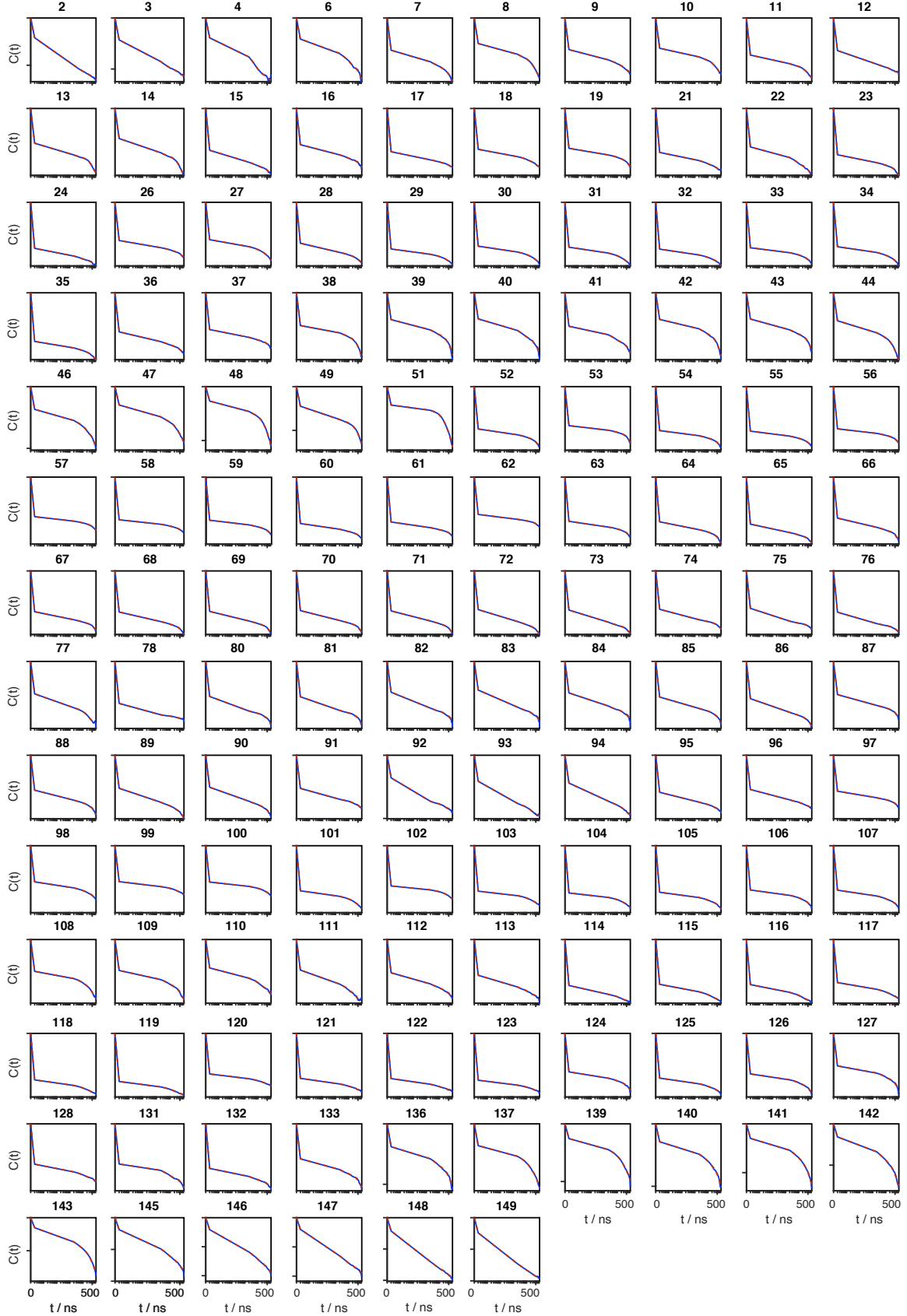
**Fig. S3:** Site-specific  $T_{1\rho}(^{15}\text{N})$  relaxation traces, extracted from 3D hCANH spectra recorded at 20.0 T external magnetic field, 110 kHz MAS (0.7mm rotor) with corresponding mono-exponential fits.



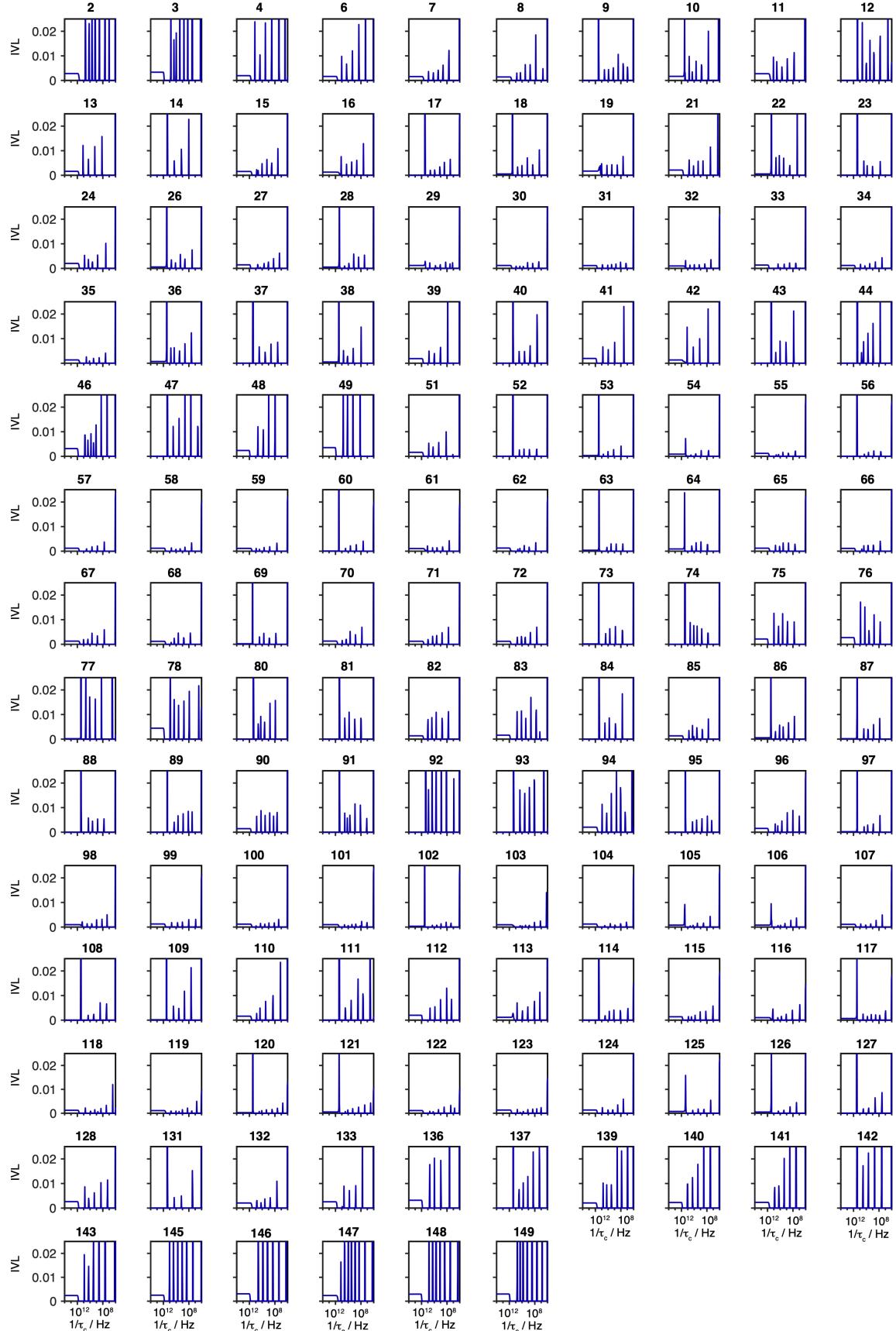
**Fig. S4:** Site-specific  $T_{1\rho}^{(15\text{N})}$  relaxation traces, extracted from 3D hCANH spectra recorded at 20.0 T external magnetic field, 80 kHz MAS (0.7mm rotor) with corresponding mono-exponential fits.



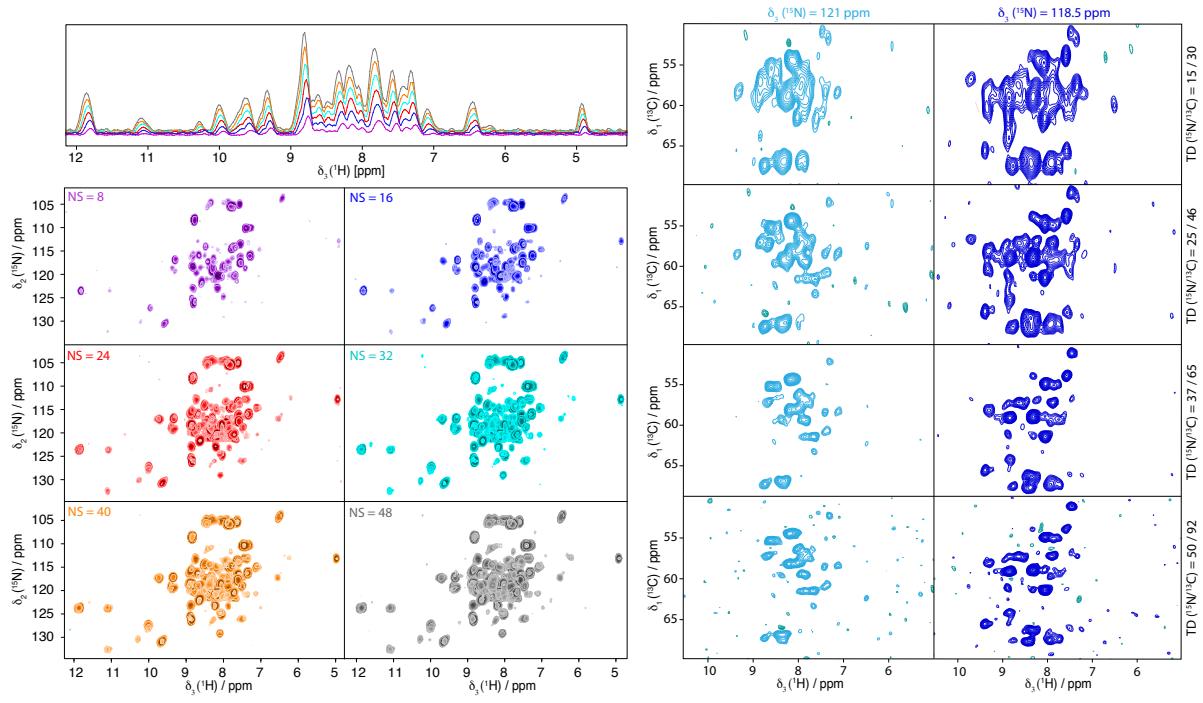
**Fig. S5:** Site-specific  $T_{1\rho}(^{15}\text{N})$  relaxation traces, extracted from 2D hNH spectra recorded at 20.0 T external magnetic field and (a) 160 kHz MAS, (b) 80 kHz MAS (0.5mm rotor) with corresponding mono-exponential fits.



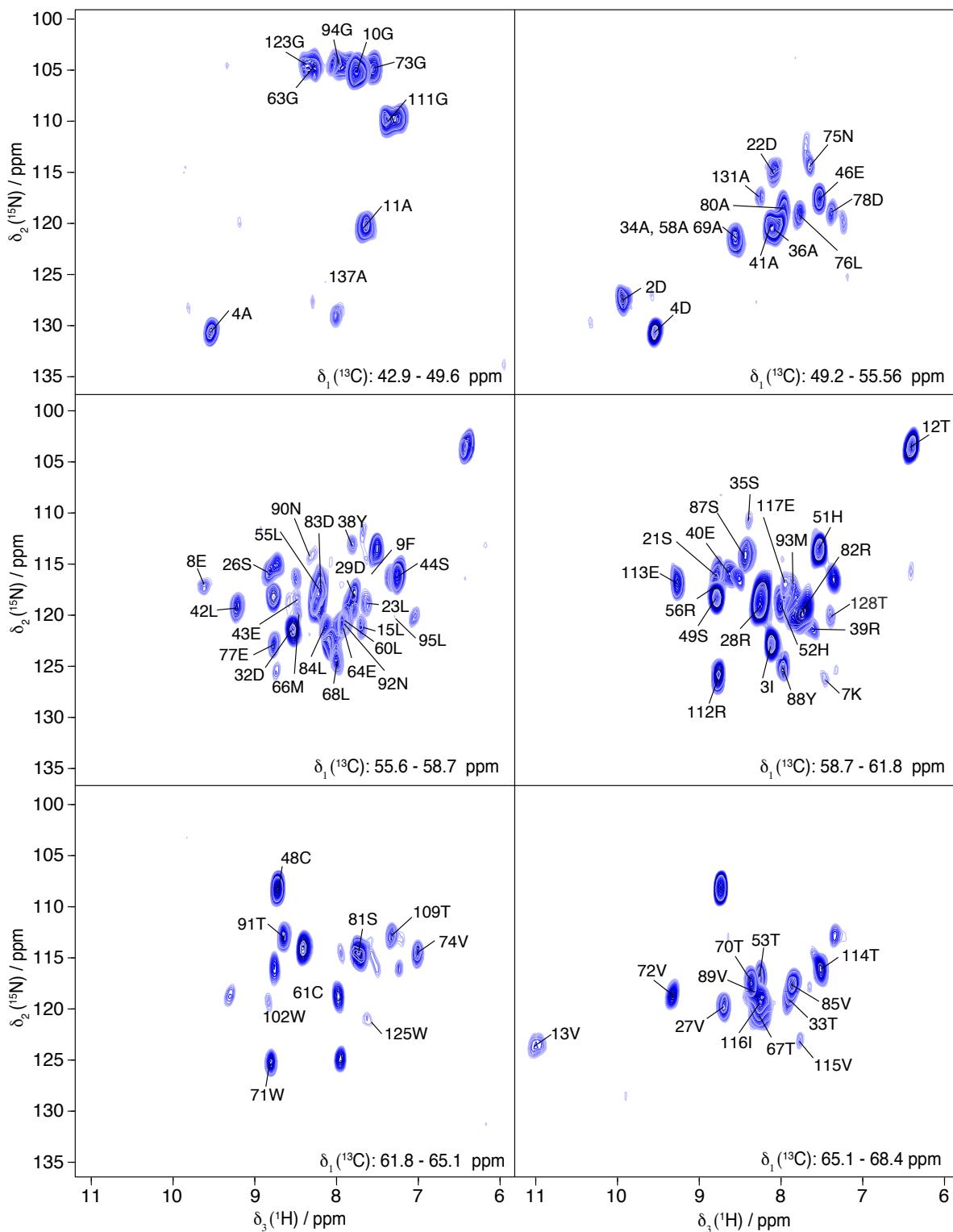
**Fig. S6:** Residue-specific correlation functions  $C(t)$  extracted from the MD trajectory for chain A (blue) and best fit, obtained solving the inverse Laplace transform problem with  $\lambda = 0$  (red).



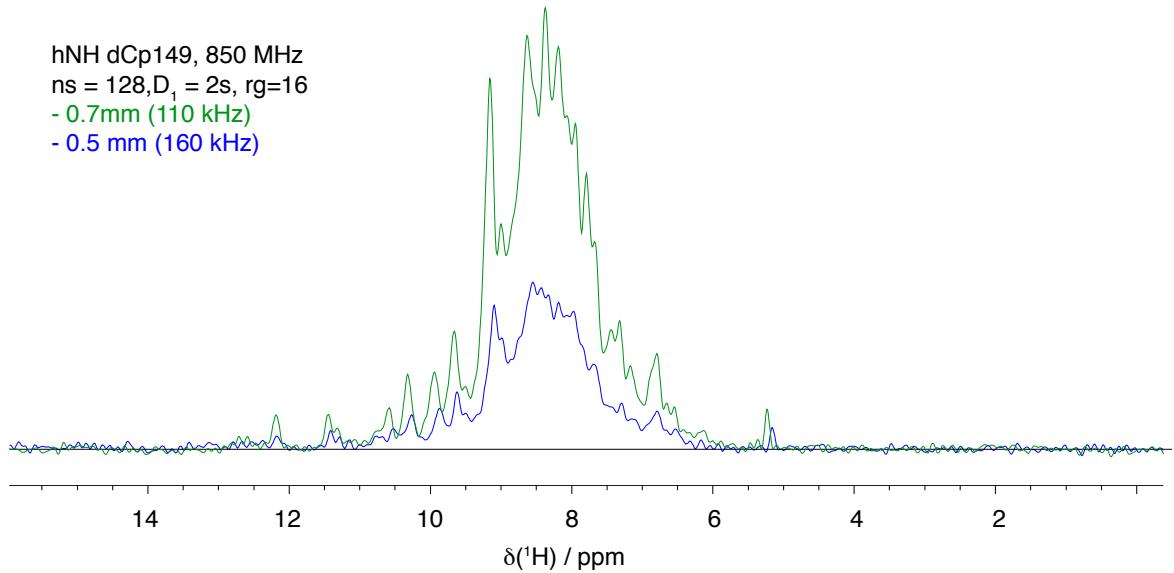
**Fig. S7:** Residue-specific inverse Laplace transforms (IVL) obtained from the correlation functions of the MD trajectory for chain A and a regularization weight of  $\lambda = 0$ . The IVL was performed with  $\tau_c = [10^{-14}, 10^{-6}] \text{ s}$ .



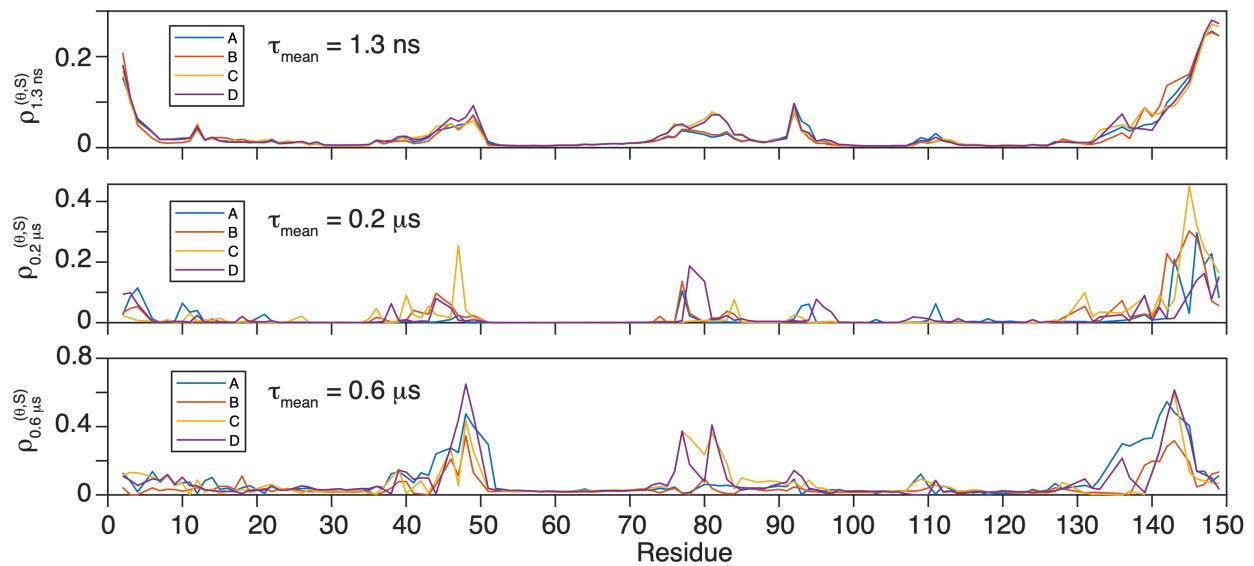
**Fig. S8:** Illustration for a comparison of number of scans and spectral resolution compromises in 3D hCANH spectra of Cp149 recorded at 110 kHz MAS and 20.0 T external magnetic field, used to reduce the experimental time of one 3D experiment to a maximum of one day of measurement time. In the final experiment used for the  $T_{1\rho}(^{15}\text{N})$  measurements, 24 scans and 30/50 points in the indirect dimensions were used.



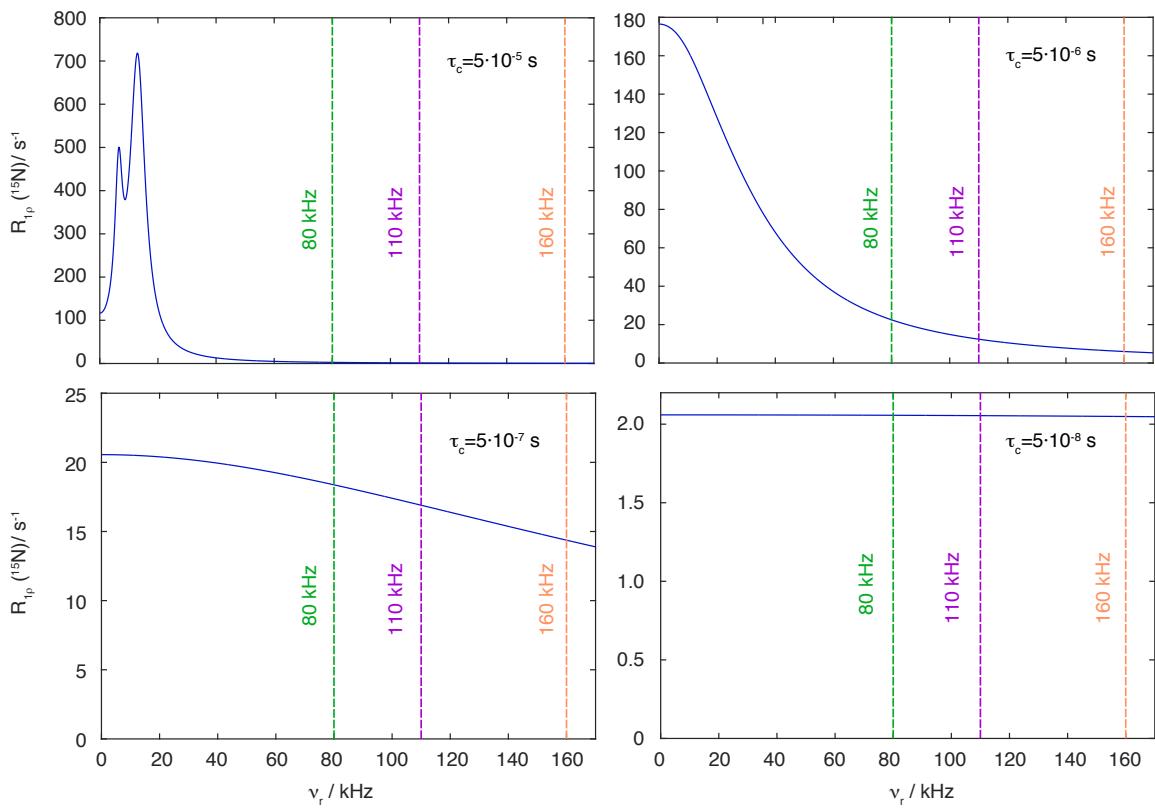
**Fig. S9:** HN slices for different regions of  $^{13}\text{C}$  chemical shifts of the 3D hCANH experiment recorded at 110 kHz MAS and 20.0 T external magnetic fields, with 24 scans and 30/50 points in the indirect dimension.



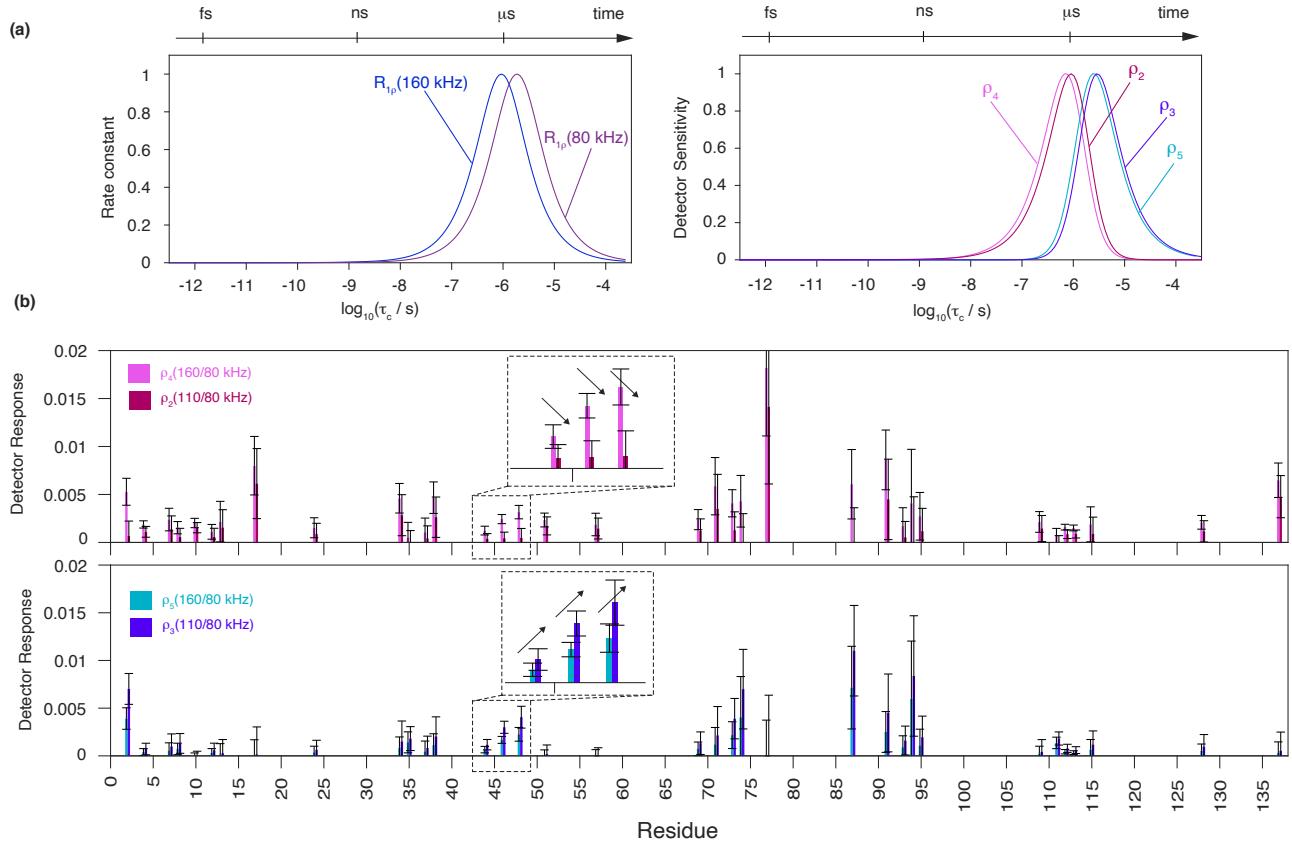
**Fig. S10:** Overlay of first free-induction decay (FID) of hNH experiments recorded at 110 kHz in an 0.7mm rotor (green) and 160 kHz in an 0.5 mm rotor (blue) at 20.0 T external magnetic field, with 128 scans, a receiver gain (RG) of 16 and 2s of inter-scan delay.



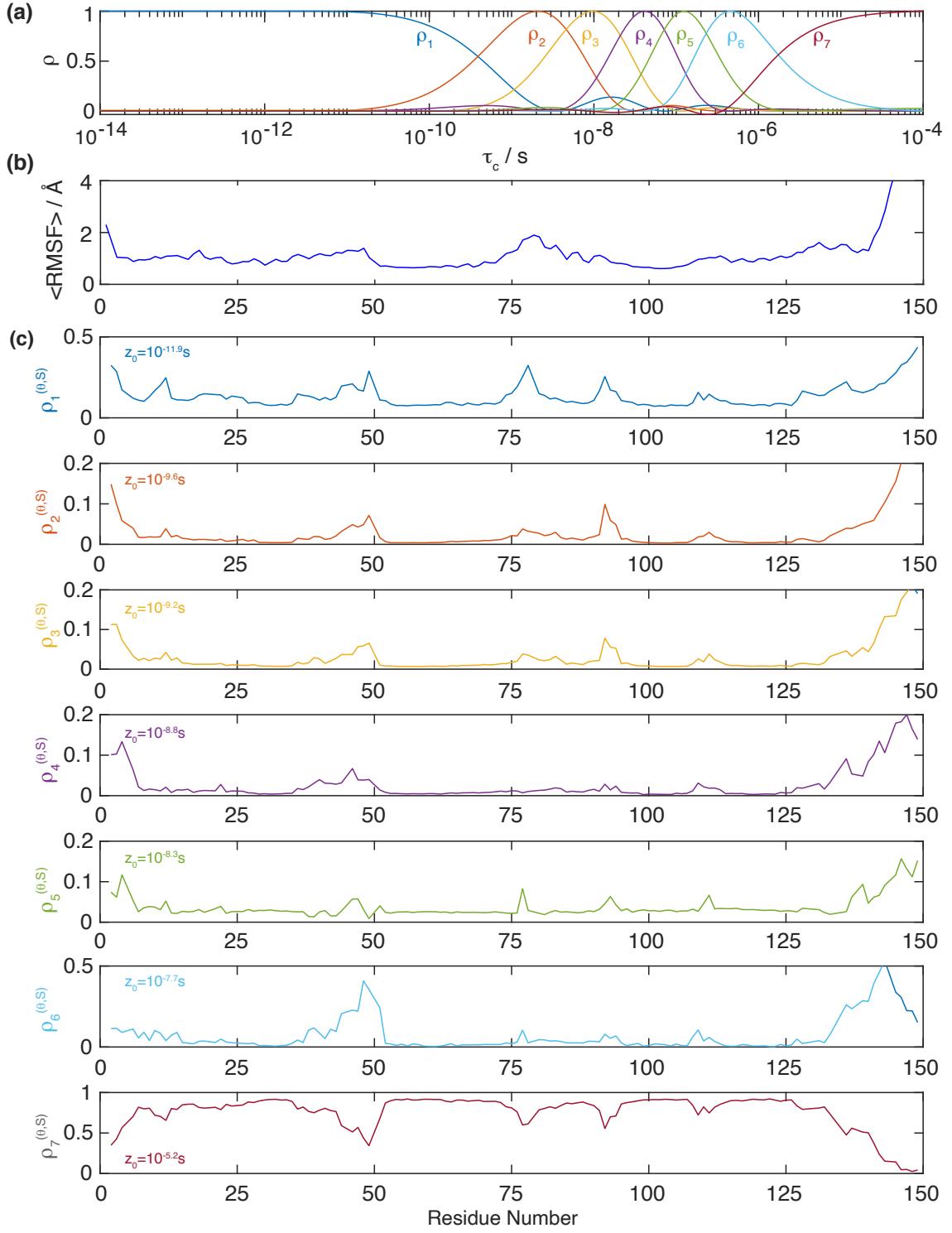
**Fig. S11:** Molecular dynamic detector responses of the four capsid subunits A,B,C,D for the  $\rho_1$ ,  $\rho_2$ ,  $\rho_3$  detectors shown in Fig.3b, constructed from the rotating-frame relaxation-rate constants at 13 kHz spin-lock strength and 110 kHz and 80 kHz MAS and the longitudinal relaxation rate constant at 850 MHz. MD detector optimization has been carried out over the correlation time interval  $\tau_c = [10^{-14}, 10^{-6}] s$ .



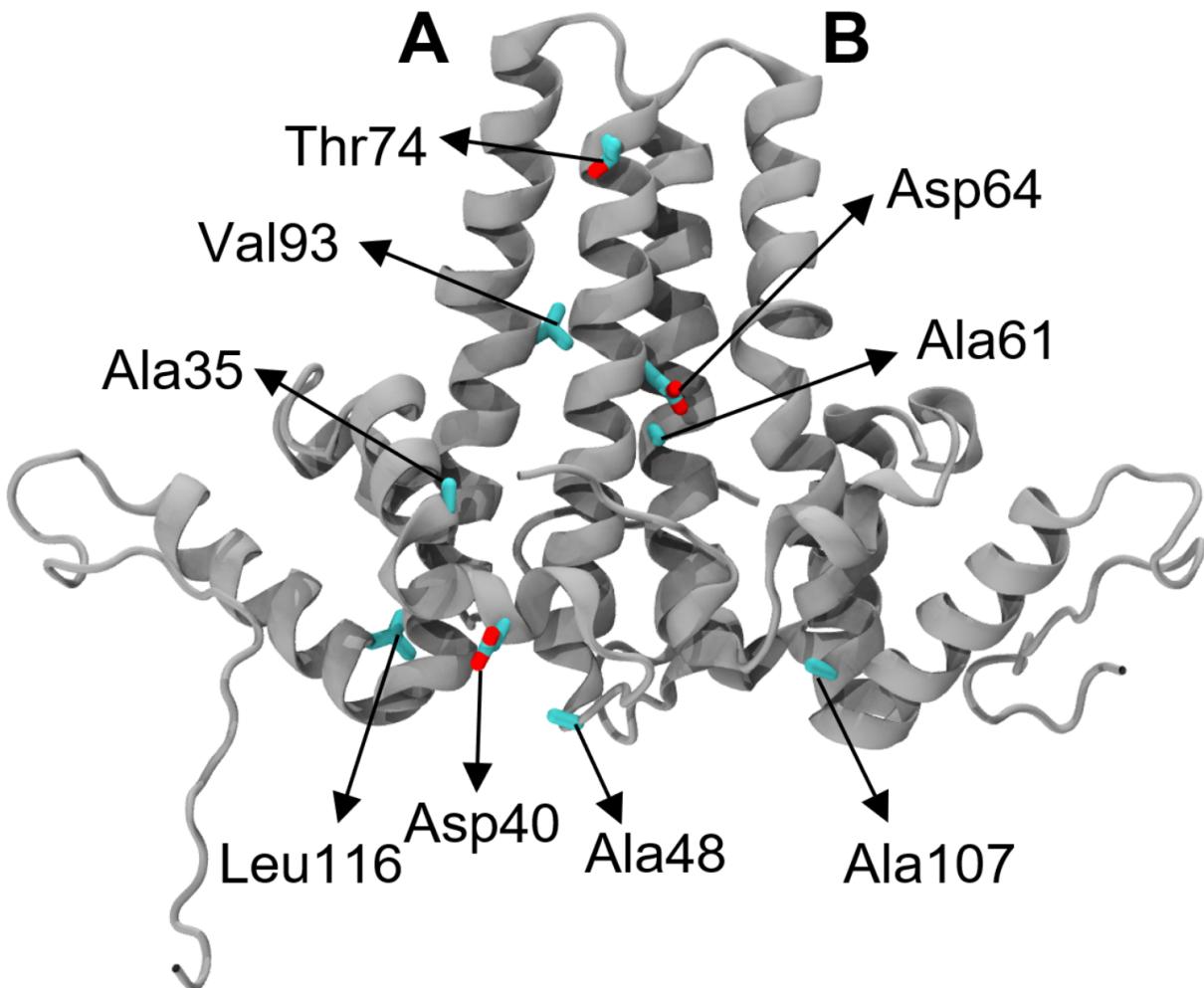
**Fig. S12:** Analytical form of the  $R_{1\rho}(^{15}\text{N})$  relaxation-rate constant for 13 kHz spin-lock strength as function of the MAS frequency and different correlation times ranging from  $5 \cdot 10^{-5} \text{ s}$  to  $5 \cdot 10^{-8} \text{ s}$ . The vertical dashed lines correspond to the used experimental conditions (80, 110 and 160 kHz MAS).



**Fig. S13:** (b) Analysis of the experimental rotating-frame relaxation rate constants recorded at 160/80 kHz, using  $\rho_2^{(\theta,S)}$ ,  $\rho_3^{(\theta,S)}$  and  $\rho_4^{(\theta,S)}$ ,  $\rho_5^{(\theta,S)}$  detector sensitivities depicted in (a).



**Fig. S14:** (a) Detector sensitivities  $\rho_1, \dots, \rho_7$  positioned at each order of magnitude on the correlation time axis. (b) root-mean-square fluctuations (RMFS) calculated for a A-B dimer in reference (*Hadden et al. 2018*) from main text. (c) Corresponding averaged detector responses extracted from the molecular dynamics trajectory for  $\rho_1, \dots, \rho_7$  with mean correlation times ranging from  $10^{-11.9}\text{s}$  to  $10^{-5.2}\text{s}$ . Inverse Laplace transforms and detector profile optimization have been performed over the correlation time interval  $\tau_c = [10^{-14}, 10^{-6}] \text{s}$ .



**Fig. S15:** The simulation model of mutated Cp149 and the wild-type sequence analysed with NMR in the present work differ by nine amino acid residues: A35/S35, D40/E40, A48/C48, A61/C61, D64/E64, T74/V74, V93/M93, A107/C107, and L116/I116. The location of each amino acid substitution is indicated on the Cp149 structure.

**Table S1:** Experimental tables for solid-state NMR acquisition and pulse program parameters used in the relaxation experiments.

Experiment	hCANH 3D, $T_{1\rho}(^{15}\text{N})$ , 110 kHz (DUL Cp149)	hCANH 3D $T_{1\rho}(^{15}\text{N})$ , 80 kHz (DUL Cp149)
MAS frequency/ kHz	110	80
Field/ T	20	20
Transfer I	HC-CP(DQ)	HC-CP(DQ)
$^1\text{H}$ field/ kHz	88	61
$^{13}\text{C}$ field/ kHz	15	15
Shape	Tangent $^1\text{H}$	Tangent $^1\text{H}$
Carrier / ppm	51	52
Time/ ms	4.5	4.5
Transfer II	CN-CP	CN-CP

<sup>13</sup> C field/ kHz	68	50
<sup>15</sup> N field/ kHz	40	29
Shape	Tangent <sup>13</sup> C	Tangent <sup>13</sup> C
Carrier/ ppm	117.5	117.5
Time/ ms	16.0	17.0
Transfer III	NH-CP	NH-CP
<sup>1</sup> H field/ kHz	83	59
<sup>15</sup> N field/ kHz	16	17
Shape	Tangent <sup>1</sup> H	Tangent <sup>1</sup> H
Carrier/ ppm	4.8	4.8
Time/ ms	2.7	2.4
<b><i>T<sub>1ρ</sub>(<sup>15</sup>N) Measurement</i></b>	13 kHz Spin-Lock <sup>15</sup> N	13 kHz Spin-Lock <sup>15</sup> N
Relaxation delays / ms	0.001, 26, 51, 101, 126, 151, 201, 251	0.001, 26, 51, 101, 126, 151, 201, 251
t1 increments	50	50
Sweep width (t1)/ ppm	30	30
Acquisition time (t1)/ ms	3.9	3.9
t2 increments	30	30
Sweep width (t2)/ ppm	40	40
Acquisition time (t2)/ ms	4.4	4.4
t3 increments	2048	2048
Sweep width (t3)/ ppm	47	47
Acquisition time (t3)/ ms	25.8	25.8
<sup>1</sup> H swfTPPM decoupling/ kHz	10	10
<sup>15</sup> N WALTZ64 decoupling/ kHz	10	5
<sup>13</sup> C WALTZ64 decoupling/ kHz	5	5
Water Suppression	MISSISSIPPI	MISSISSIPPI
<sup>1</sup> H field / kHz	20	20
Time / ms	120	120
Interscan delay/ s	2.19	2.19
Number of scans	24	24
Measurement time/ h	184	184

<b>Experiment</b>	<b>hNH 2D <i>T<sub>1ρ</sub>(<sup>15</sup>N)</i> 160 kHz (DUL Cp149)</b>	<b>hNH 2D <i>T<sub>1ρ</sub>(<sup>15</sup>N)</i> 80 kHz (DUL Cp149)</b>
MAS frequency/ kHz	160	80
Field/ T	20	20
<b>Transfer I</b>	HN-CP (DQ)	HN-CP (DQ)
<sup>1</sup> H field/ kHz	118	63
<sup>15</sup> N field/ kHz	30	15
Shape	Tangent <sup>1</sup> H	Tangent <sup>1</sup> H
Time / ms	1.1	1.4
<b>Transfer II</b>	NH-CP (DQ)	NH-CP (DQ)
<sup>1</sup> H field/ kHz	118	63

<sup>15</sup> N field/ kHz	30	15
Shape	Tangent <sup>1</sup> H	Tangent <sup>1</sup> H
Time / ms	1.2	1
<b>T<sub>1ρ</sub>(<sup>15</sup>N) Measurement</b>	13 kHz Spin-Lock <sup>15</sup> N	13 kHz Spin-Lock <sup>15</sup> N
Relaxation delays / ms	0.001, 51, 101, 151, 201, 251, 321, 371	0.001, 51, 101, 151, 201, 251, 321, 371
<sup>1</sup> H carrier/ ppm	4.7	4.7
<sup>15</sup> N carrier/ ppm	107	107
t1 increments	324	324
Sweep width (t1)/ ppm	70	70
Acquisition time (t1)/ ms	26.8	26.8
t2 increments	2048	2048
Sweep width (t2)/ ppm	40	40
Acquisition time (t2)/ ms	30.0	30.0
Water Suppression	MISSISSIPPI	MISSISSIPPI
<sup>1</sup> H field / kHz	5	5
Time / ms	30	30
<sup>1</sup> H swftpm decoupling power/ kHz	5	5
<sup>15</sup> N WALTZ64 decoupling power/ kHz	10	10
Interscan delay/ s	2	2
Number of scans	32	32
Measurement time/ h	50	50

<b>Experiment</b>	<b>CP-hNH T<sub>1</sub>'(<sup>15</sup>N) (DUL Cp149)</b>
MAS frequency/ kHz	108
Field/ T	20
<b>Transfer I</b>	HN-CP (DQ)
<sup>1</sup> H field/ kHz	86
<sup>15</sup> N field/ kHz	17
Shape	Tangent <sup>1</sup> H
Time / ms	1.3
<b>Transfer II</b>	NH-CP (DQ)
<sup>1</sup> H field/ kHz	86
<sup>15</sup> N field/ kHz	17
Shape	Tangent <sup>1</sup> H
Time / ms	1.3
<b>T<sub>1</sub>(<sup>15</sup>N) Measurement</b>	z storage on <sup>15</sup> N
Relaxation delays / s	0.02 1.02 2.02 4.02 6.02 8.02 10.02 12.02
<sup>1</sup> H carrier/ ppm	4.7
<sup>15</sup> N carrier/ ppm	107
t1 increments	160
Sweep width (t1)/ ppm	55
Acquisition time (t1)/ ms	16.9
t2 increments	2048
Sweep width (t2)/ ppm	40.1
Acquisition time (t2)/ ms	30.0
Water Suppression	MISSISSIPPI
<sup>1</sup> H field / kHz	10
Time / ms	30
<sup>1</sup> H swftpm decoupling power/ kHz	10
<sup>15</sup> N WALTZ64 decoupling power/ kHz	10
Interscan delay/ s	1.2
Number of scans	16
Measurement time/ h	38

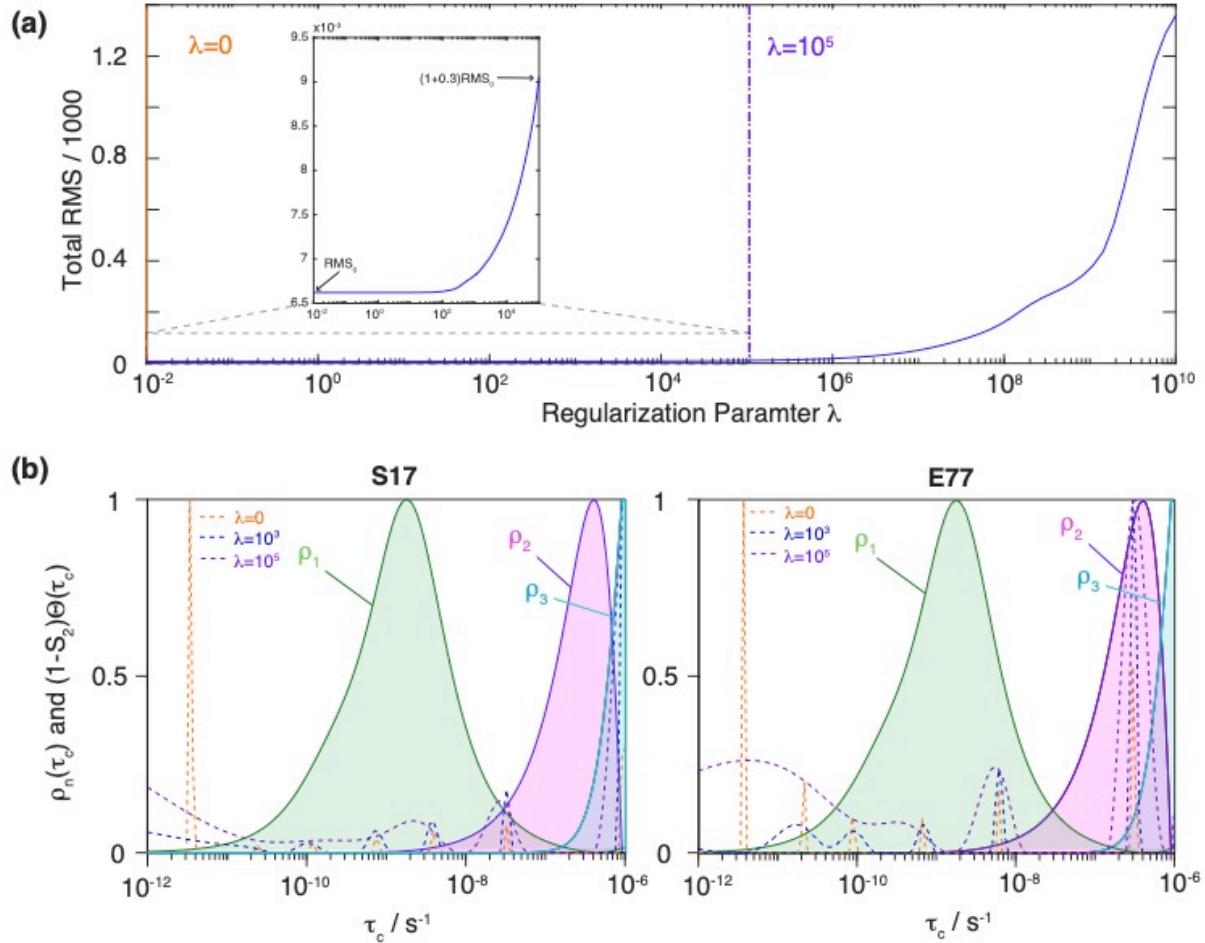
## **Section S1: Error Analysis on the detector responses obtained from the molecular dynamic trajectory.**

In this section we show the considerations that have been used to construct the estimated error on the quantitative MD prediction of amplitudes of motions at different timescales shown in the main text Fig. 3 (grey shaded area). We will discuss in detail two contributions leading to uncertainty, when trying to extract quantitative predictions on motions from the  $1\ \mu s$  MD trajectory and will illustrate how both errors grow significantly for timescales approaching the end of the trajectory in particular for the detectors constructed from the experimental conditions over a correlation time range from  $[10^{-14}, 10^{-6}]$  s. Doing the error investigation on a detector analysis that uses a set of seven almost non-overlapping detector sensitivities spaced at each order of magnitude (with mean correlation times ranging from  $10^{-11.6}$  s to  $10^{-5.0}$  s) allows for monitoring the increase of the error sources for slower timescales and thereby proposing the position of the boundary between the green and grey zones of the MD prediction, discussed in the main text. The whole analysis has been performed on chain A of the Cp149 capsid.

### **S1.1 Detector Responses for detectors approaching the end of the trajectory are less well defined (influence of the regularization parameter)**

As sketched in the main text Fig.3(a) amplitudes of motion at all timescales are extracted from the correlation function of the  $1\ \mu s$  MD trajectory by computing an inverse Laplace transform. This is mathematically an ill-posed problem, solved with the help of a regularization function, consisting in a minimization of the second derivative of the distribution of motion  $\Theta$ . The relative priority of minimizing the second derivative, compared to obtaining a good fit of the input correlation function is determined by the regularization parameter  $\lambda$ . Increasing the regularization parameter will in general lead to a smoother/broader distribution of motion. If the regularization parameter is increased too much, the obtained distribution of motion will broaden to such an extent that it will not lead to a good fit of the input correlation function anymore. This behaviour of the total fit error on the regularization parameter translates into an L-shaped curve typical of regularization problems. The L-curve for the dependence of the total fit error as function of the regularization weight in chain A of Cp149 is shown in Fig.S16(a). In the zoom given in the insert of Fig.S16(a) it can be seen that up to around  $\lambda = 10^5$ , the total fit error increases less than 50% of the initial error at  $\lambda = 0$ , while it starts growing substantially afterwards. Fig. S16(b) shows the expected broadening of the distribution of motion with increasing the regularization parameter ( $\lambda = 0, 10^3, 10^5$ ). It has been shown by (Smith et al. 2019) on HET-s(218-289) that this change in the form of the distribution of motion has in general only a minor impact on the extracted detector responses. Finding the distributions of motion is an ill-posed problem, but the step from the distributions of motion to the actual detector responses is such, that the result of the two-step process is a rather well-defined problem. Indeed, the detector sensitivities are convoluted with the distribution of motion and serve as a regularization function themselves, being much broader

than the individual peaks in the distribution of motion (green, pink and blue areas in Fig.S16(b)).



**Fig. S16:** Influence of the regularization parameter on the distributions of motion. **(a)** Total fit error between MD correlation function and correlation function obtained by solving the inverse Laplace problem as function of the regularization parameter  $\lambda$ . **(b)** Illustration of the effect of the regularization parameter on the distributions of motion  $(1-S^2)\Theta$  of two different residues for  $\lambda = 0, 10^3, 10^5$ . The coloured green, pink and blue areas correspond to the sensitivities  $\rho_n$  of the experimental detector sensitivities shown in the main text Fig.3(a), which are convoluted with the distribution of motion to obtain the detector responses.

However, motions at long correlation times, approaching the end of the trajectory can be indistinguishable from each other based on the time points available in the correlation function. This can lead to ill-defined detector responses introducing a dependence of the detector responses to the regularization function. Thus, a variety of slower motions can be fitted equally well, because the trajectory is not long enough and does not contain sufficient points to define a clear peak position. Fig.S16(b)) illustrates how the systematic broadening of the peaks in the distribution of motion due to an increase in regularization parameter can lead to an additional overlap with a neighbouring detector and thereby alter the detector responses, as it is particularly evident for E77. This is a possible source of error when describing slow motions. Note that this effect does not come from a bad fitting of the input

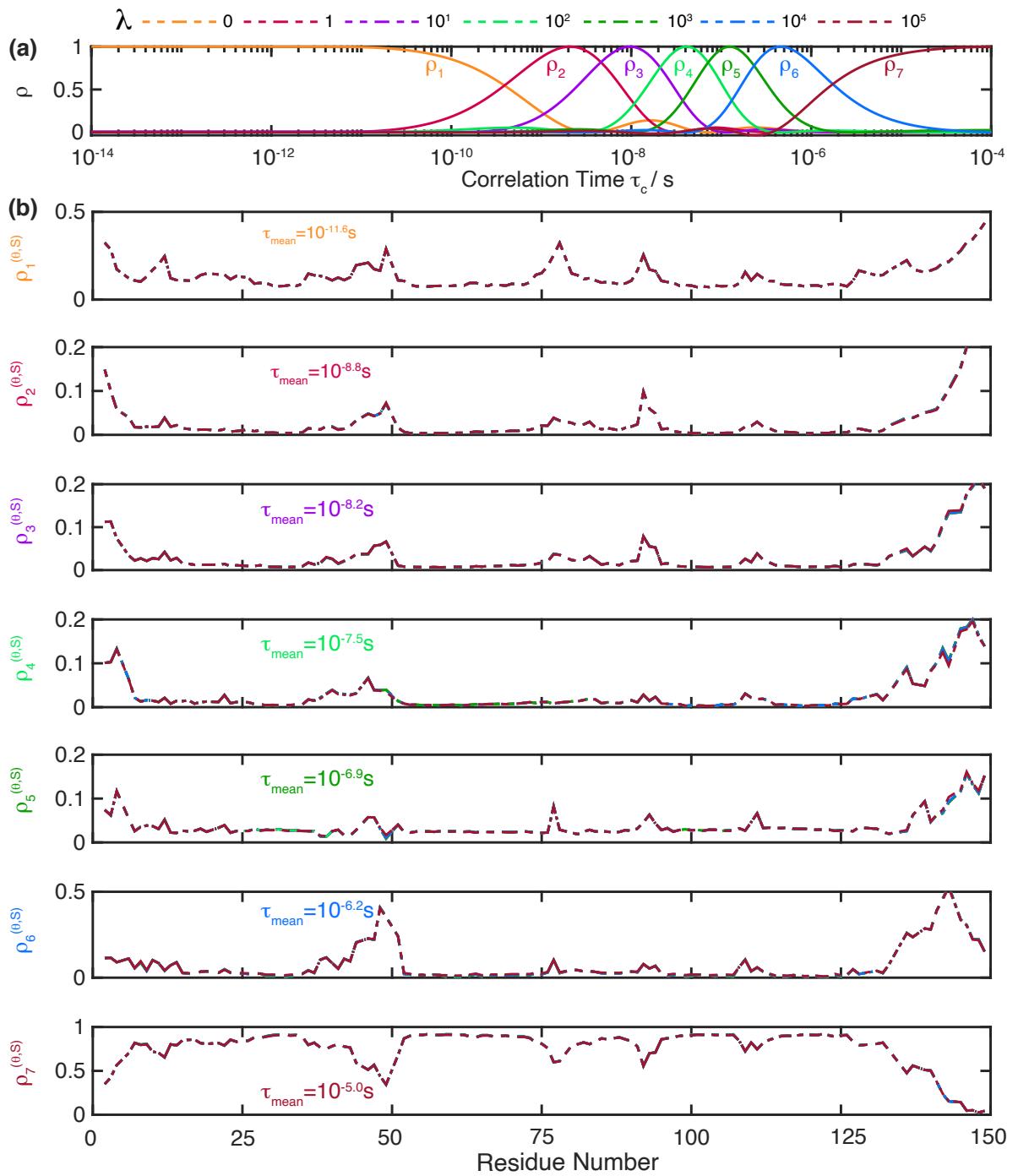
correlation function. All chosen  $\lambda$  values correspond to regularization weights for which the total RMS of the fit is small (see Fig.S16(a)).

This effect was systematically investigated over a large range of timescales, by using the seven detector sensitivities spaced at each order of magnitude shown in Fig.S17(a) together with distribution of motions obtained with regularization parameters between  $\lambda = 0$  and  $\lambda = 10^5$  (chosen such that the RMS increases less than 50% of its initial value for  $\lambda = 0$ ). We find that the curves are almost indistinguishable from each other at all timescales. Note however that this finding is also influenced by the choice and number of detectors used for the analysis. The seven detectors given in Fig.S17(a) are specifically optimized over the full correlation time range  $[10^{-14}, 10^{-4}]$  s based on the available data (the  $1 \mu\text{s}$  trajectory), such that it generates broad detectors where there is uncertainty and narrow detectors where there is more confidence. The situation is different when performing the regularization influence analysis specifically for the three detectors optimized from the experimental conditions and on the restricted correlation time range from  $[10^{-14}, 10^{-6}]$  s (Fig.S18). Indeed, we now observe a significant increase in regularization error at the slower timescales, which is likely due to the narrowness of the  $\rho_{0.6\mu\text{s}}$  detector, which is restricted by the ending of the MD trajectory.

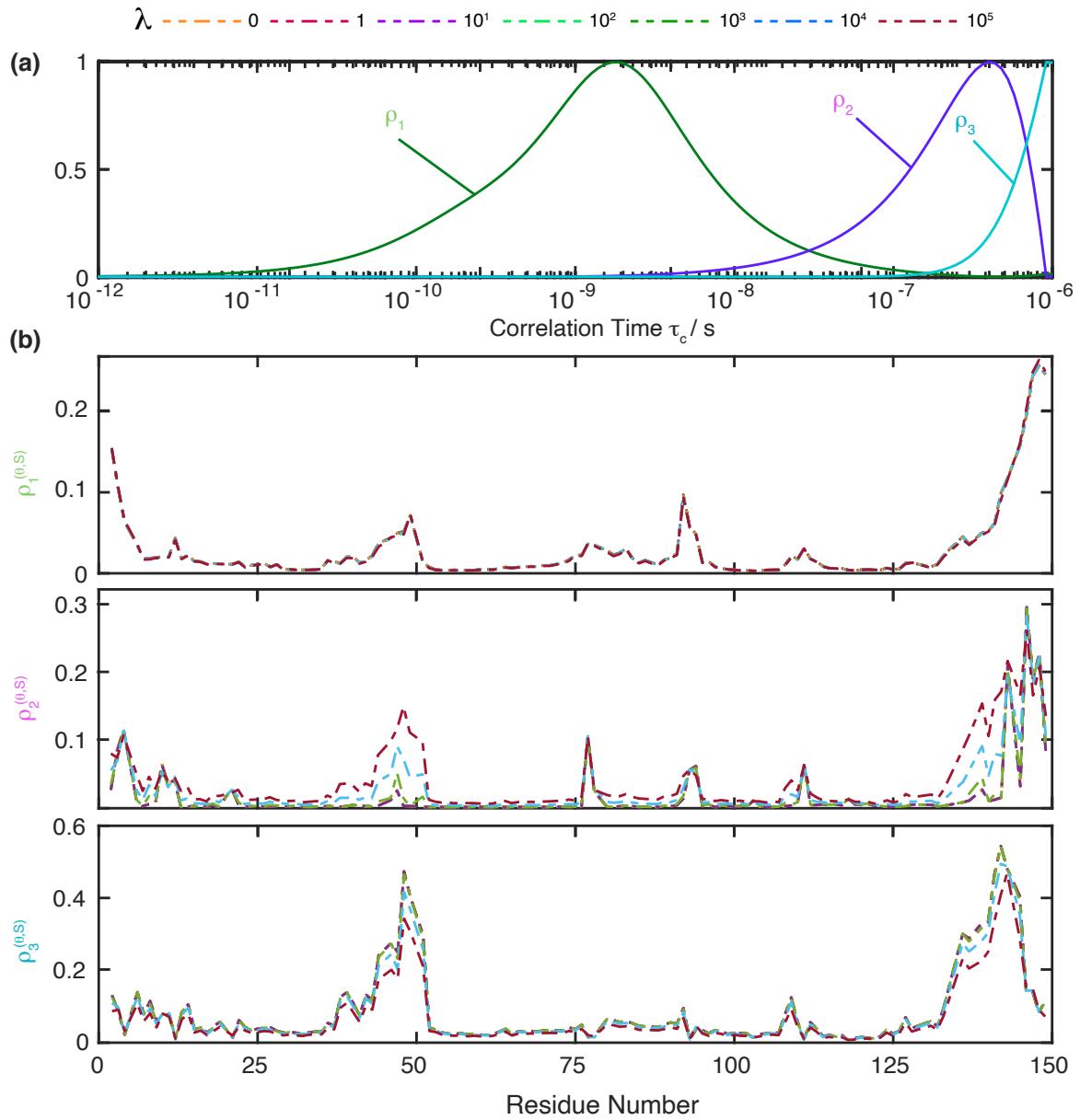
### S1.2 Motions approaching the longest correlation times are not equilibrated

The MD trajectory contains 60 copies of the Cp149 molecule. This allows to calculate the correlation function and inverse Laplace transform for all 60 copies and use it to obtain the corresponding detector responses for the set of seven detector sensitivities spaced at each order of magnitude (shown in Fig.S17(a)). The site-specific standard deviations over the different detector responses for all 60 copies could thus be calculated and compared between the different orders of magnitude. For better comparison the squared standard deviation (variance) was normalized by the mean over all 60 copies ( $\sigma^2/\text{mean}$ ) in Fig.S19, accounting for the amplitude of each motion. The variance was divided by the mean, instead of the standard deviation directly, since if one assumes an underlying Poisson statistic, the variance equals the expectation value for the mean.

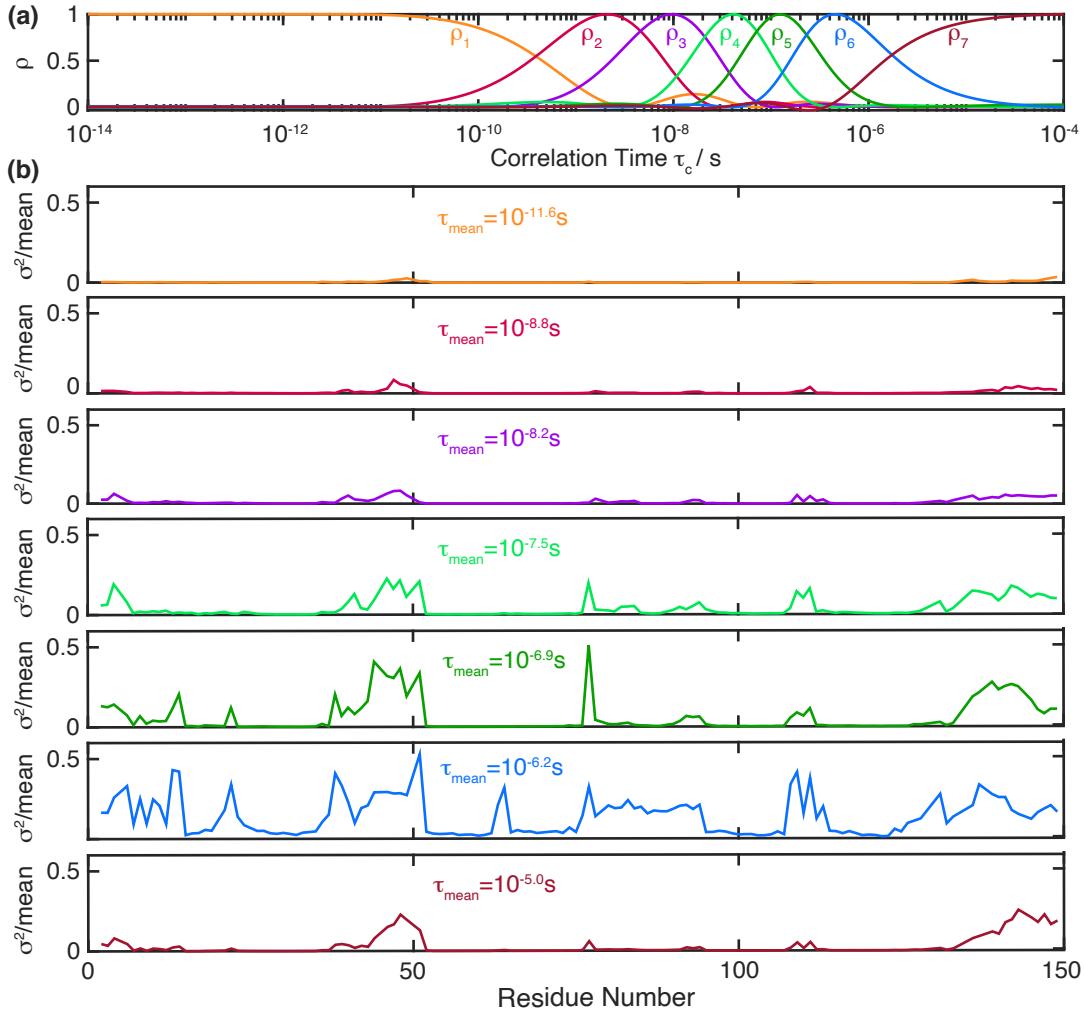
We observe, as expect, the total error to be smaller for fast motions and higher for slow motions approaching the end of the trajectory, due to the slower motions not being well equilibrated in the given trajectory. The only exception represents the detector centred at  $10^{-5}\text{s}$ , which is however special because it describes motions outside of the actual trajectory length, which are therefore purely extrapolated by the  $1 \mu\text{s}$  trajectory.



**Fig. S17:** Influence of the regularization parameter on detector responses. **(a)** Detector sensitivities spaced at each order of magnitude. **(b)** Detector responses for the detectors spaced at each order of magnitude from (d), using the inverse Laplace transforms obtained with a regularization parameters from  $\lambda = 0$  to  $10^5$ . Inverse Laplace transforms and detector profile optimization have been performed over the correlation time interval  $\tau_c = [10^{-14}, 10^{-4}] \text{s}$ .



**Fig. S18:** Influence of the regularization parameter on detector responses for the detector sensitivities obtained from the experimental conditions. **(a)** Detector sensitivities  $\rho_1, \rho_2, \rho_3$  constructed from the experimental conditions as used in the main text. **(b)** Detector responses for the experimental detector sensitivities, using the inverse Laplace transforms obtained with regularization parameters from  $\lambda = 0$  to  $10^5$  and computed over the correlation time interval  $\tau_c = [10^{-14}, 10^{-6}] \text{s}$ .



**Fig. S19:** Error statistics over 60 copies of Cp149. Site-specific squared standard deviation over the detector responses from the 60 copies of Cp149 divided by the respective mean for the detectors spaced at each order of magnitude from Fig.S16(a). Inverse Laplace transforms and detector profile optimization have been performed with  $\lambda = 0$  over the correlation time interval  $\tau_c = [10^{-14}, 10^{-4}] \text{ s}$ .

### S1.3 Statistical confidence in detector responses at longer correlation times can be improved by averaging over 60 copies of the molecule.

Using a  $1 \mu\text{s}$  trajectory implies that the statistical significance of nanosecond predictions should be better than the one of microsecond predictions, since the nanosecond motions are much better monitored during the length of the  $1 \mu\text{s}$  trajectory. This improves the extraction of these motions from the correlation function. A possibility to estimate and improve the statistical significance of predictions at all timescales, would be to run a large number of molecular dynamic trajectories with different starting conditions and compare the obtained results. Due to the heavy computational demands for a large viral assembly, this cannot be applied on Cp149. However, since the published MD trajectory contains information on the trajectory of 60 copies of the Cp149 molecule, it is alternatively possible to increase the statistical significance of the predictions for slower motions by averaging over the distribution

of detector responses generated by all 60 copies, which we eventually also used to define the estimated error on the MD prediction (grey shaded area in Fig.3).

Prior to use such an approach, it is important to find out if the motion of the different copies is correlated, as strong correlations between the copies could bias the result. In the following we will discuss how to investigate the presence of such correlations and that they don't play a significant role for Cp149 in particular for slower motions, supporting the validity of an averaging approach over the 60 molecular copies.

We identify correlations based on calculations of the 60 detector responses for the correlation functions calculated from the *first* and the *second* half of the trajectory (both extending over a length of 500 ns). Two methods of error statistics are then performed on these  $2 \times 60$  detector responses on different timescales. The first method will lead to the correct variance of the statistics, while the second will lead to a different value only if the copies can be considered to be correlated.

We generalize the problem and consider  $2 \times N$  samplings of a distribution (mean:  $\mu$ , standard deviation:  $\sigma$ , in our case  $N=60$ ). There, the two copies of the sample are independent of each other ( $1^{\text{st}}$  and  $2^{\text{nd}}$  half of the trajectory being independent), but each set of  $N$  samplings may not be (there might be a correlation between copies which we want to find out).

Method 1, consists in calculating the mean over the copies and variance over  $1_{\text{st}}/2_{\text{nd}}$  half. We consider that  $E[x_{i,1}] = E[x_{i,2}] = E[x_{j,1}] = E[x_{j,2}]$  (all the variables are distributed the same way). Further note that  $E[x_{i,1}x_{i,2}] = E[x_{i,1}]E[x_{i,2}] = E[x_{i,1}]^2$  as we assume that the two copies are independent. We thus obtain from this approach the correct variance  $\sigma^2$  of the distribution (result of equation [1]).

$$\begin{aligned}
 & E \left[ \frac{1}{N} \sum_i^N \left( x_{i,1} - \frac{1}{2}(x_{i,1} + x_{i,2}) \right)^2 + \left( x_{i,2} - \frac{1}{2}(x_{i,1} + x_{i,2}) \right)^2 \right] \\
 &= E \left[ \frac{1}{N} \sum_i^N \left( x_{i,1}^2 - x_{i,1}^2 - x_{i,1}x_{i,2} + \frac{1}{4}x_{i,1}^2 + \frac{1}{4}x_{i,2}^2 + \frac{1}{2}x_{i,1}x_{i,2} \right) \right. \\
 &\quad \left. + \left( x_{i,2}^2 - x_{i,2}^2 - x_{i,1}x_{i,2} + \frac{1}{4}x_{i,1}^2 + \frac{1}{4}x_{i,2}^2 + \frac{1}{2}x_{i,1}x_{i,2} \right) \right] \\
 &= E \left[ \frac{1}{N} \sum_i^N \left( \frac{1}{2}x_{i,1}^2 + \frac{1}{2}x_{i,2}^2 - x_{i,1}x_{i,2} \right) \right] = \frac{1}{N} \sum_i^N E[x_{i,1}^2] - E[x_{i,1}]E[x_{i,2}] \\
 &= E[x_{i,1}^2] - (E[x_{i,1}])^2 = \sigma^2
 \end{aligned} \tag{1}$$

In method 2, the variance is taken over the 60 potentially correlated copies, and then the average over the uncorrelated pair of data ( $1^{\text{st}}$  and  $2^{\text{nd}}$  half). The variance over the 60 copies is yield by  $E[x_{i,1}^2] - E[x_{i,1}]^2$  as following (the expectation for the average over both halves

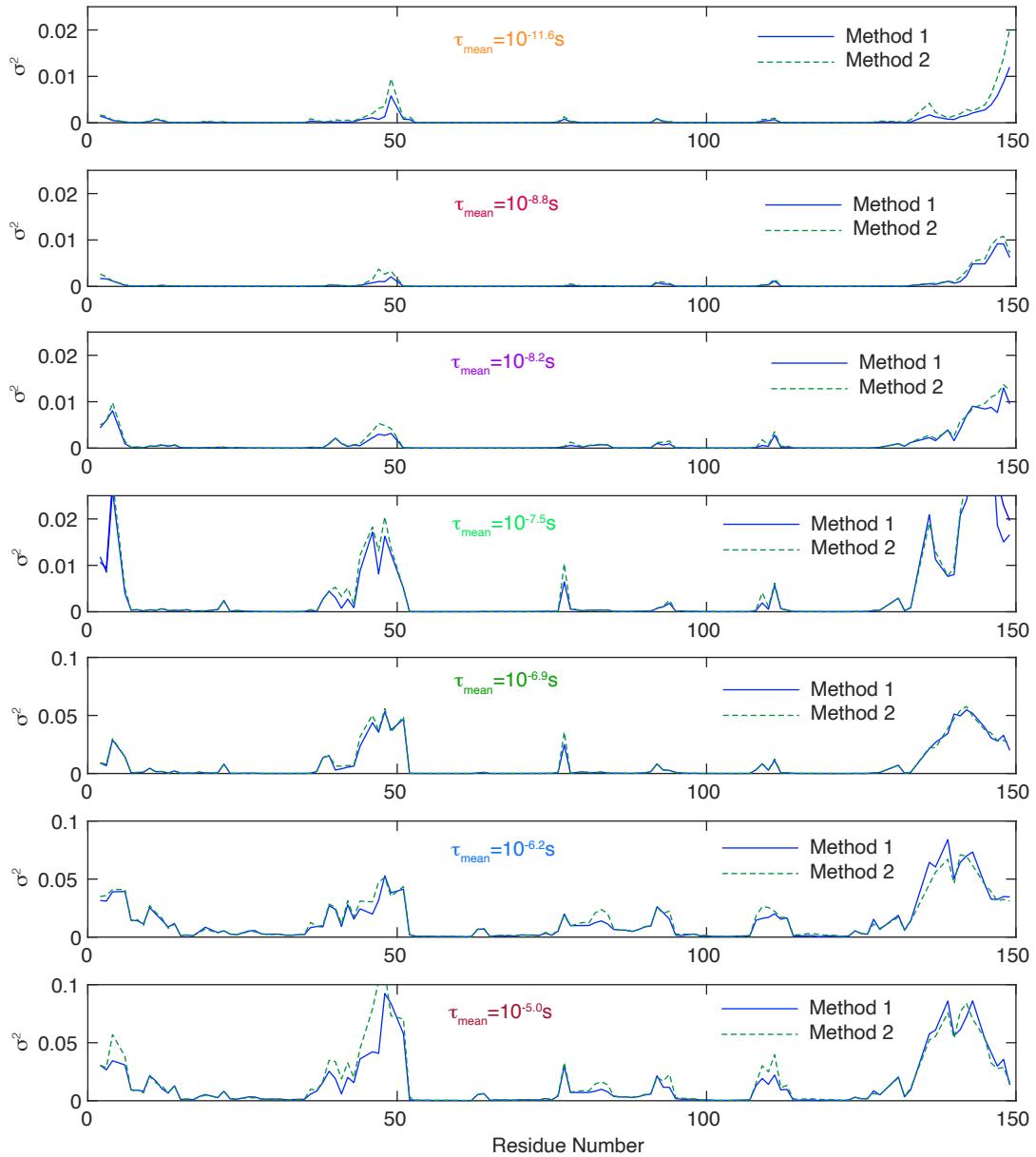
should be the same as the expectation for just one half). We can assume that the sum over all covariances relative to some index  $i$  is the same regardless of the starting index.

$$\begin{aligned}
E \left[ \frac{1}{N-1} \sum_i^N \left( x_{i,1} - \frac{1}{N} \sum_i^N x_{j,1} \right)^2 \right] &= \frac{1}{N-1} E \left[ \sum_i^N \left( x_{i,1}^2 - \frac{2}{N} \sum_i^N x_{i,1} x_{j,1} + \frac{1}{N^2} \sum_i^N x_{j,1} x_{k,1} \right) \right] \\
&= \frac{1}{N-1} \sum_i \left( E[x_{i,1}^2] - \frac{2}{N} E[x_{i,1}^2] - \frac{2}{N} \sum_{j \neq i} E[x_{i,1} x_{j,1}] + \frac{1}{N^2} \sum_j E[x_{j,1}^2] + \frac{1}{N^2} \sum_{\substack{j,k \\ k \neq j}} E[x_{j,1} x_{k,1}] \right) \\
&= \frac{1}{N-1} (N-2+1) E[x_{i,1}^2] + \\
&+ \frac{1}{N-1} (-2) \sum_{j \neq i} \left( E[x_{i,1}]^2 + \text{cov}(x_{i,1}, x_{j,1}) \right) + \frac{1}{N} \sum_{\substack{j,k \\ k \neq j}} \left( E[x_{j,1}]^2 + \text{cov}(x_{j,1}, x_{k,1}) \right) \\
&= E[x_{i,1}^2] + \frac{1}{N-1} (-2(N-1) + (N-1)) E[x_{i,1}]^2 \\
&\quad + \frac{1}{N-1} \left( -2 \sum_{j \neq i} \text{cov}(x_{i,1}, x_{j,1}) + \frac{1}{N} \sum_{\substack{j,k \\ k \neq j}} (\text{cov}(x_{j,1}, x_{k,1})) \right) \\
&= E[x_{i,1}^2] - E[x_{i,1}]^2 + \frac{1}{N-1} \left( -2 \sum_{j \neq i} \text{cov}(x_{i,1}, x_{j,1}) + \frac{1}{N} \sum_{\substack{j,k \\ j \neq k}} \text{cov}(x_{j,1}, x_{k,1}) \right) \\
&= \sigma(x_{i,1}^2) + \frac{1}{N-1} \left( -2 \sum_{j \neq i} \text{cov}(x_{i,1}, x_{j,1}) + \sum_{j \neq k} \text{cov}(x_{j,1}, x_{k,1}) \right) \\
&= \sigma(x_{i,1}^2) - \frac{1}{N-1} \sum_{j \neq i} \text{cov}(x_{i,1}, x_{j,1}). \tag{2}
\end{aligned}$$

According to equation [2] if the covariance (correlation) is zero between copies, the two methods should yield the same result. However, if there is a positive/negative covariance between copies then the variance calculated by method 2 is smaller/higher than the one calculated with method 1.

Fig.S20 displays the results of the error statistics applying method 1 (blue, correlation unbiased) and method 2 (green dashed, correlation biased) to the detector responses of the 60 copies at the seven detector sensitivities shown in Fig.S17(a). The results for both methods

agree to a very good degree over most portions of the molecule. The only regions where the two methods show larger deviations are the loop from 44-50 and the last part of the molecule approaching the C-ter (on the other side, the N-ter does not show any large correlations). We observe slight tendency to an increased difference between the two methods for the slowest detector, hinting to an increased correlation between the motions. Overall these results indicate that the MD error analysis can be improved, by averaging over the copies without having to account for covariance between copies, avoiding the computationally demanding step of performing a multitude of different MD  $1 \mu\text{s}$  trajectories using different starting conditions.



**Fig. S20:** Cp149 MD error statistics using method 1 (correlation unbiased) and method 2 (correlation biased) discussed in section S1.3 for the detector sensitivities in Fig.S17(a). Inverse Laplace transforms and detector profile optimization have been performed over the correlation time interval  $\tau_c = [10^{-14}, 10^{-4}] \text{ s}$ .

#### S1.4: Estimation of the total MD error at different timescales

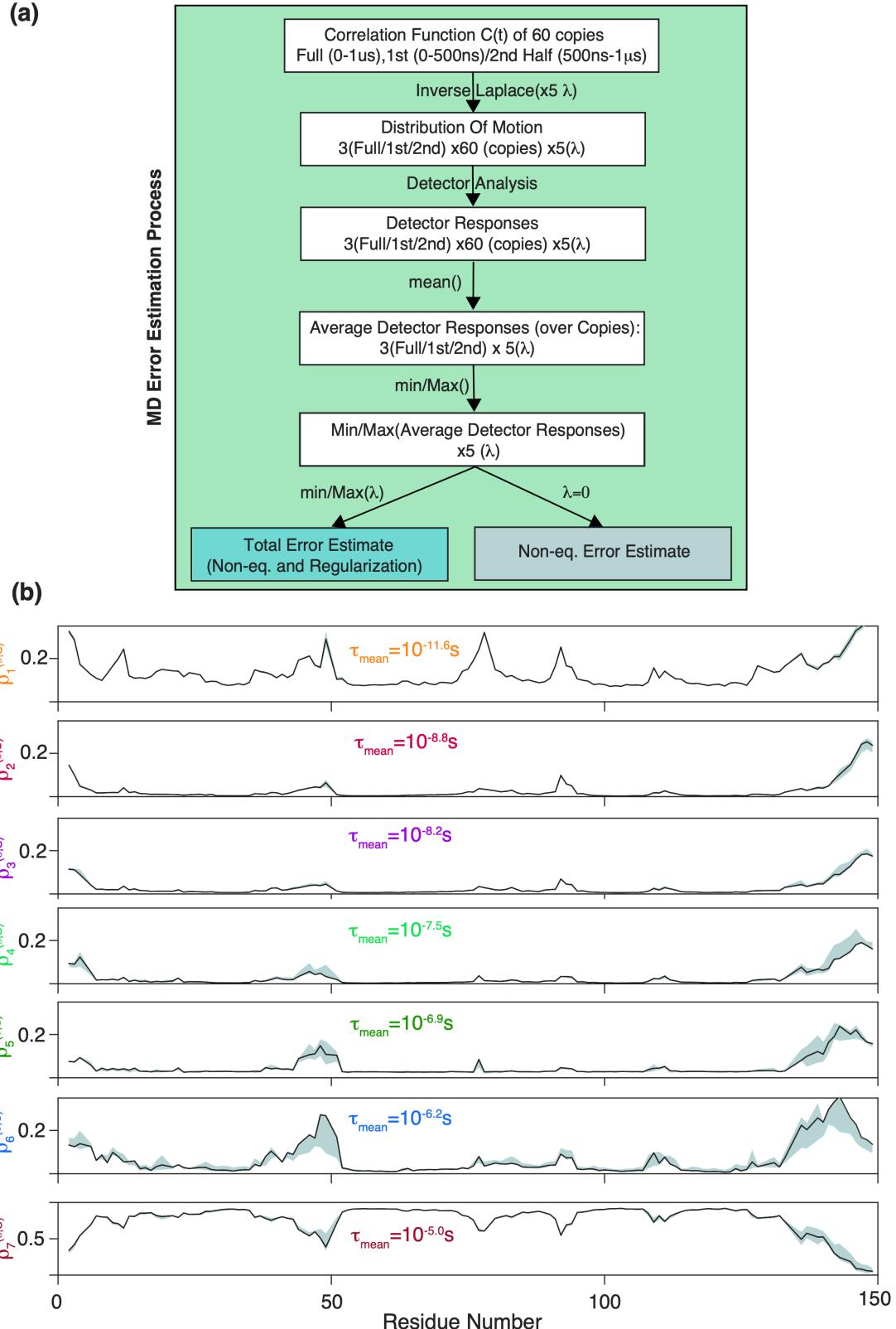
The total estimated error on the quantitative MD prediction of amplitudes of motions at different timescales (grey shaded area in Fig.3) could be constructed combining the two effects we described above (S1.1 and S1.2).

The general procedure for the error estimation is sketched in Fig.S21(a). We start by extracting for each of the 60 copies a correlation function for the first and second half of the trajectory as well as for the full trajectory. Each correlation function is inverse Laplace transformed with 5 logarithmically-spaced regularization parameters  $\lambda$  ranging from  $[0, 10^5]$ . Note that this range of values corresponds to regularization parameters which lead to a good fit of the input correlation function and therefore representing a valid solution of the inverse-Laplace transform problem (as it has been shown in Fig.S16(a)).

The result comprises  $3 \times (1^{\text{st}}/2^{\text{nd}}, \text{full}) \times 60(\text{copies}) \times 5(\lambda) = 900$  distributions of motion. We now extract detector responses for all distributions of motion and perform an average over all 60 copies for each regularization parameter and for the  $1^{\text{st}}/2^{\text{nd}}$  half and the full trajectory. This leaves us with 3x5 averaged responses. We have shown in section S1.3 that it is possible to perform such an averaging over copies, without having to account for strong correlation between copies. For each regularization parameter  $\lambda$  we define a grey shaded area, which ranges from the minimum to the maximum of all detector responses for the  $1^{\text{st}}/2^{\text{nd}}$  half and the full trajectory. This represents the amount of the MD error that can account for the non-equilibration of slower motions within the length of the trajectory (see section S.1.2). The grey shaded area in Fig.S21 shows this portion of the MD error for  $\lambda = 0$ . As it can be seen the non-equilibration error grows systematically for slower motion compared to the one for faster motions. The blue shaded area in Fig.S21 is created by taking the minimum and maximum value of the averaged detector responses over the grey shaded areas calculated for all regularization parameters  $\lambda$ . For this reason, the grey shaded area for  $\lambda = 0$  shown in Fig.S21 is always contained in the blue shaded area and the difference between blue and grey shaded area accounts for the portion of the MD error that can be attributed to the instability of the detector responses with respect to the choice of the regularization parameter. As illustrated in section S.1.1 this second error source remains however much smaller. Fig. S22 estimates the MD error according to the same procedure for the detectors  $\rho_{1.3\text{ns}}$ ,  $\rho_{0.2\mu\text{s}}$ ,  $\rho_{0.6\mu\text{s}}$  constructed from the experimental conditions and used for the analysis in the main text. As already seen in Fig. S18, due to slower motion being ill-defined, based on the time points available in the trajectory and the narrowness of the  $\rho_{0.6\mu\text{s}}$  detector due to the ending of the trajectory, the regularization error (blue-shaded) is substantially larger for the slower motions approaching  $1\ \mu\text{s}$ , thereby becoming an important contribution for the total error.

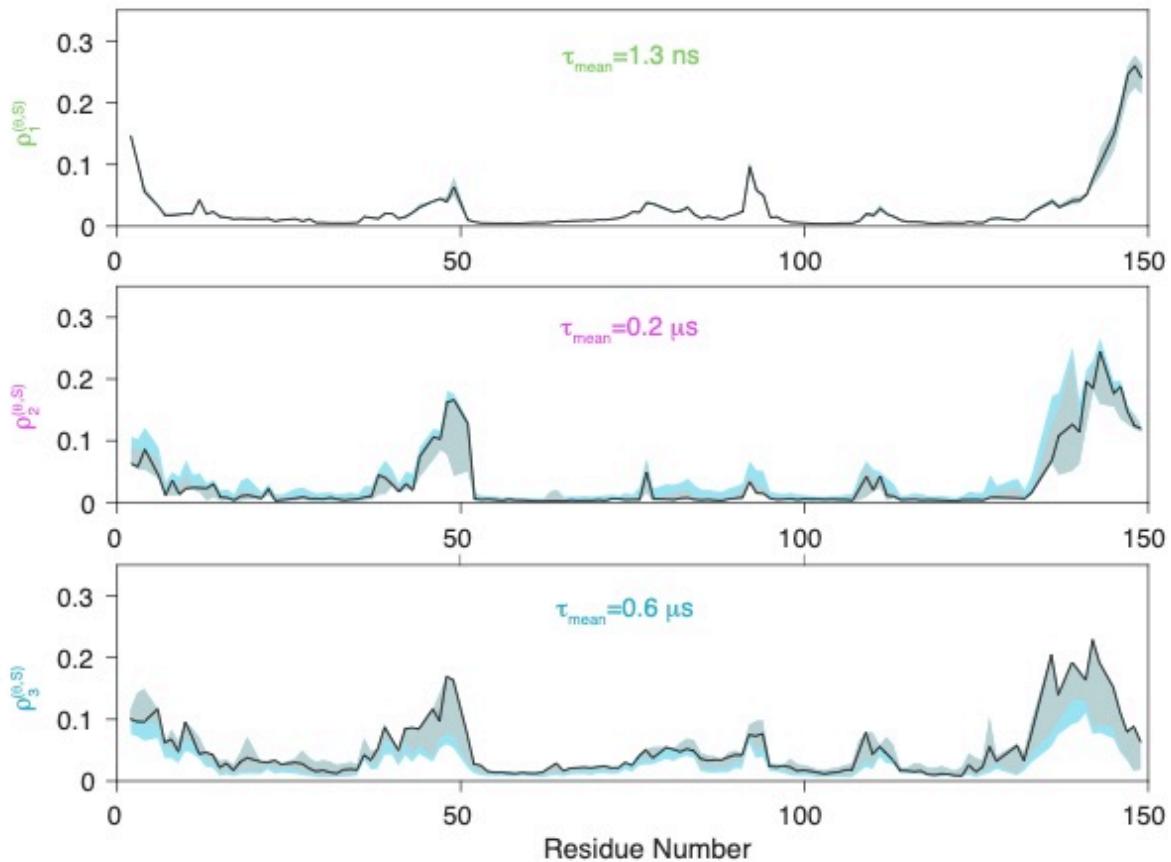
In summary, the combination of both sources of MD error in a total MD error estimate for detector responses obtained from independent detectors spaced at each order of magnitude (blue shaded area in Fig.S21) shows that around timescales in the order of  $10^{-7}\text{s}$  the MD uncertainty starts growing significantly and becomes particularly large when using the detectors approaching the end of the trajectory obtained from the experimental conditions used in the main text. For these reasons, we propose to set the boundary of the

green and grey areas for quantitative MD predictions for Cp149, using a detectors approach to  $10^{-7}$ s, which is around one order of magnitude smaller than the length of the trajectory.

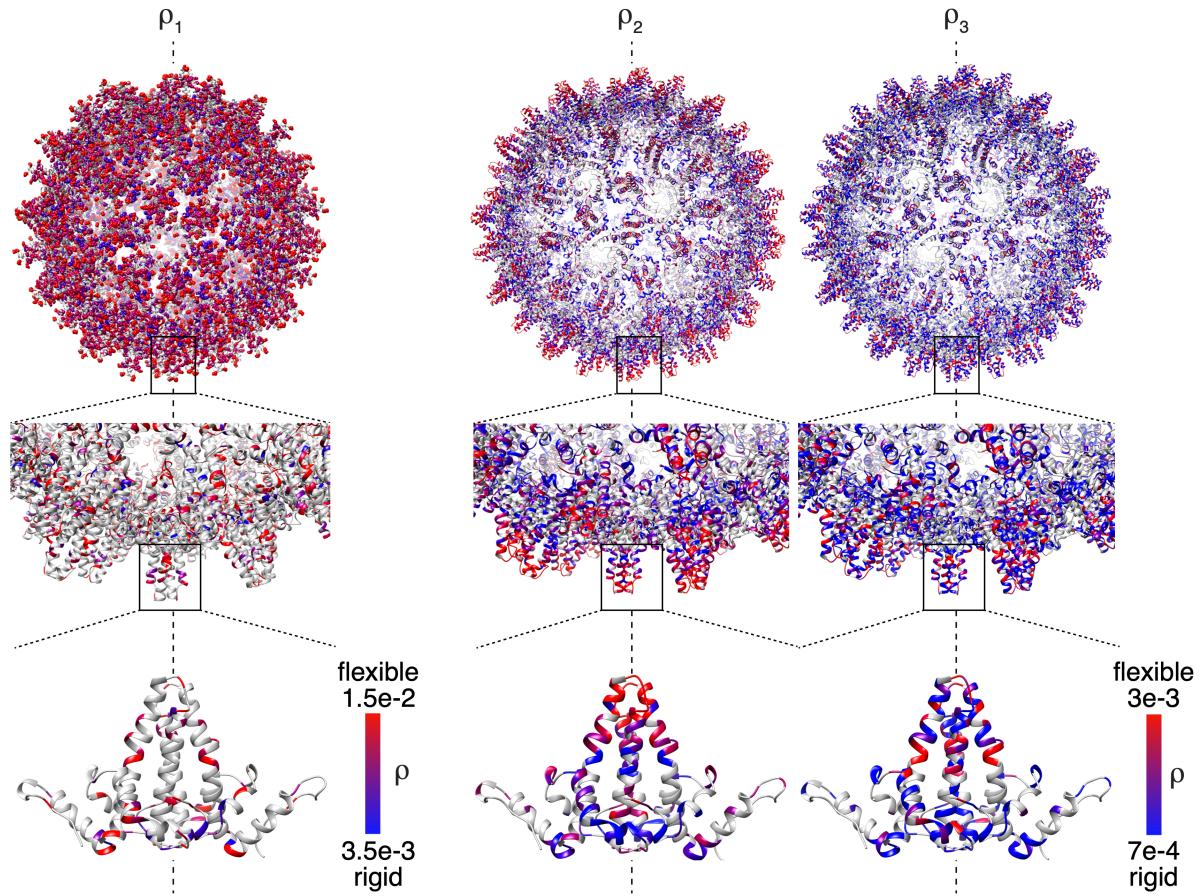


**Fig. S21:** Schematic summary of the procedure used for the MD error estimation **(a)**. Detector responses and corresponding MD error estimate for the seven detectors spaced at each order of magnitude from Fig.S17(a). The black line corresponds to the average detector response over all 60 copies for the full trajectory for  $\lambda = 0$ . The grey shaded area is comprised in between the maximum and minimum values of the average detector

response for the 1<sup>st</sup>/2<sup>nd</sup> half and the full trajectory for  $\lambda = 0$ . The blue shaded area corresponds to the maximum and minimum values of all grey shaded areas that have been calculated over the 5  $\lambda$  values.



**Fig. S22:** Detector responses and corresponding MD error estimate for the detectors  $\rho_1, \rho_2, \rho_3$  constructed from the experimental conditions and used in the main text. The black line corresponds to the average detector response over all 60 copies for the full trajectory for  $\lambda = 0$ . The grey shaded area is comprised in between the maximum and minimum values of the average detector response for the 1<sup>st</sup>/2<sup>nd</sup> half and the full trajectory for  $\lambda = 0$ . The blue shaded area corresponds to the Maximum and Minimum values of all grey shaded areas that have been calculated over the 5  $\lambda$  values.



**Fig. S23:** NMR amplitude of motions  $\rho_1^{(0,S)}$ ,  $\rho_2^{(0,S)}$  and  $\rho_3^{(0,S)}$  shown in a red to blue gradient on the Cp149 dimer structure (PDB 1QGT) at different scales from the full capsid (top) to the dimer (bottom). Residues not characterized by NMR are shown in grey.









