

**The preference signature of the SARS-CoV-2 Nucleocapsid NTD
for its 5'-genomic RNA elements**

Sophie Marianne Korn^{1,2}, Karthikeyan Dhamotharan^{1,2}, Andreas Schlundt^{1,2,*}

¹Institute for Molecular Biosciences and

²Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe-University Frankfurt, Max-von-Laue-Str. 7-9, 60438 Frankfurt/M., Germany.

* To whom correspondence should be addressed: schlundt@bio.uni-frankfurt.de

Supplementary Information

This document contains:

- 1 Supplementary Table
- 13 Supplementary Figures

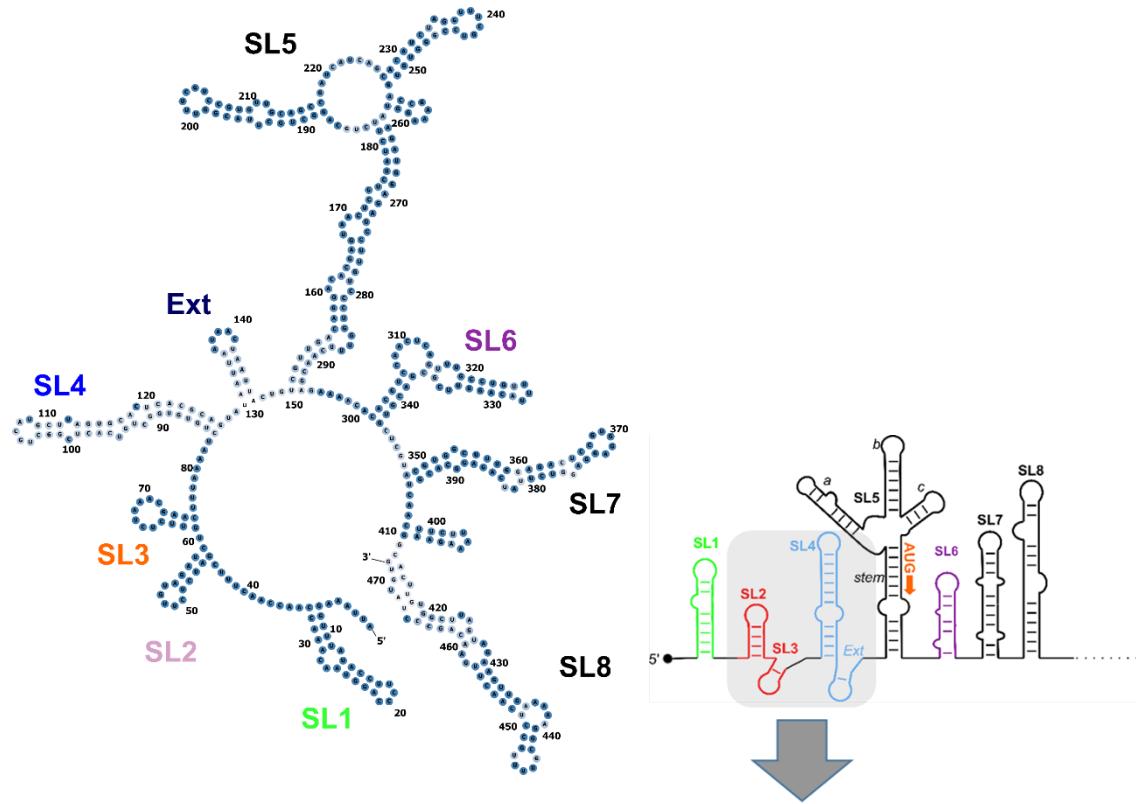
Supplementary Table 1, related to Table 1:

Table 1: Overview of RNAs and sequences used in this study. Small letters indicate non-genomic nucleotides artificially added to facilitate T7 *in vitro* transcription.

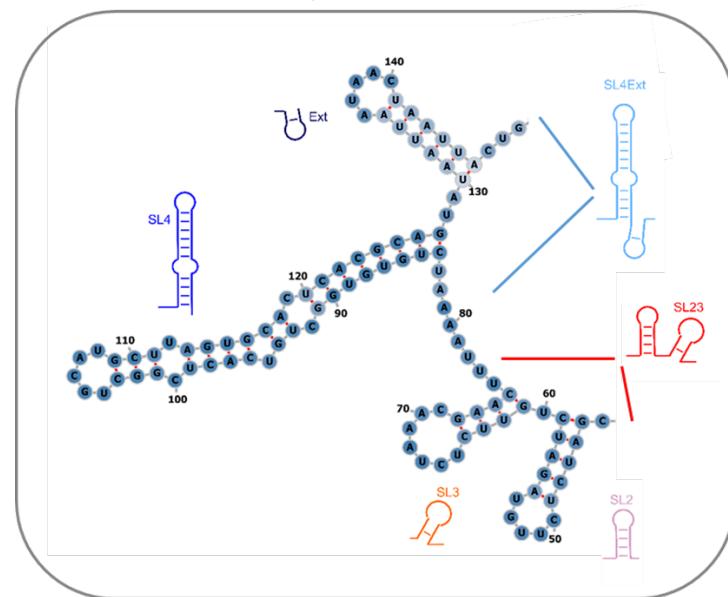
RNA	Genomic position	nts	Sequence 5'-3'
SL1	7-33	27	GGUUUAUACCUUCCCAGGUACAAACC
SL2+3	45-75	32	gGAUCUCUUGUAGAUCUGUUCUAAACGAAC
SL4	86-125	44	ggGUGUGGCUGUCACUCGGCUGCAUGCUCUAGUGCACUCACG Ccc
SL4ext	86-149	64	gUGUGGCUGUCACUCGGCUGCAUGCUCUAGUGCACUCACGCA GUAUAAAUAUAACUAAUACUG
Ext	127-148	22	ggAUAAUUAUAACUAAUACU
SL5	149-297	150	gGUCGUUGACAGGACACGAGUAACUCGUCUAUCUUCUGCAG GCUGCUUACGGUUUCGUCCGUGUUGCAGCCGAUCAUCAGCA CAUCUAGGUUUCGUCCGGGUGUGACCGAAAGGUAGAUGGA GAGCCUUGUCCUCGGUUCAACGAGAA
SL6	302-343	46	ggCACGUCCAACUCAGUUUGCCUGUUUACAGGUUCGCGAC GUGC
P2	726-756	33	ggAUGAAGAUUUCAAGAAAACUGGAACACUAA
ss19T	-	19	GGCACAAUUAACGUCGCC
ss19B	-	19	GGCGACGUUAUAUUGUGCC
ds19	-	2x19	Annealed ss19T and ss19B
SL_AUA	-	20	GGCGCGCAUUAAGCGCGCC

Supplementary Fig. 1, related to Fig. 1:

a

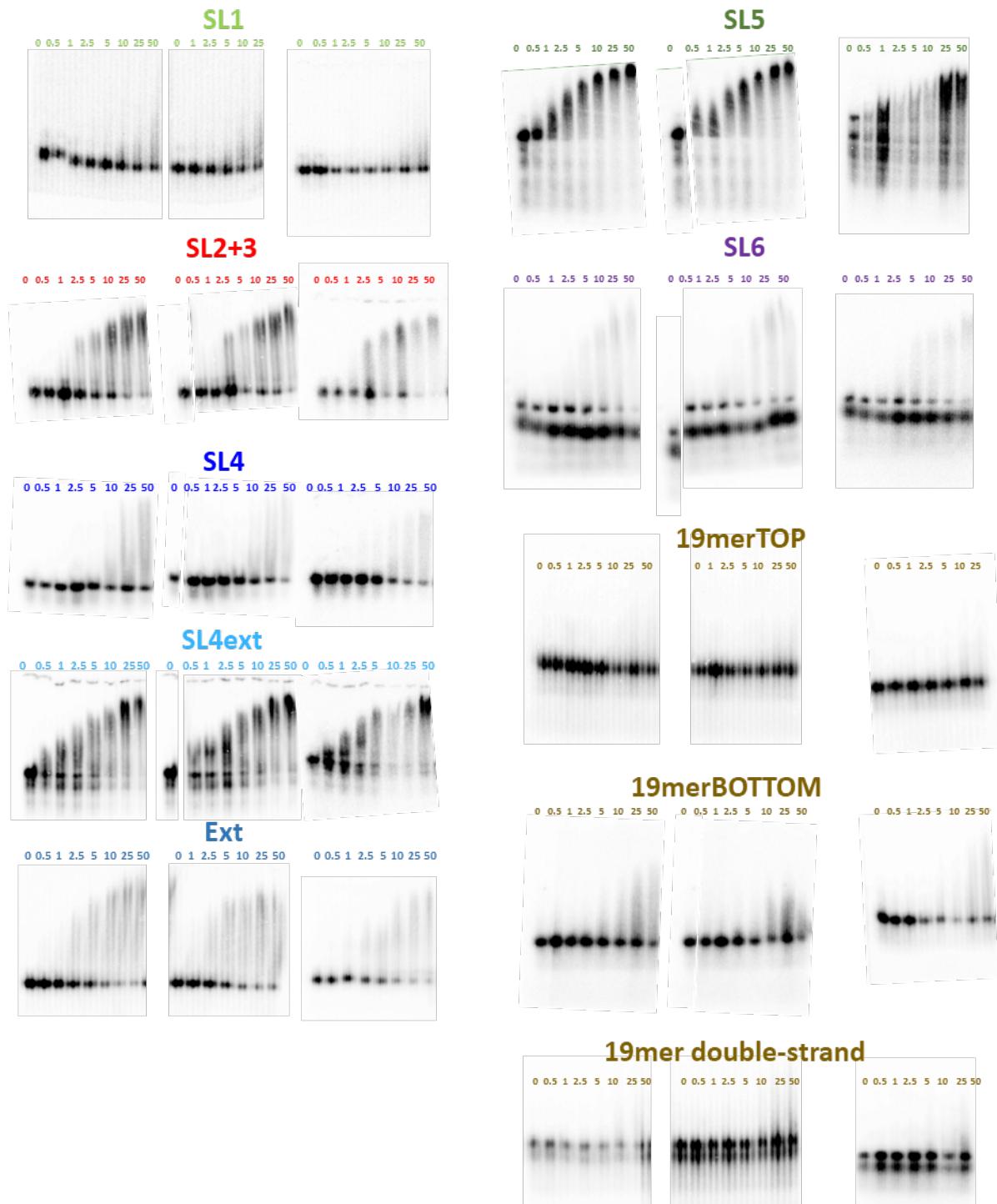


b



Supplementary Figure 1: The SARS-CoV-2 5'-genomic end, including the 5'UTR and SL6 to SL8. **a** Secondary structure prediction as obtained from Vienna RNA fold suite¹ supported by available experimental restraints² and visualized using the FORNA tool including the genomic numbering from SL1 to SL8. Individually explored elements are labelled with their colour code. **b** Zoom-in highlighting the SL2+3 and the SL4 hubs within the scheme of panel a as well as its position in the linear depiction on top.

Supplementary Fig. 2, related to Fig. 1:

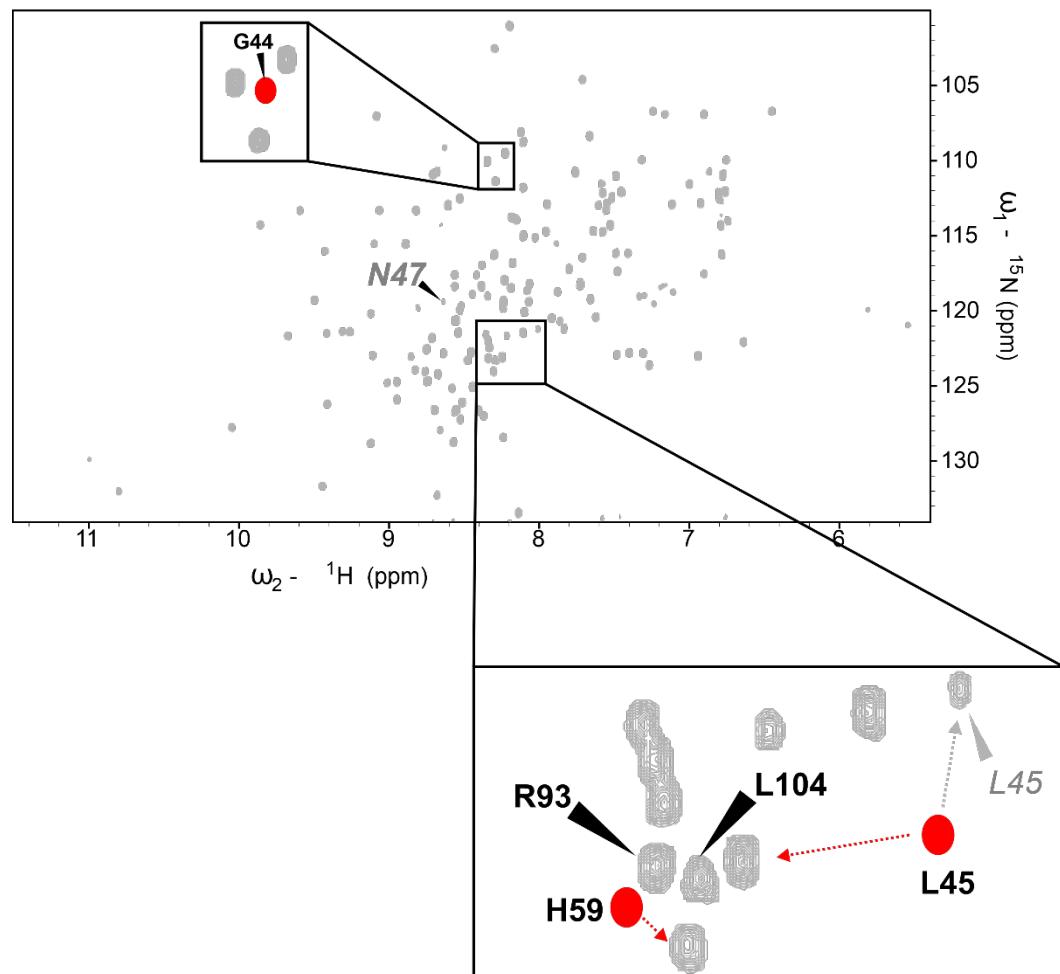


Supplementary Fig. 2: EMSAs of NTD with SARS-CoV-2 5'-genomic RNA elements (comprising RNAs from SL1 to SL6) and control RNAs (19mer RNAs). Plot of EMSAs as triplicate experiments. Shown are radiographs of native PAGEs with ^{32}P -labelled RNAs as given and titrated with different concentrations of NTD shown above lanes.

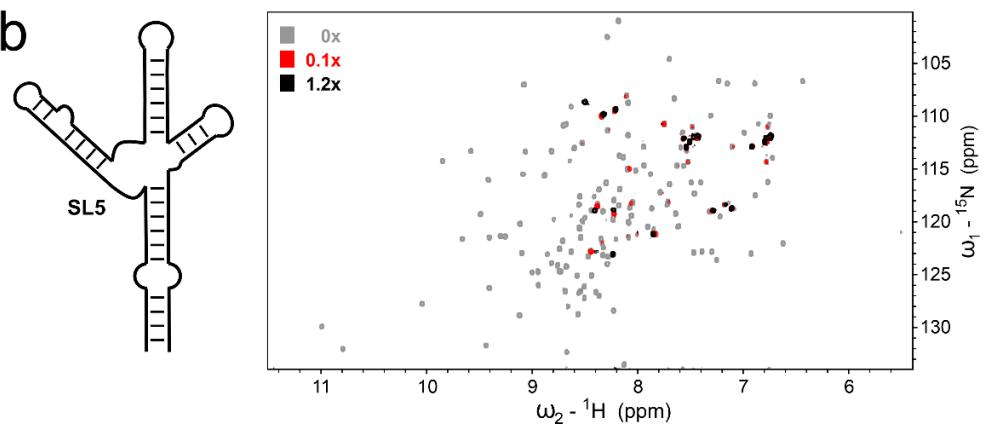
Supplementary Fig. 3, related to Fig. 2:

a

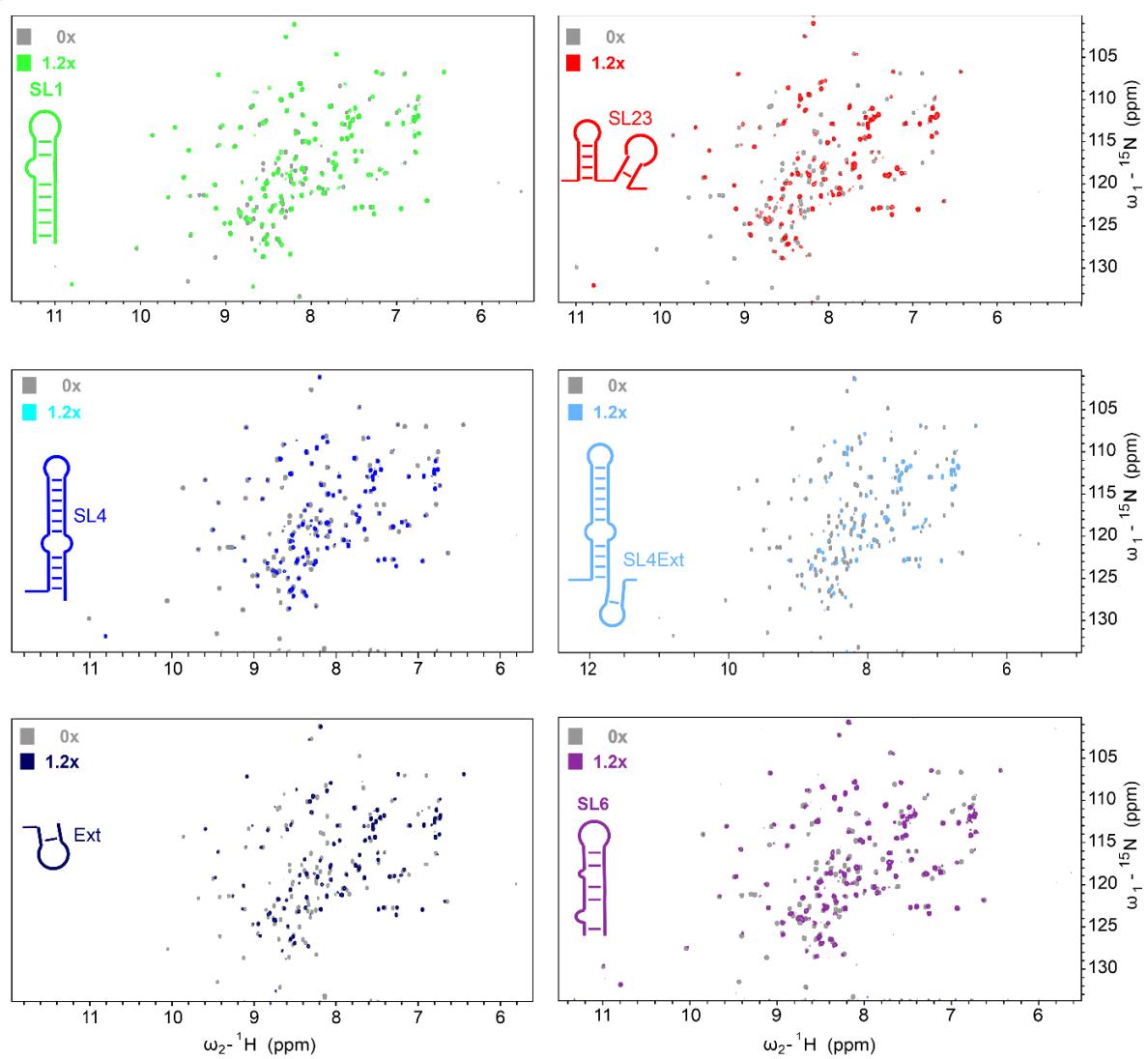
BMRB 34511 assigned construct: **GAMA** - NTD 44-180
 Herein used construct: **G** - NTD 44-180

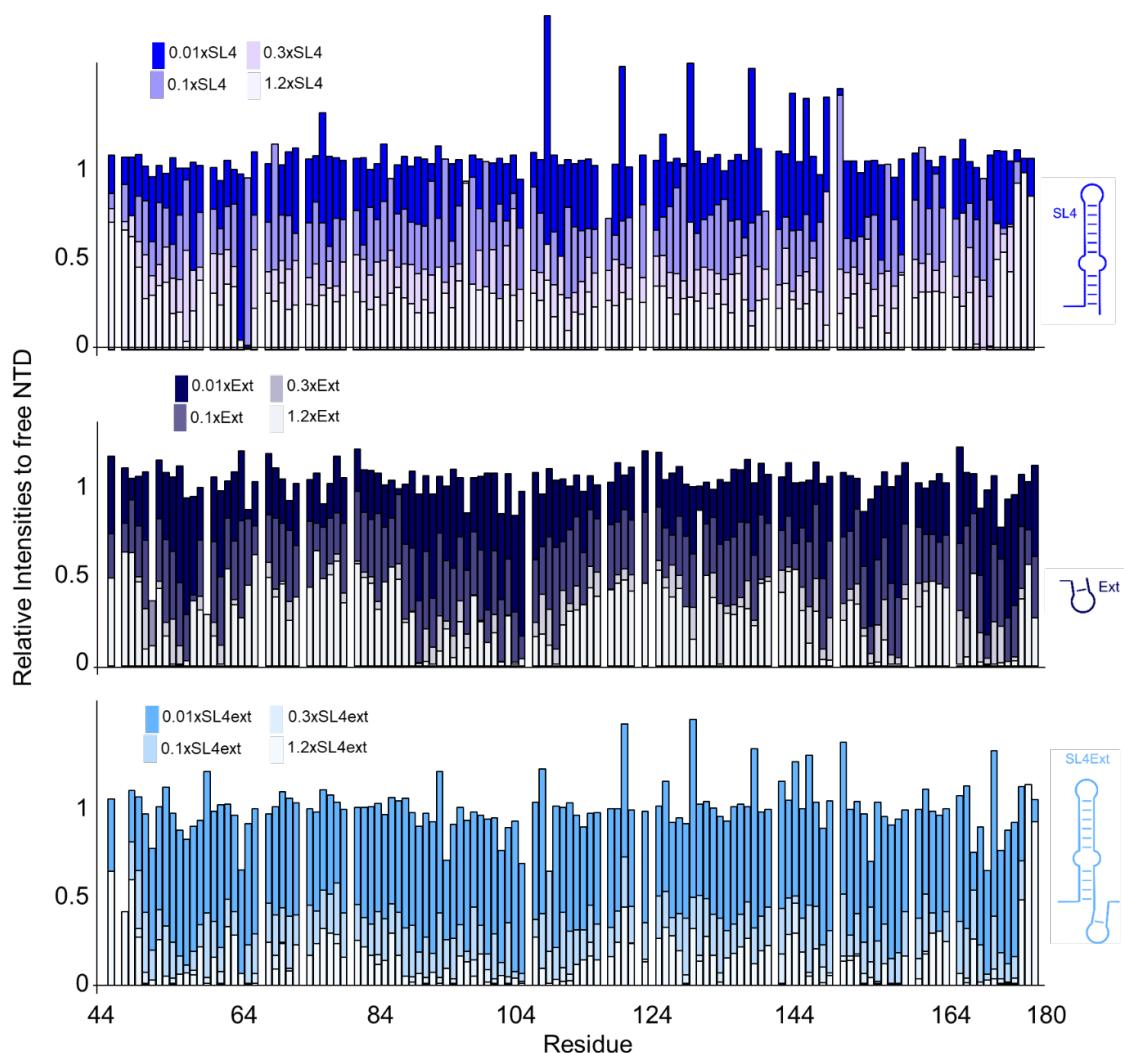


b



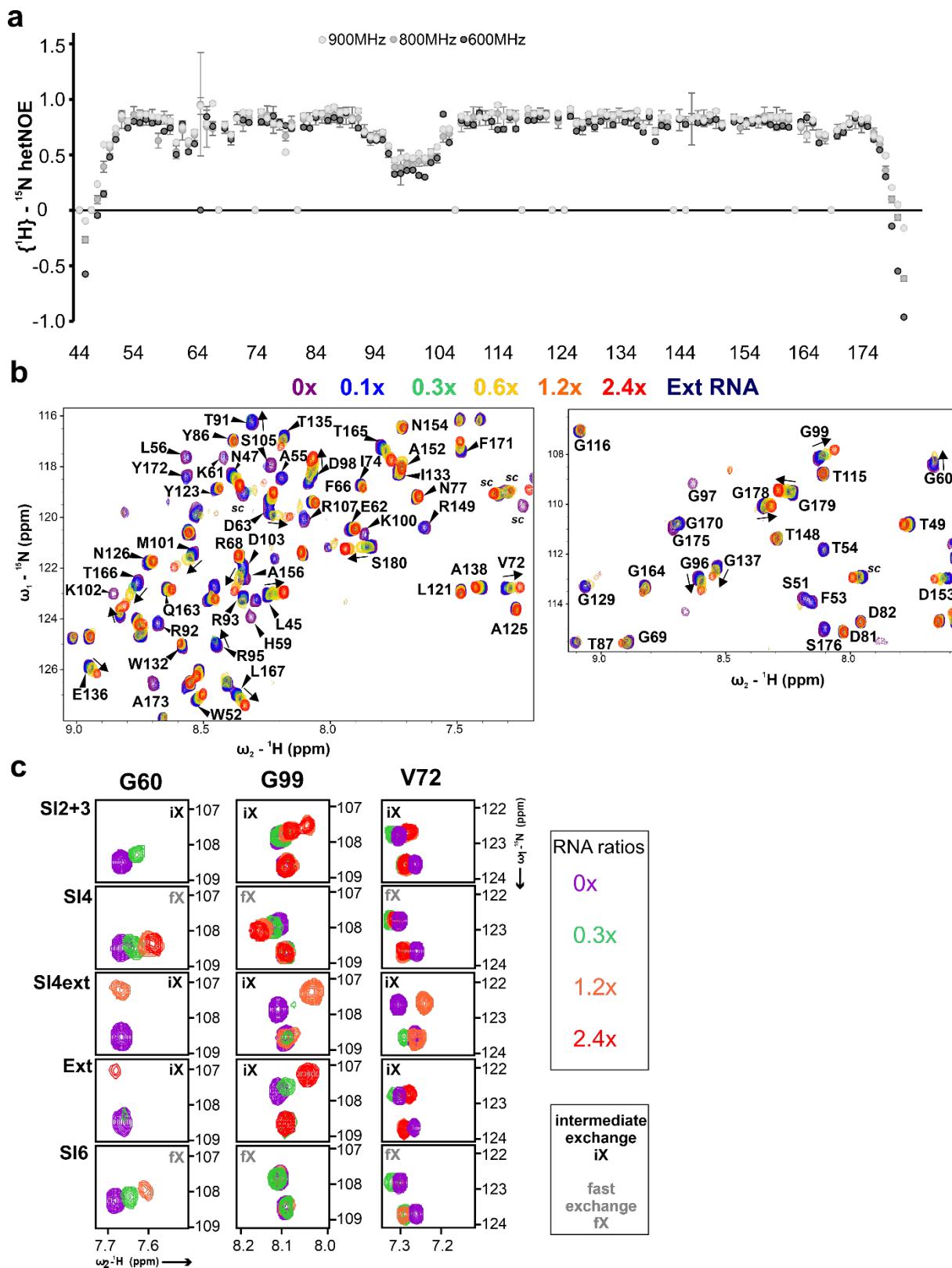
C



d

Supplementary Fig. 3. Titrations of NTD with SARS-CoV-2 RNAs. **a** Comparison of the NTD 44-180 (with N-terminal artificial G residue from cleavage) in the herein used NTD buffer compared to the BMRB entry 34511³, highlighting differences of NH-residues. Red circles represent NH-resonances where they are found in the BMRB for the NTD 44-180 (including 4 artificial residues GAMA upon cleavage). **b** Titration of the large, branched SL5 RNA to the NTD led to severe line broadening of protein peaks even at 0.1 molar equivalents. In line with the results from EMSAs (see Fig. 1) this suggests SL5 (157 nt, 48 kDa) is simultaneously bound by multiple NTD molecules leading to an immense increase of MW, making it invisible for analysis by solution NMR. Of note, visible peaks can be correlated to side chain NHs or the N- and C-terminal amino acids, that tumble independently from the huge RNP complexes. **c** Overlay of apo NTD ^1H - ^{15}N -HSQC spectra and its complexes with selected SARS-CoV-2 5'-geRNA elements as iconized showing the ratios as given in insets. **d** Intensity ratios obtained from titration points of NTD with the given equivalents of SL4, SL4ext and Ext RNA. Data were measured at 70 μM NTD concentration and afterwards apo peaks normalized to 100% (not included in plot). RNA titration points are shown for 0.01x, 0.1x, 0.3x and 1.2x SL4 stoichiometries. The comparison shows that line broadening of NTD resonances upon SL4 titration is evenly distributed as compared to titrations with SL4ext and Ext, respectively, where pronounced “dips” in signal intensity can be assigned to the first loop region (55-70) and the β -hairpin element (89-107).

Supplementary Figure 4, related to Fig. 3:

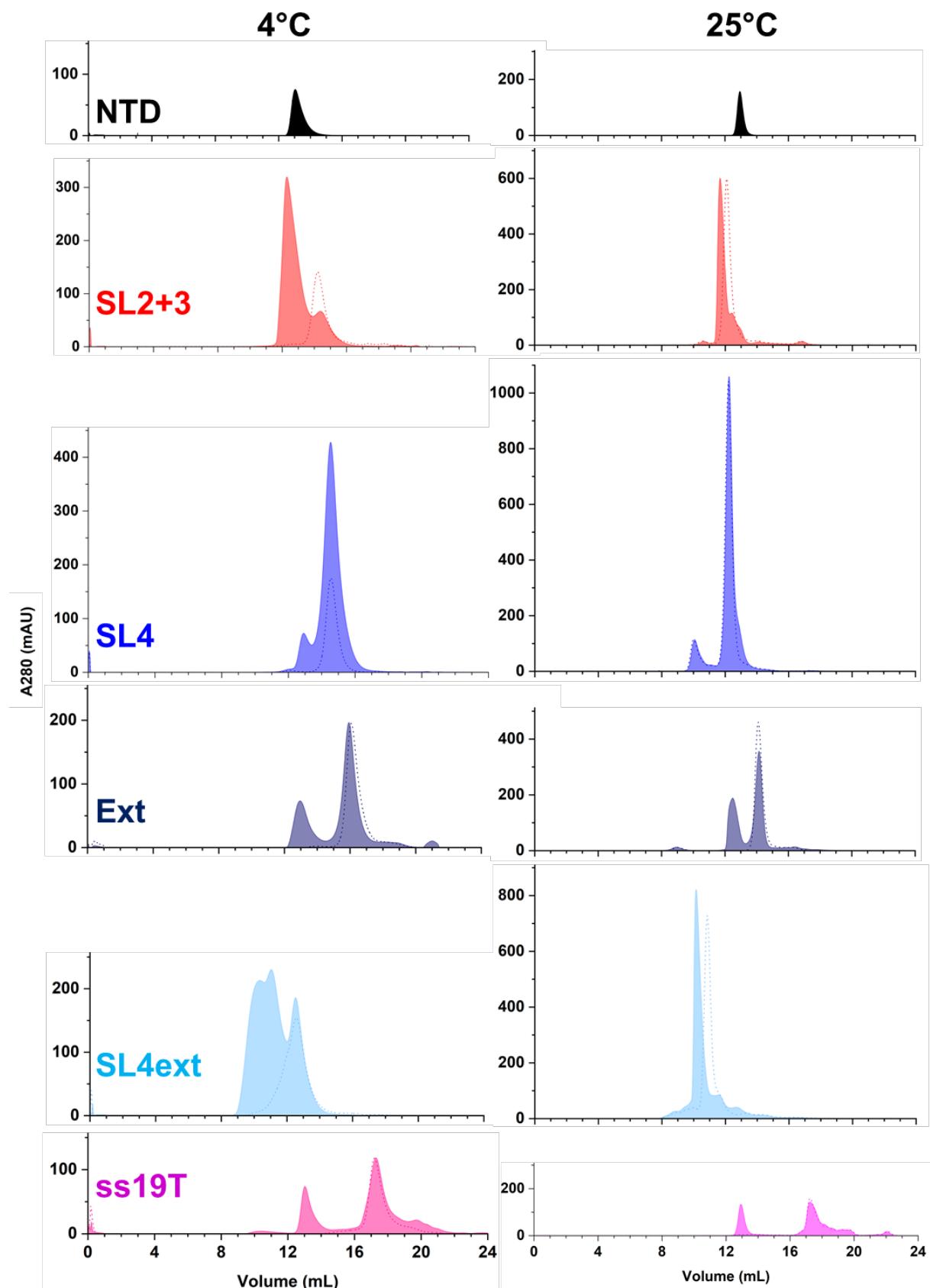


Supplementary Fig. 4. NTD dynamics in its apo and RNA-bound form. **a** ^1H - ^{15}N -hetNOE spectra recorded at proton Larmor frequencies of 600MHz, 800MHz and 900MHz, respectively. **b** Two representative zoom-ins of ^1H - ^{15}N -HSQC overlays showing NTD titrated

with Ext RNA at molar ratios depicted above. NTD amide resonance assignments are shown.

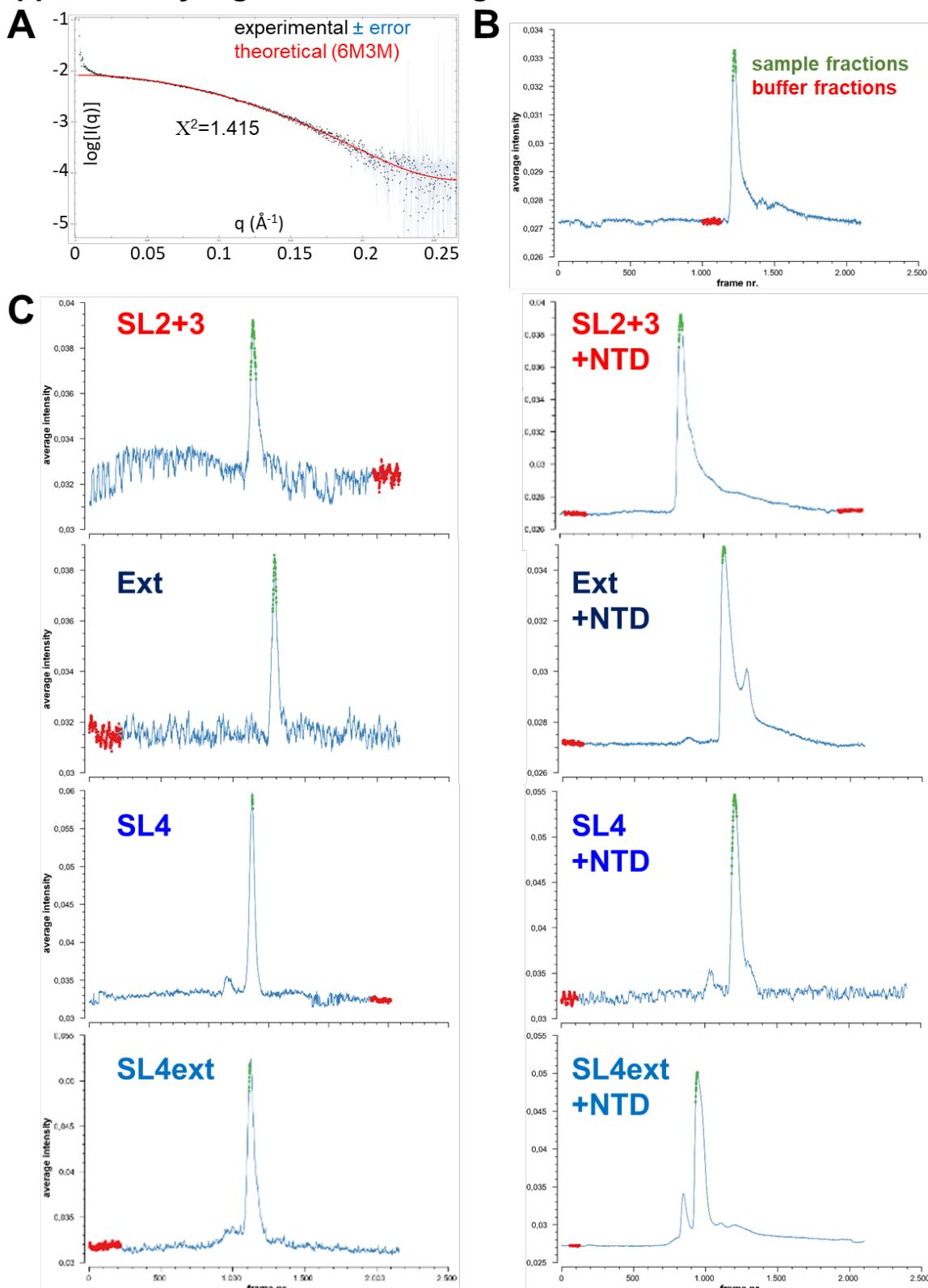
c Overview of protein NMR-observed exchange regimes between NTD and the 5'geRNA elements of this study. The zoom-ins show NTD amide groups in apo and RNA-bound forms with ratios as given. For the different RNAs, the spectral comparisons show non-identical exchange regimes as indicated by the respective peak patterns. The type of exchange regime is denoted, and intermediate exchange (iX) regimes indicate dissociation constants of approximately 1-10 μ M, while pure fast exchange (fX) is found for constants significantly above 10 μ M⁴.

Supplementary Figure 5, related to Fig. 3:



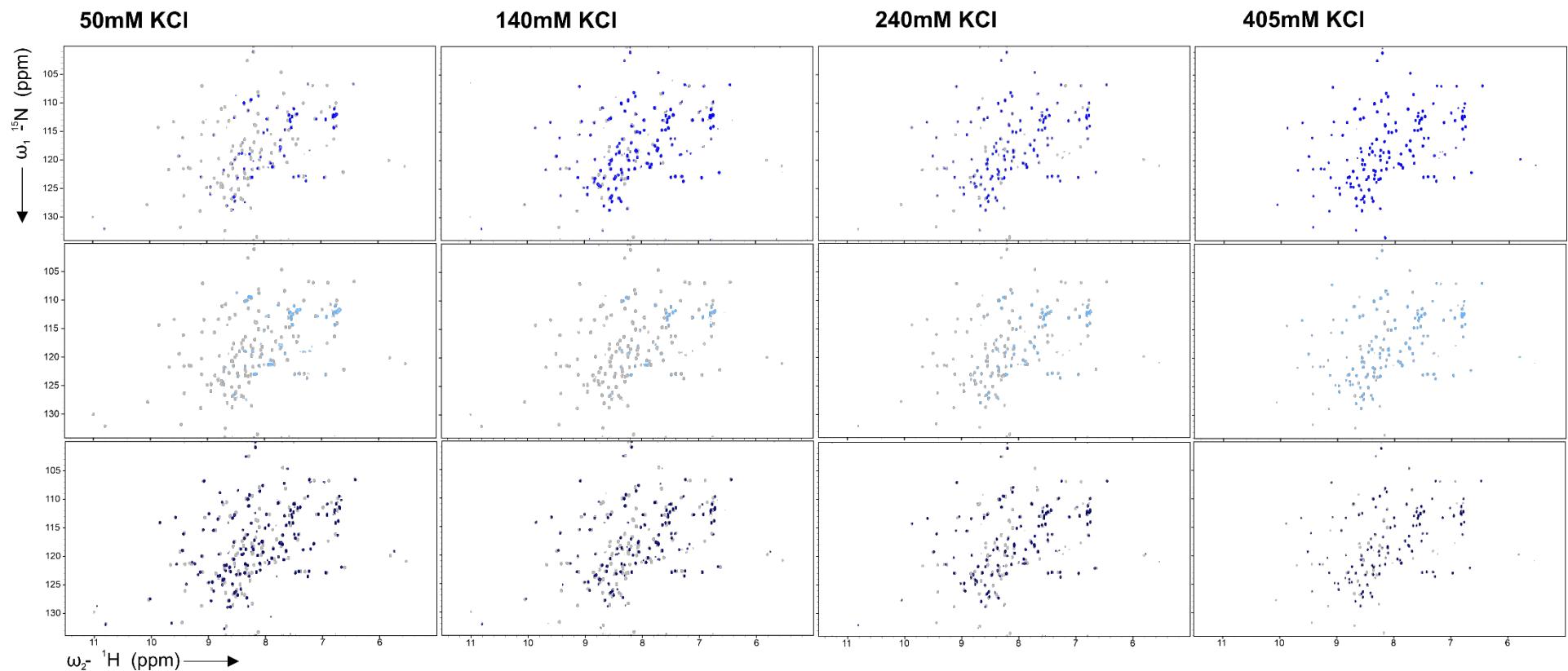
Supplementary Fig. 5. Comparison of analytical SEC runs at 4°C (left column) and RT (right) for NTD, selected RNAs of this study as shown (dotted lines, open curves) and their complexes with NTD (filled curves).

Supplementary Fig. 6, related to Fig. 4:



Supplementary Fig. 6. (SEC-)SAXS analysis of NTD and SARS-CoV-2 5'geRNAs and their complexes. **a** SAXS raw scattering curve of NTD after buffer subtraction (black) overlaid with a theoretical scattering curve (red) derived from PDB entry 6M3M⁵. The fit quality is given by the X^2 value. **b** SAXS profile of a SEC run for the NTD. The retention volume is 25mL spread over 2500 frames. **c** The same as in b for RNAs as given, either alone (left) or in complex with NTD (right). For b and c, green frames indicate fractions used for downstream analyses of the sample, while red frames were used for internal buffer subtraction.

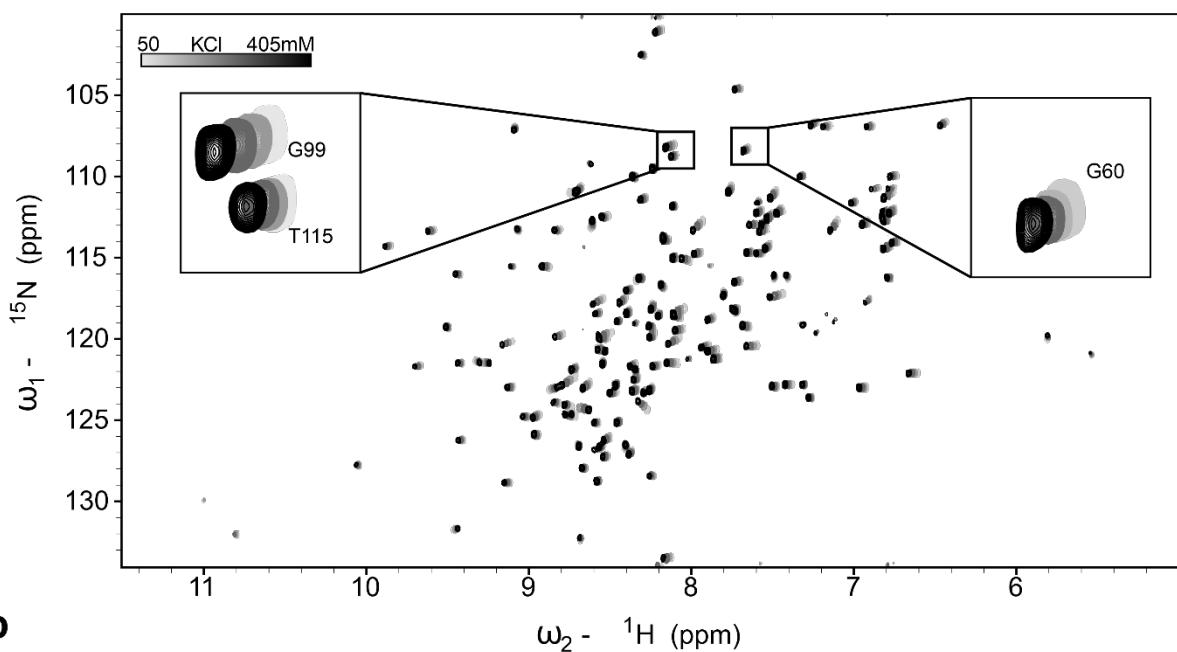
Supplementary Fig. 7, related to Fig. 5:



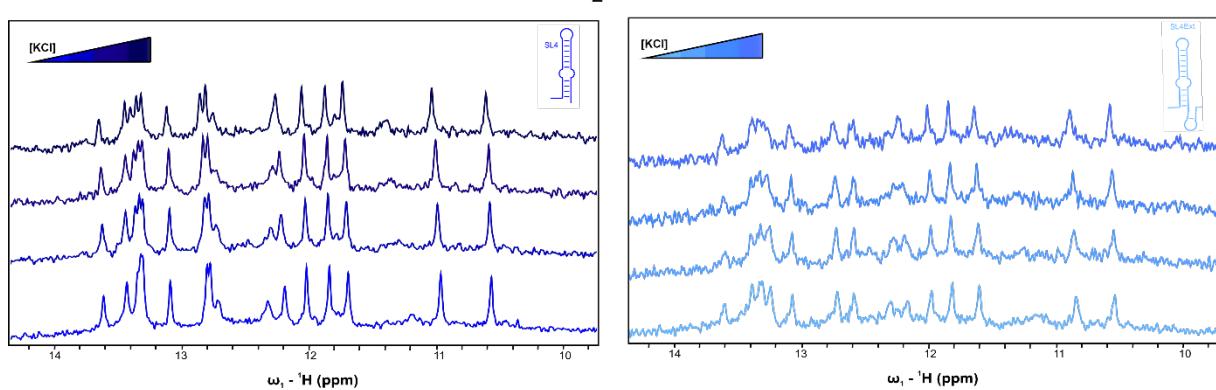
Supplementary Fig. 7: HSQC-based titrations of 5'geRNA elements to NTD with salt. ^1H - ^{15}N -HSQC overlays comparing complexes of NTD with 1.2x molar excess of SL4 (blue), SL4ext (light blue) and Ext (dark blue), respectively, on top of 60 μM apo NTD (light grey). Increase in salt concentration was achieved by adding respective volumes from a 3M KCl stock solution, with subsequent incubation for four hours to allow samples to reach equilibrium.

Supplementary Fig. 8, related to Fig. 5:

a

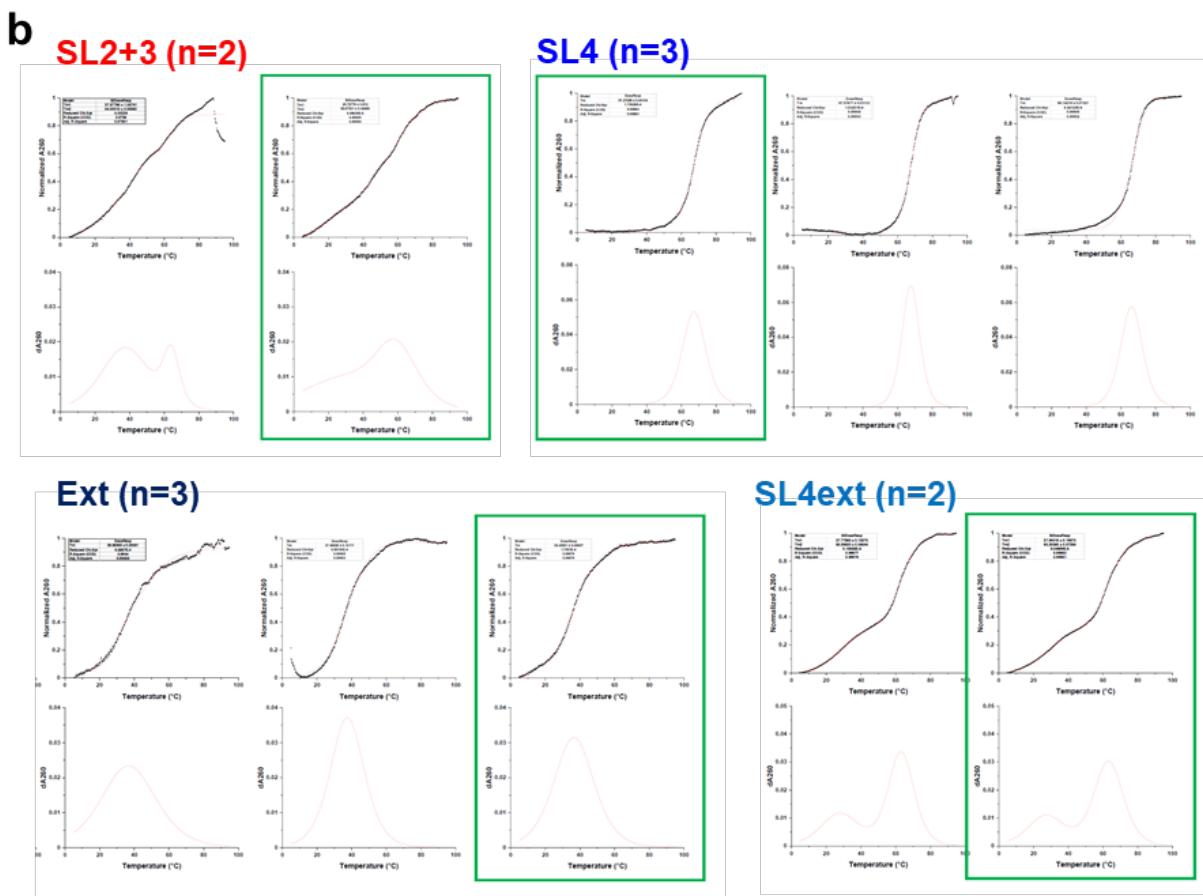
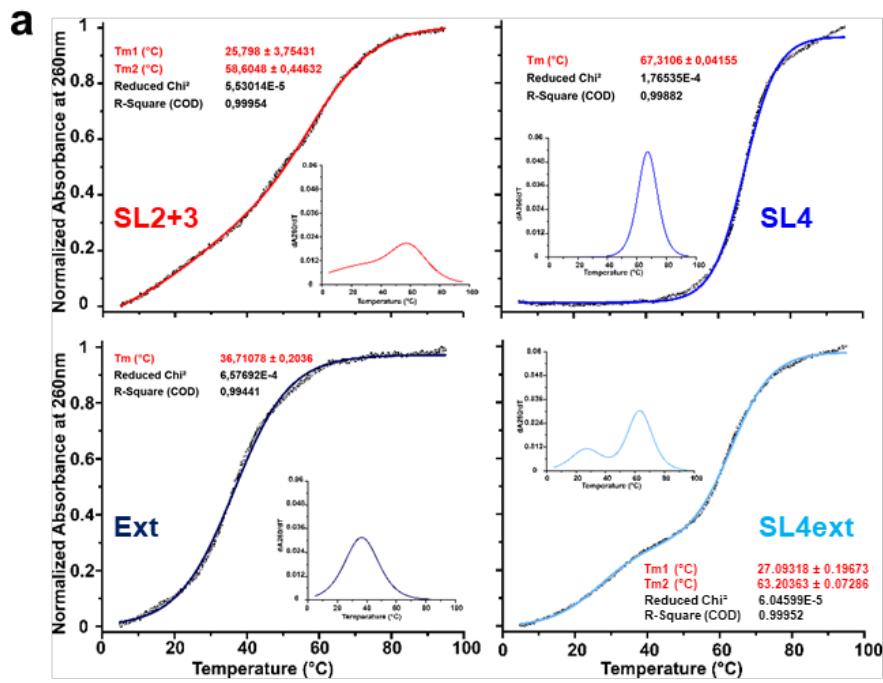


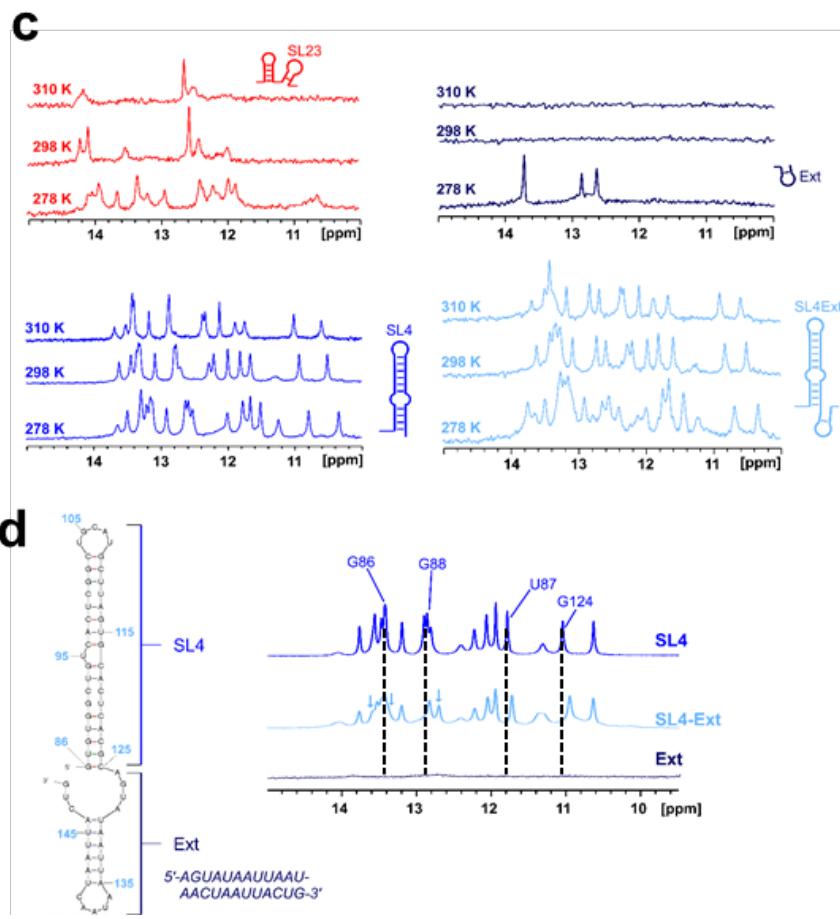
b



Supplementary Fig. 8. Effect of KCl concentration on NTD and RNAs alone. **a** ^1H - ^{15}N -HSQC overlays of free NTD (60 μM) showing spectra at increasing concentrations of salt: 50 mM KCl (light grey), 140 mM KCl (medium grey), 240 mM KCl (dark grey) and 405 mM KCl (black). **b** Imino- ^1H jump-return spectra of SL4 and SL4ext at increasing salt concentrations. Left, SL4 (40 μM) with 50 mM KCl (blue), 140 mM KCl (red), 240 mM KCl (green) and 405 mM KCl (purple). Right, SL4ext (20 μM) with 50 mM KCl (blue), 140 mM KCl (red), 240 mM KCl (green) and 405 mM KCl (purple).

Supplementary Fig. 9, related to Fig. 6:



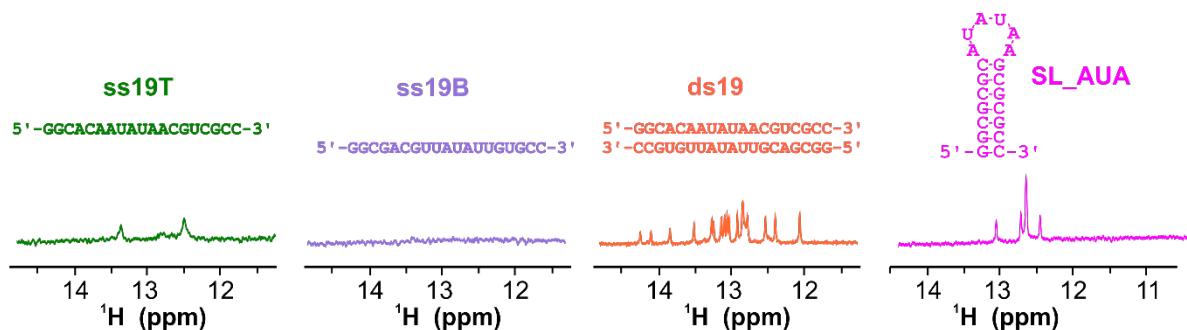


Supplementary Fig. 9. Stability of RNA secondary structures in the SL2+3 and SL4-hubs.

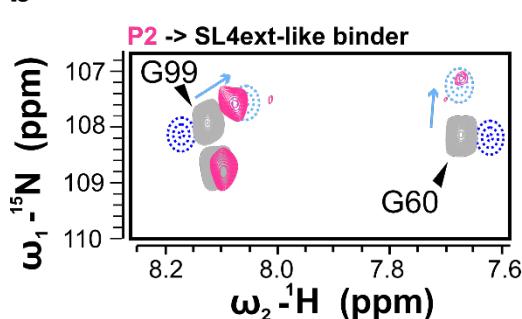
a Determination of melting points for RNA elements as shown via normalized absorbance and fits at 260nm in a CD spectrometer. Curves of first derivatives as used in **Fig. 6** are shown as insets. Fit qualities are included with reduced Chi-2 and R² values. All measurements as used for values in **Fig. 6** are summarized in panel b. **b** Summary of all CD measurements for RNAs in a with replicates as given. Panels show 260-nm absorption in dependence of temperature (top) and first derivates (bottom), respectively. Green-boxed pairs represent the runs used for panel a and **Fig. 6**. **c** Comparison of Imino-¹H-1D spectra recorded from Ext and SL3 RNAs at 278K, 298K and 310K, respectively. **d** Imino-¹H-1D spectra recorded from SL4 and SL4ext. Assignments within SL4^{2,6} are used to highlight CSP between RNAs and to show that the Ext region attached to SL4 does not interfere with SL4 per se, while slight shifts of peaks map to base-pairs close to the site of SL4-Ext transition. **e** ITC-derived binding of NTD to SL4ext. Top, representative thermogramm and integrated data fit using a two-site binding model with transitions at approximately 0.5 and 1.0 molar ratios and K_D values as given, i.e. 7 μM for the first (Ext) binding and 52 μM for the second (SL4) binding. Bottom, three additional ITC experimental replicates used for determining standard deviations of K_D.

Supplementary Fig. 10, related to Fig. 7:

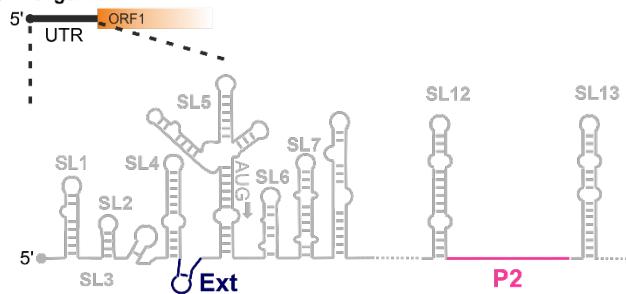
a



b



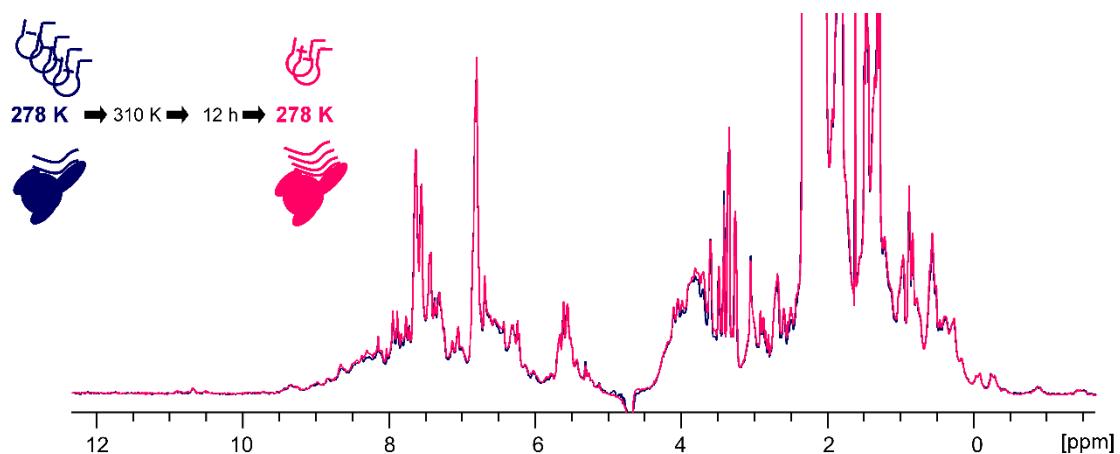
c



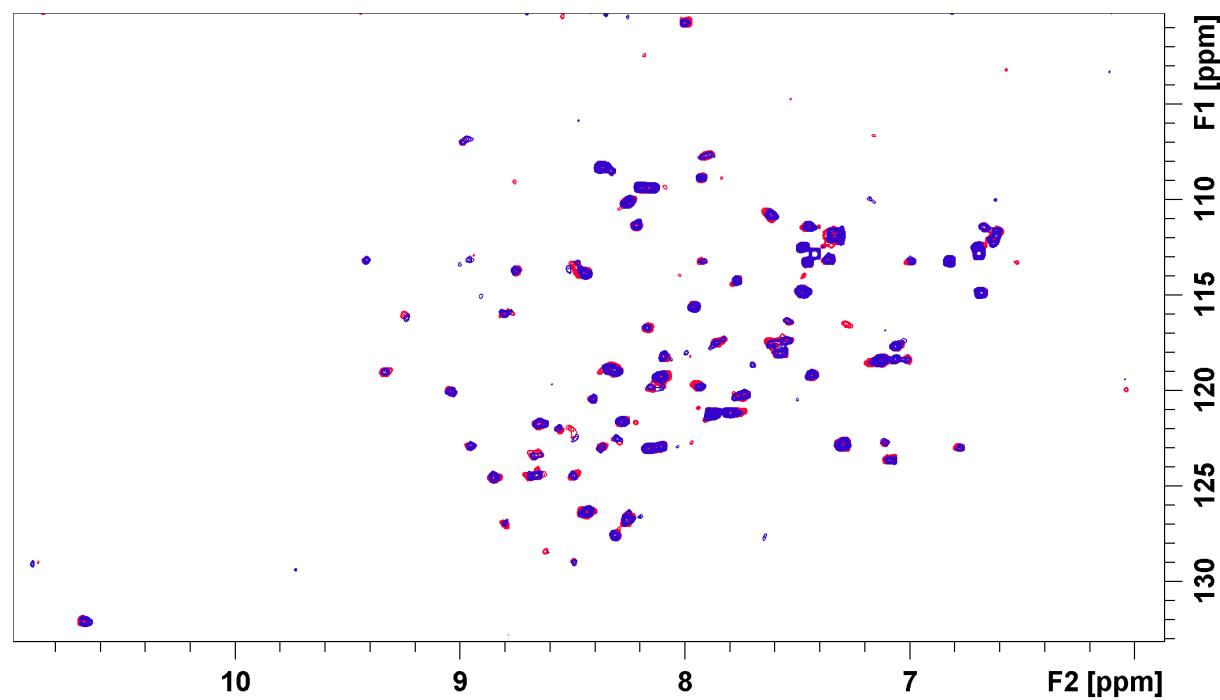
Supplementary Fig. 10: **a** ^1H -1D spectra of non-viral RNAs used in this study. **b** Spectral excerpt of ^1H - ^{15}N -HSQC spectrum overlaying free NTD (grey) with viral RNA element P2, including reporters G99 and G60. For comparison of reference “shift modes” chemical shifts of 1.2x SL4 (dotted blue lines) and SL4ext (dotted light blue lines) are schematized, respectively. **c** Schematic depiction of genomic SCoV-2 RNA (nt 1-822). Two principal N-protein binding sites (as predicted by SHAPE coupled to mutational profiling)⁷ are indicated with respective colors.

Supplementary Fig. 11, related to Fig. 8:

a

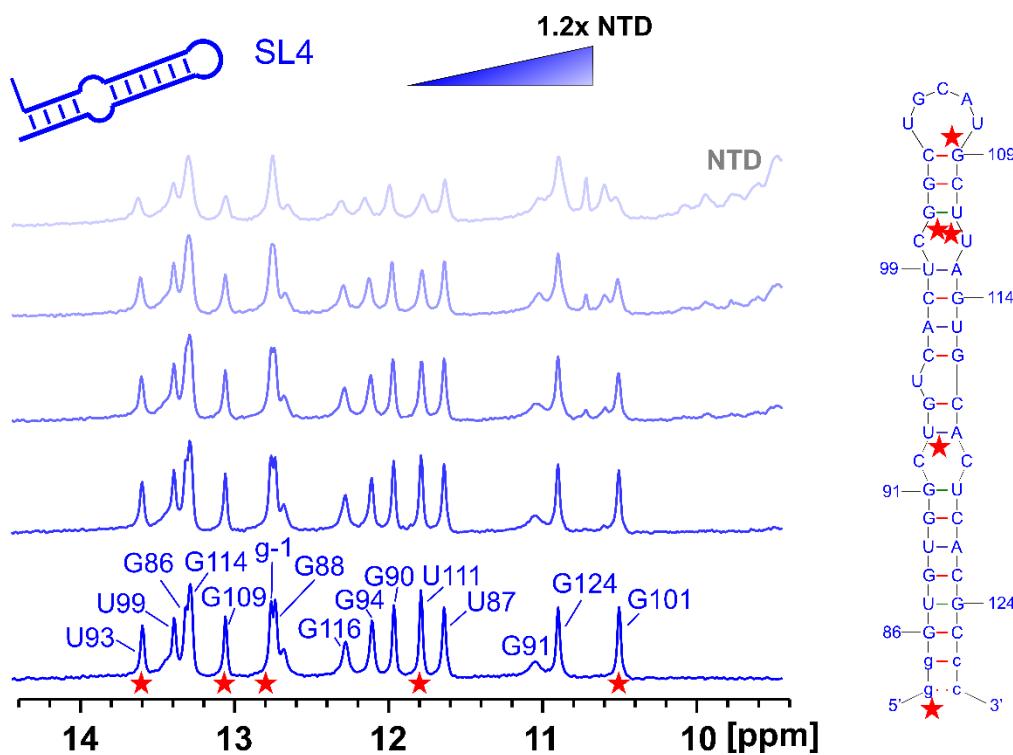


b



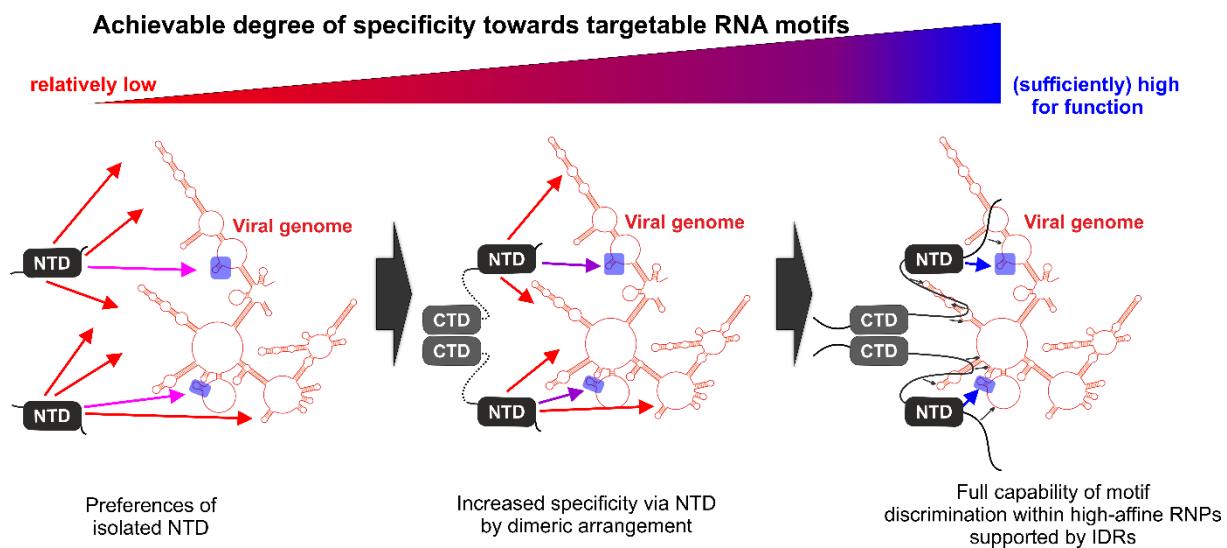
Supplementary Fig. 11: Verification of model in main text Fig. 8c. **a** Overlay of ¹H-1D spectra of NTD in complex with Ext RNA acquired at 278K before and after the temperature cycling. The spectra are paralleled with the imino proton spectra shown in panel d of main text Fig. 8. Spectra of the identical complex sample with NTD have been recorded while undergoing the temperature route as given in Fig. 8c and show that the overall sample state is unaltered. **b** Full ¹H-¹⁵N-HSQC spectral view showing the overlay underlying the spectral excerpts in main text Fig. 8e. For convenience the first spectrum (before cycling, dark blue) is shown on top of the re-measured spectrum (after cycling, red).

Supplementary Fig. 12:



Supplementary Fig. 12: Mapping of NTD-binding sites on SL4. Display of imino-¹H-spectra of SL4 alone (bottom) and at increasing stoichiometric ratios of NTD. Right, nucleotides whose iminos are visibly affected in CS and/or line-width (asterisks) are marked on the secondary structure model of SL4. The mapping supports the weak binding of SL4 to the NTD. Despite its stable and robust structure, SL4 imino peaks^{2,6} adjacent to bulges or loops are found affected by line-broadening in contrast to those of base paired regions. We conclude that even in the stable SL4 context, the NTD is capable of recognizing regions with ssRNA character, which again emphasizes NTD's ability to precisely differentiate between available RNA motifs.

Supplementary Fig. 13:



Supplementary Fig. 13: Hypothetical model of N specificity as hidden in multi-modular interactions of NTDs in cluster-RNPs. Specificity of N is displayed as a measurable distinctness between all potentially targetable RNA motifs in viral genomic RNA. From left to right, in an initial scenario N faces a large excess of heterogeneous RNA elements and its capability of distinguishing between them - using its NTD - will only result in low distinguishability, as e.g. in the preferred recognition of sites like Ext (here indicated in blue) over less relevant sites like SL1 or SL6 (all others). Increased distinctness will be achieved via dimers/oligomers of N, including e.g. the cooperative recognition of preferred, close-by sites. Also, a general increase in affinity within full-length N versus isolated NTDs is achieved via IDR1 and IDR2, which might proportionally increase the distinctness between available motifs to engage with.

Supplementary References

- 1 Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429-3431, doi:10.1093/nar/gkg599 (2003).
- 2 Wacker, A. et al. Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res*, doi:10.1093/nar/gkaa1013 (2020).
- 3 Dinesh, D. C. et al. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* **16**, e1009100 (2020).
- 4 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* **73**, 1-16, doi:10.1016/j.pnmrs.2013.02.001 (2013).
- 5 Kang, S. S. et al. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B* **10**, 1228-1238, doi:10.1016/j.apsb.2020.04.009 (2020).
- 6 Vogelee, J. et al. (1)H, (13)C, (15)N and (31)P chemical shift assignment for stem-loop 4 from the 5'-UTR of SARS-CoV-2. *Biomol NMR Assign*, doi:10.1007/s12104-021-10026-7 (2021).
- 7 Iserman, C. et al. Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell.* **80**, 1078-1091 (2020).