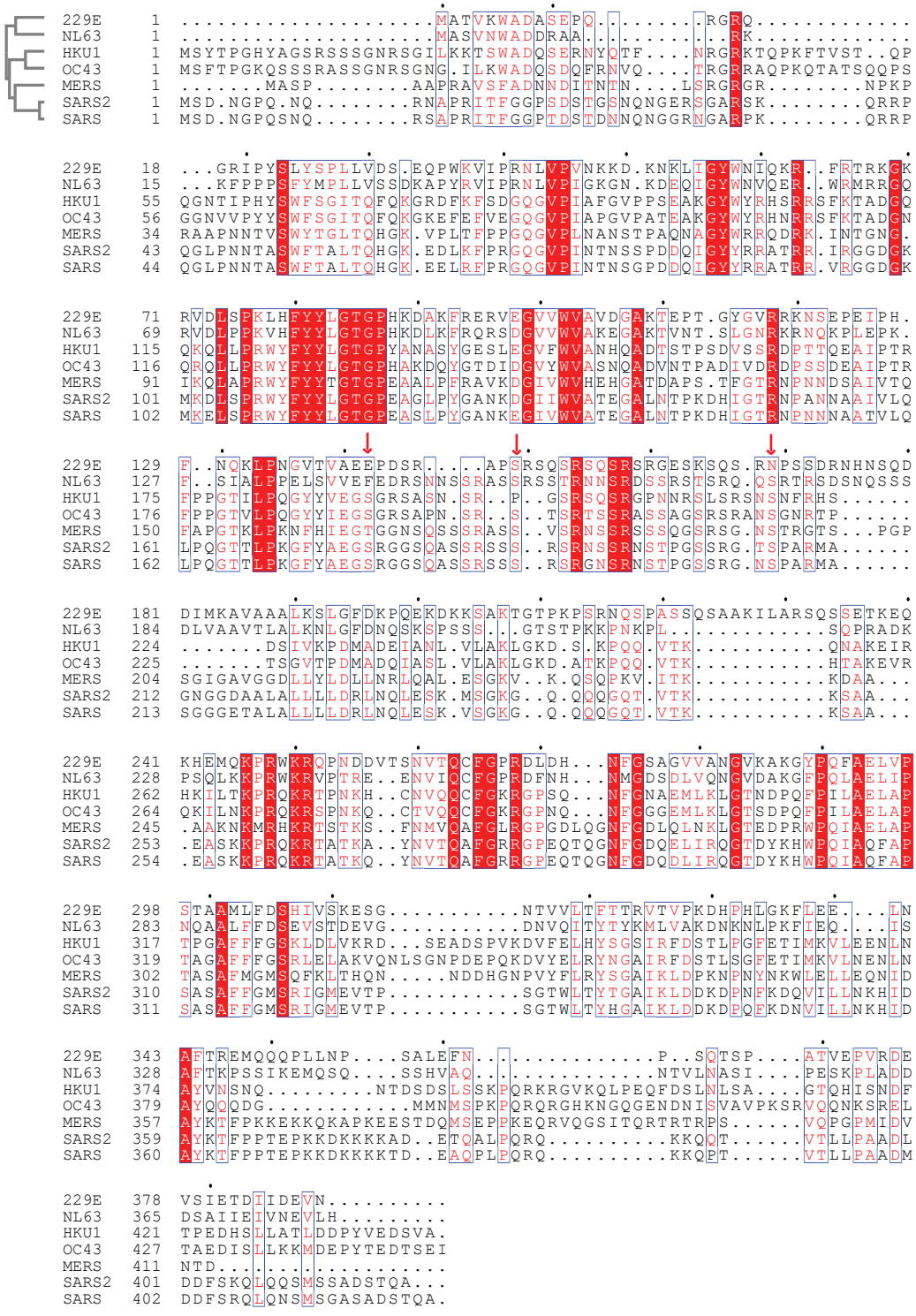
iScience, Volume *24*

Supplemental information

Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain

Chao Wu, Abraham J. Qavi, Asmaa Hachim, Niloufar Kavian, Aidan R. Cole, Austin B. Moyle, Nicole D. Wagner, Joyce Sweeney-Gibbons, Henry W. Rohrs, Michael L. Gross, J. S. Malik Peiris, Christopher F. Basler, Christopher W. Farnsworth, Sophie A. Valkenburg, Gaya K. Amarasinghe, and Daisy W. Leung

A



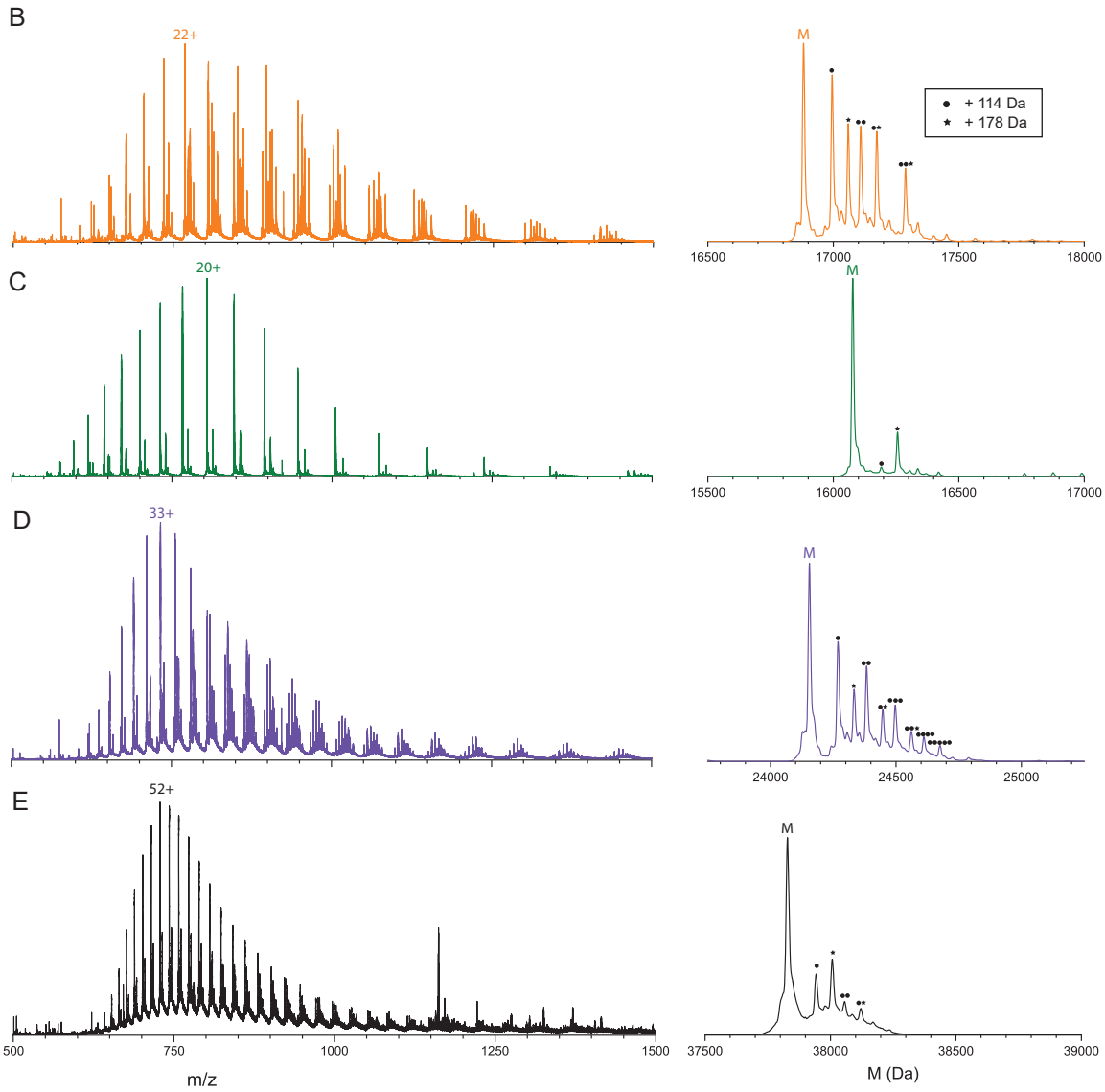
B

	229E	NL63	HK1	OC43	MERS	SARS
FL	26	28	34	36	49	90
Narm	12	0	17	29	22	84
NTD	32	30	44	43	60	92
LKR	26	36	31	34	46	89
CTD	30	28	36	37	54	96
Carm	10	21	15	18	14	74

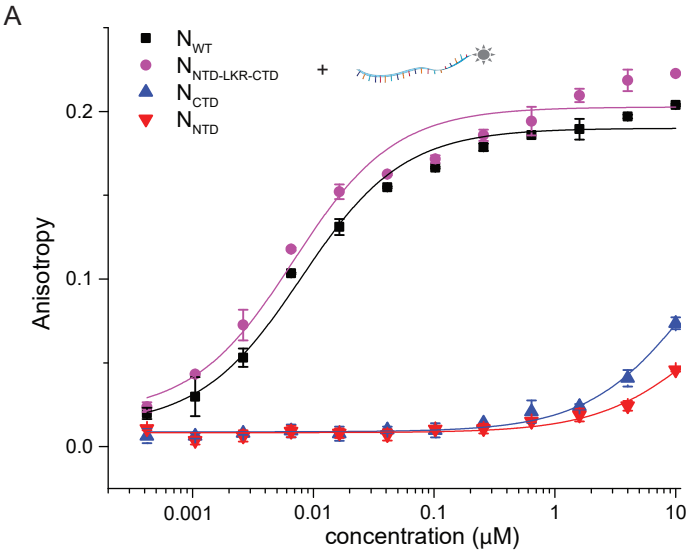
Supplementary Figure 1, related to Figure 1. Multiple sequence alignment of coronavirus nucleocapsids. A. Multiple sequence alignment of coronavirus nucleocapsids. Sequences were aligned using Clustal Omega. Accession numbers used are 229E (APT69891.1), NL63 (YP_003771.1), HK1 (AAT98585.1), OC43 (AAR01019.1), MERS (AKL80590.1), SARS (AAP30037.1), SARS2 (YP_009724397.2). Alignments were analyzed using ESPript3. The three serines (176, 188, and 206) are labeled with red arrows. **B**. Sequence identity between SARS-CoV-2 N and that of common cold coronaviruses and MERS and SARS. FL, full length. All units are in %. Percent identity matrixes for corresponding domains of N are generated using Clustal2.1.

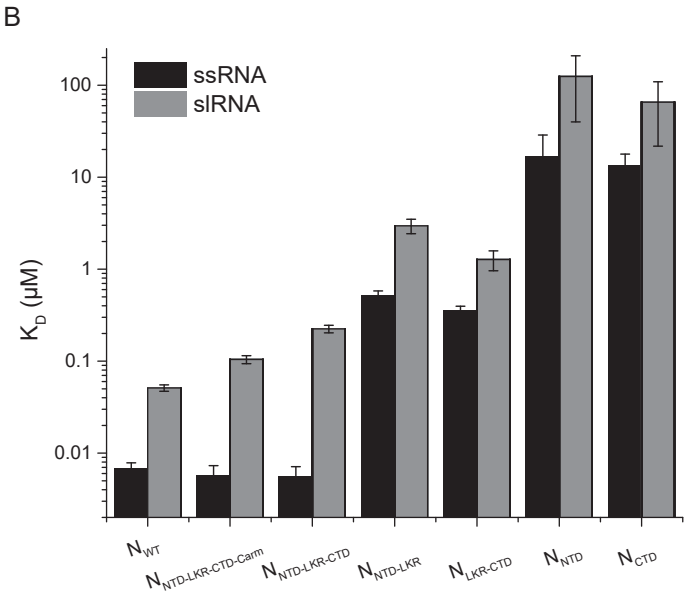
A

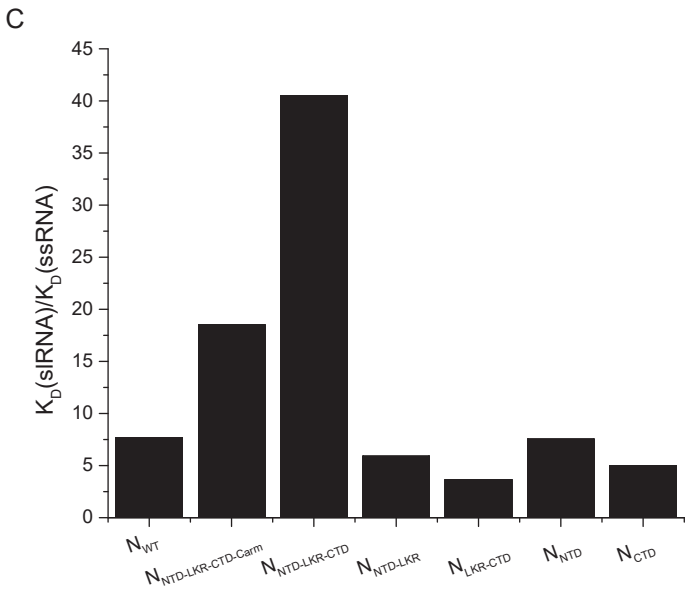
	MW (kDa)	polydispersity (%)	
		peak 1	peak 2
N_{WT}	46	20 ± 1	40 ± 20
$N_{NTD-LKR-CTD}$	35	17 ± 7	30 ± 20
$N_{NTD-LKR}$	22	23 ± 3	
N_{NTD}	14	7 ± 2	
N_{CTD}	14	15 ± 1	



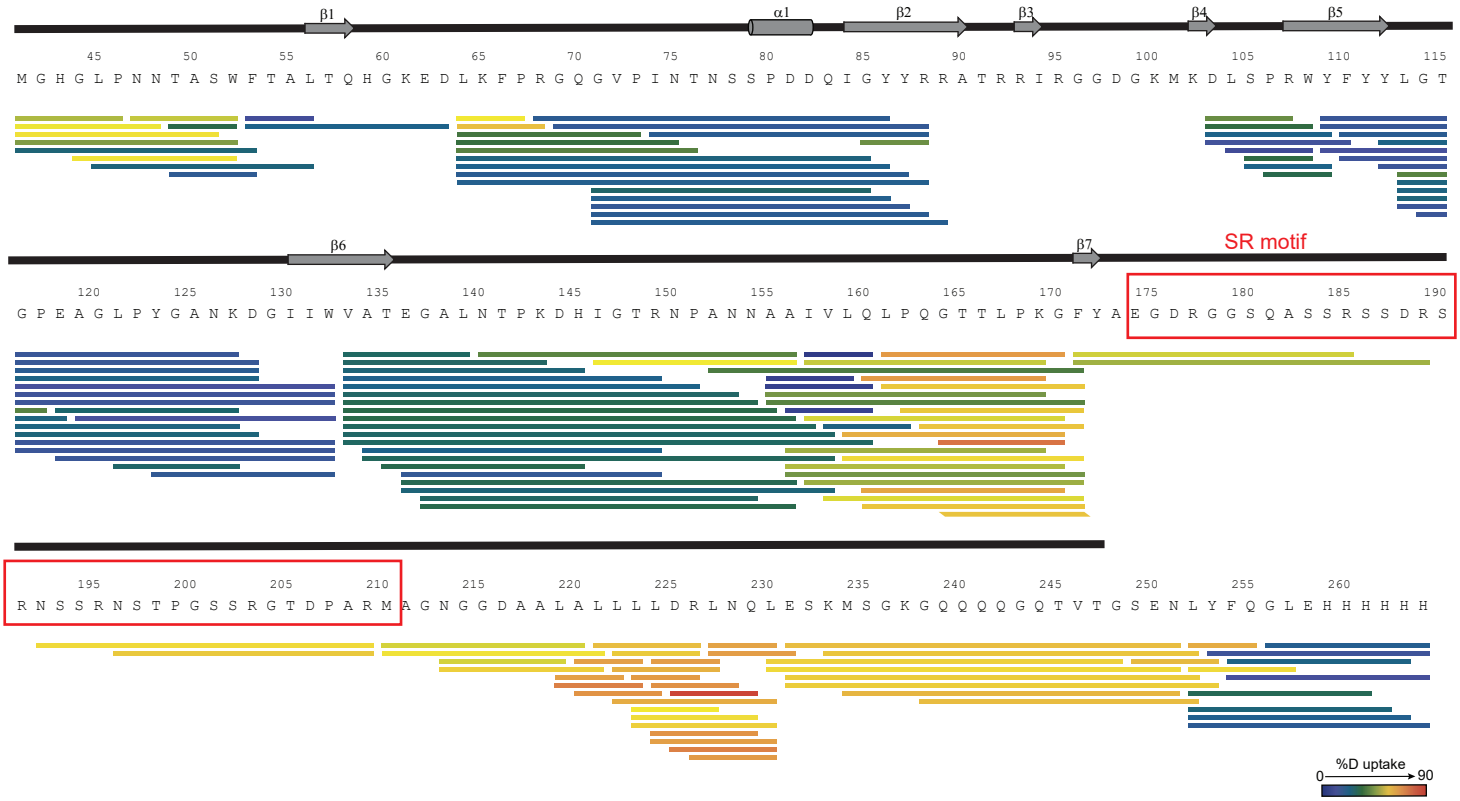
Supplementary Figure 2, related to Figure 1. Denaturing mass spectra of N protein truncations N_{NTD}, N_{CTD}, $N_{NTD-LKR}$, and $N_{NTD-LKR-CTD}$. A. DLS polydispersity table for N constructs. Higher values indicate broader size distributions. Numbers are reported as average and standard deviation of three experiments. Deconvolution yields experimental M values of 16,881 Da, 16,078 Da, 24,155 Da, and 37,829 Da for (**B**) N_{NTD}, (**C**) N_{CTD}, (**D**) $N_{NTD-LKR}$, and (**E**) $N_{NTD-LKR-CTD}$ respectively, matching theoretical values within 1 Da, based on protein sequence. Deconvoluted mass spectra (right) and adduct series corresponding to pervasive trifluoroacetic adducts (delta mass 114 Da, circle) and a-N-gluconoylation (delta mass 178 Da, star). TFA adducts are introduced by the ion pairing reagent in solvent, while α-N-gluconoylation is a common modification occurring on His-tagged proteins. Native spray of NTD (not pictured) yielded no peaks with delta mass 114 Da, but retained a single delta mass 178 Da, confirming transient TFA adducts are an artifact of the denaturing experiment, but the α-N-gluconoylation of the His-tag is covalent.


A
N_WT
N_NTD-LKR-CTD
N_CTD
N_NTD
+
Anisotropy
0.0
0.1
0.2
0.001
0.01
0.1
1
10
concentration (μM)

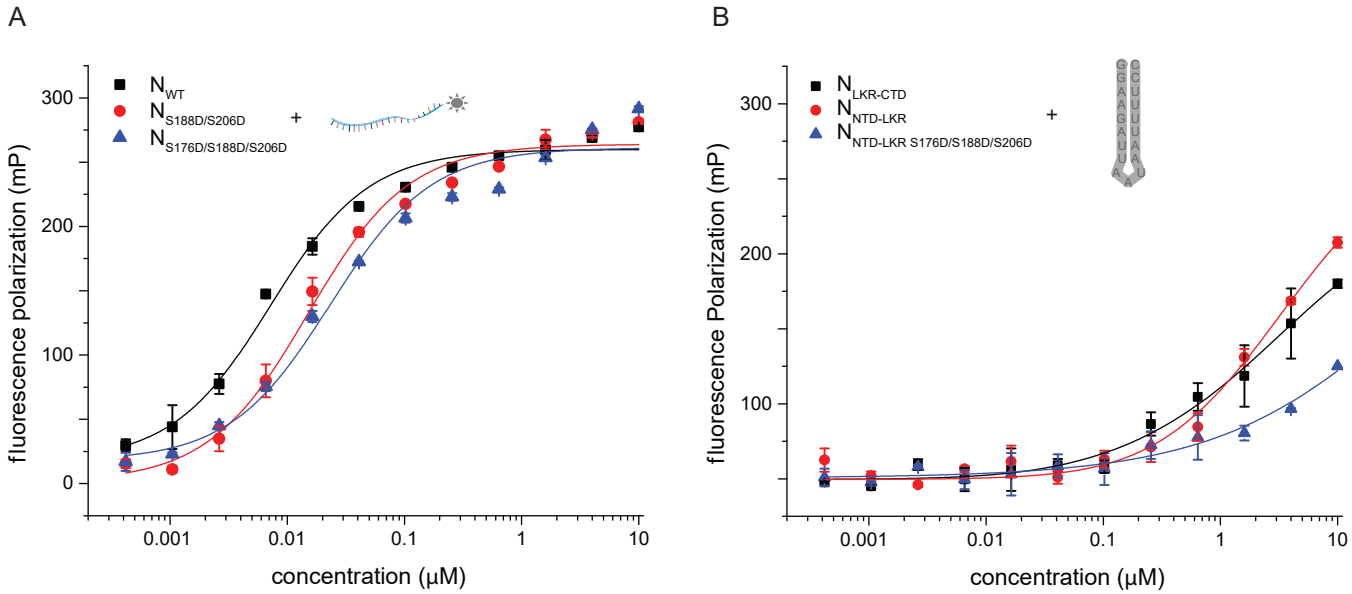

B
ssRNA
slRNA
K_D (μM)
100
10
1
0.1
0.01
N_WT
N_NTD-LKR-CTD-Carm
N_NTD-LKR-CTD
N_NTD-LKR
N_LKR-CTD
N_NTD
N_CTD


C
K_D(slRNA)/K_D(ssRNA)
0
5
10
15
20
25
30
35
40
45
N_WT
N_NTD-LKR-CTD-Carm
N_NTD-LKR-CTD
N_NTD-LKR
N_LKR-CTD
N_NTD
N_CTD

Supplementary Figure 3, related to Figure 2. Nucleocapsid binds stem-loop RNA with reduced affinity. A. Fluorescence anisotropy binding curves of N constructs to a 20-nt ssRNA. Anisotropy values were converted from polarization according to previous research (Kozlov et al., 2012). The fitted K_D values are 0.007 ± 0.001 μM (N_{WT}, black square), 0.006 ± 0.002 μM ($N_{NTD\text{-}LKR\text{-}CTD}$, magenta circle), 14 ± 5 μM (N_{CTD}, blue up triangle) and 18 ± 14 μM (N_{NTD}, red down triangle). These values are very close to those of polarization. In this system, binding monitored by anisotropy is similar to that of polarization. **B.** Fitted K_D values for N constructs binding to ssRNA (black) and slRNA (grey). **C.** Ratio of K_D of slRNA over that of ssRNA for N constructs. The reduced binding to slRNA is around 5-fold for most N constructs. The reduction is higher for those of $N_{NTD\text{-}LKR\text{-}CTD\text{-}Carm}$ and $N_{NTD\text{-}LKR\text{-}CTD}$, suggesting Narm and Carm are more involved in slRNA binding.


β1
α1
β2
β3
β4
β5
β6
β7
SR motif
41 M G H G L P N N T A S W F T A L T Q H G K E D L K F P R G Q G V P I N T N S S P D D Q I G Y Y R R A T R R I R G G D G K M K D L S P R W Y F Y Y L G T
116 G P E A G L P Y G A N K D G I I W V A T E G A L N T P K D H I G T R N P A N N A A I V L Q L P Q G T T L P K G F Y A E G D R G G S Q A S S R S S D R S
191 R N S S R N S T P G S S R G T D P A R M A G N G G D A A L A L L L L D R L N Q L E S K M S G K G Q Q Q Q G Q T V T G S E N L Y F Q G L E H H H H H H
%D uptake
0
90

Supplementary Figure 4, related to Figure 3. Sequence coverage of $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$ in HDX-MS and HDX of the unbound state. Protein coverage map of unbound state $N_{NTD\text{-}LKR\ S176D/S188D/S206D}$ HDX yielding 152 peptides with 93.3% sequence coverage. Peptide bars are colored according to their average %HDX relative to the color bar, where cooler colors depict low average %HDX and warmer colors depict high average %HDX. The secondary structure reported by PDB 6M3M is shown above the sequence. Overall, the HDX of the unbound state is largely consistent with the reported secondary structure and a well-ordered tertiary structure; regions outside of the reported structure undergo relatively rapid HDX, consistent with a lack of backbone hydrogen bonding. Interestingly, despite a lack of reported secondary structure in the region of 155-160, relatively low HDX was observed, consistent with either hydrogen bonding of secondary/tertiary structure or a hydrophobic pocket. SR-motif in LKR are boxed in red.


A
B
N WT
N S188D/S206D
N S176D/S188D/S206D
fluorescence polarization (mP)
concentration (μM)
N LKR-CTD
N NTD-LKR
N NTD-LKR S176D/S188D/S206D
fluorescence Polarization (mP)
concentration (μM)

Supplementary Figure 5, related to Figure 4. Phosphomimics of N reduce RNA binding. A. Fluorescence polarization binding curves of N constructs to a 20-nt ssRNA. The fitted K_D values are 0.007 ± 0.001 μM for N_{WT}, 0.015 ± 0.002 μM for $N_{S188D/S206D}$, and 0.023 ± 0.006 μM for $N_{S176D/S188D/S188D}$. **B.** Fluorescence polarization binding curves of N constructs to a 19-nt slRNA. The fitted K_D values are 1.3 ± 0.3 μM ($N_{LKR-CTD}$, black square), 3.0 ± 0.5 μM ($N_{NTD-LKR}$, red circle), and 2.9 ± 1.4 μM ($N_{NTD-LKR\ S176D/S188D/S206D}$, blue up triangle).