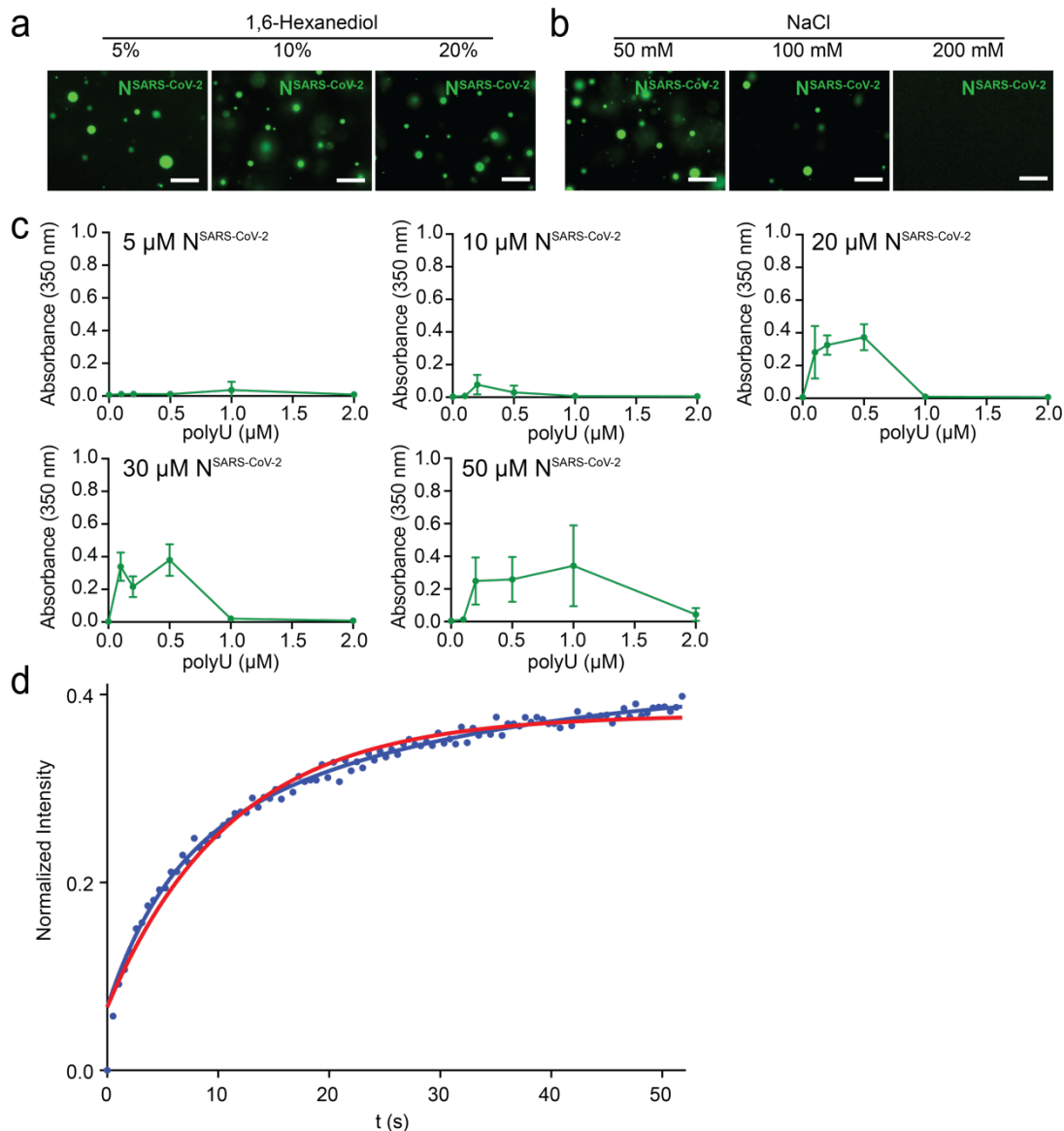


Supplementary Information for

Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates

Adriana Savastano, Alain Ibáñez de Opakua, Marija Rankovic, and Markus Zweckstetter



Supplementary Fig. 1. RNA-induced LLPS of $\text{N}^{\text{SARS-CoV-2}}$. **a-b**, Fluorescence microscopy of droplets formed by 50 μM $\text{N}^{\text{SARS-CoV-2}}$ and 1 μM polyU in 20 mM NaPi, pH 7.5, at increasing concentrations of 1,6-hexanediol (a) or NaCl (b). Scale bars, 20 μm . Micrographs are representative of two independent biological replicates. **c**, Turbidity at 350 nm of solutions of $\text{N}^{\text{SARS-CoV-2}}$ in 20 mM NaPi, pH 7.5, at different protein concentrations (5-50 μM) and increasing concentrations of polyU. Average values from three independent measurements are shown and also displayed in Figure 1b. Error bars, std. **d**, Fit of a mono-exponential (red) and bi-exponential (blue) function to FRAP data obtained for $\text{N}^{\text{SARS-CoV-2}}$ /polyU droplets incubated for one hour.

SARS-CoV-2 1 MSDNG-PQ-NQRNA-----PR-ITFGGPTDSTGNSQNGERSGARSKQ---RR-PQ-GL
SARS-CoV 1 MSDNG-PQSNQRSA-----PR-ITFGGPTDSTDNQNGGRNGARPKQ---RR-PQ-GL
MERS-CoV 1 MASPA-----A-----PRAVSFADNNDI TNTNLSRGR-GRNP-----K-PR-AA
HCoV-HKU1 1 MSYTPGHYAGSRSSSGNRSGLKKTSWADQSERNYQTFRGR-KTQPKFTVSTQ-PQ-GN
HCoV-OC43 1 MSFTPGKQSSRASSGNRS-VNGI LKWADQSDQFRNVQTGR-RAQPKQTATSQQSSGN
HCoV-NL63 1 MASV-----NWAADRAA-----R-KK-----
HCoV-299E 1 MATV-----KWADASEPQ-----RGR-QG-----

SARS-CoV-2 46 PNNTASWFTALT-QHGM-EDLKFPRGQGVPI NTNSSPDDQI GYRRATR-RI RGGDGKMK
SARS-CoV 47 PNNTASWFTALT-QHGM-EELRFPARGQGVPI NTNSGFDDQI GYRRATR-RVRRGGDGKMK
MERS-CoV 37 PNNTVSWYGLT-QHGM-VPLTFPPGQGVPLNANSTPAQNAGYWRQDR-KI NTGNG-I K
HCoV-HKU1 58 TI PHYSWFSGIT-QFQKGRDFKPSDQGVPI AFGVPSSEAKGYWRHRSRRSFKTADGQQK
HCoV-OC43 59 VVPPYWSFSGIT-QFQKGFEEFAEGQGVPI APGVPATEAKGYWRHNRRSFKTADGNQR
HCoV-NL63 16 -FPPPSFYMLLVSSDK-APYRVI PRNLVPI GKGN-KDEQI GYWNVQER--WRMRGQRV
HCoV-299E 19 -RI PYSLSYPLLV-DSE-QPWKVI PRNLVPI NKKD-KNKL GYWNVQKR--FRTRKGRV

SARS-CoV-2 103 DLSRWFYFLLGTGPEAGLPYGANKDGI I WATEGAL-NTPKDHI GTRNPANNAI VLQL
SARS-CoV 104 ELSRWFYFLLGTGPEASLPYGANKEGI V WATEGAL-NTPKDHI GTRNPANNAI VLQL
MERS-CoV 93 QLAPRWYFYYTGTGPEAALPFRAVKDDGI VWHEDGAT-DAPST-FGTRNPNNDSDAI VTQF
HCoV-HKU1 117 QLLPRWFYFLLGTGPRYANASYESLEGVFWANHQADTSTPSD-VSSRDTTQEI PTRF
HCoV-OC43 118 QLLPRWFYFLLGTGPHAKDQYGTDI DGVYVWASNAQADVNTPAD-I VDRDFSSDEAI PTRF
HCoV-NL63 71 DLPPKVFHYFLLGTGPHKDLKFRQRSDGVVWAKEGAK-TVNTS-LGNRKRKQKPLEP-KF
HCoV-299E 73 DLSPKLHFYFLLGTGPHKDAKFRERVEGVVWAVDGAK-TEPTG-YGVRKNSSEPI P-HF

SARS-CoV-2 162 PQGTTLPKGFYA-EGSRGGSQASSRSSRS--RNSSRNSTPG-SSRGTSPARMAGNGG--
SARS-CoV 163 PQGTTLPKGFYA-EGSRGGSQASSRSSRS--RNSSRNSTPG-SSRGTSPARMASGGG--
MERS-CoV 151 APGTKLKPNFHI-EGTGNSSQSSSRASSLS--RNSSRSSSQG-SRSGNS-TRGTSPGP--
HCoV-HKU1 176 PPGTILPQGYIV-EGS-GRSAPNSRPSRS--QSRGPNRSLSRSSNFRHSDSI V--
HCoV-OC43 177 PPGTVLPQGYII-EGS-GRSAPNSRSTRT--SSRASSAQRSRANSNRTPTSGV--
HCoV-NL63 128 S--IALPPELSVVEFE-DRSNSSSRASSRSTRNNSRDSSRS-TSRQQSRTSRSNSQSSS
HCoV-299E 130 N--QKLPNGVTVVE-E-PDSRAPSRSSQSSRSG--RGESKP-QSRNPSSDRNHNS--QD

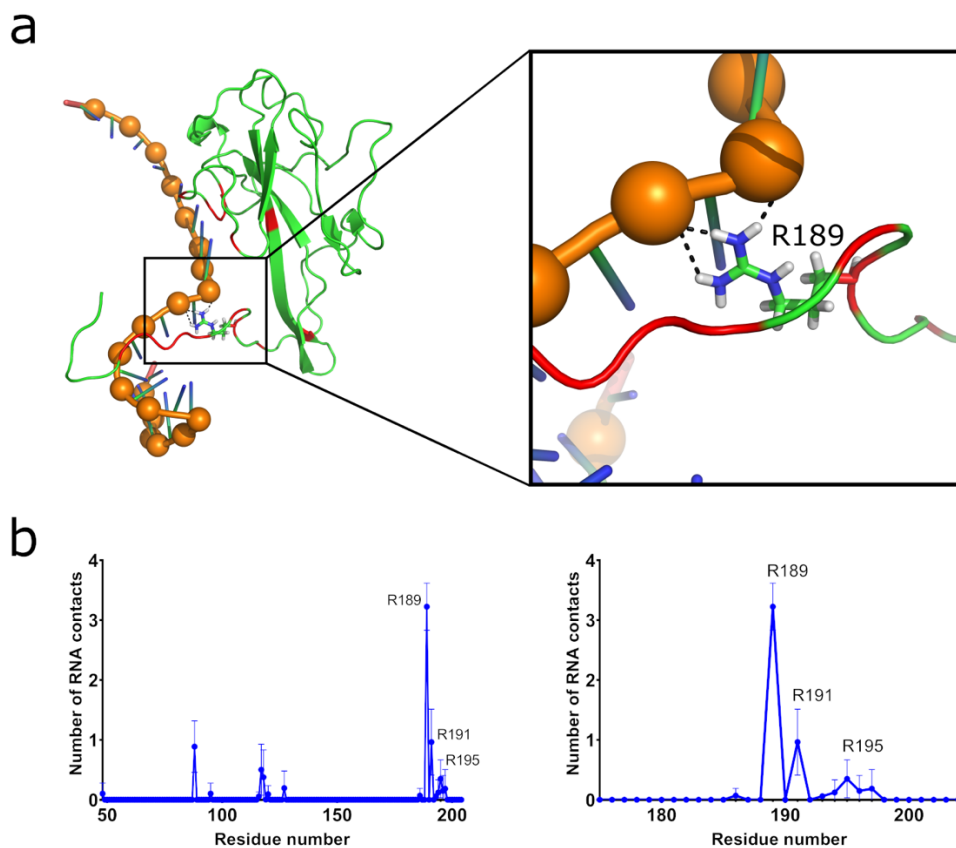
SARS-CoV-2 216 ----DA--ALALLLDRLNQLLESKMSGK-GQQ-QQG-Q-----TV-----TK
SARS-CoV 217 ----ET--ALALLLDRLNQLLESKMSGK-GQQ-QQG-Q-----TV-----TK
MERS-CoV 204 SGI GAV--GGDL LYL D LLNRLQALESGK-VKQ-SQP-K-----VI-----TK
HCoV-HKU1 228 ----K-----PDMAD E I ANLVLA KL GKD-SKP-Q-----QV-----TK
HCoV-OC43 229 ----T-----PDMAD Q I ASLVLA KL GKDAT KP-K-----QV-----TK
HCoV-NL63 184 DLVAAVTLALKNLGF DN--QSKSPSSSGTSTP-KKPNK-----PLSQPR
HCoV-299E 181 DI MKAVAAALKSLGFDPK-DEKDKKSAKTGTP--KPSRNQSPASSQTSAKSLARSQSSET

SARS-CoV-2 249 KSAAE---ASKKPRQKRTATKA--YNYTQAFGRRGPEQTQGNFGDQELI RQGTDYKHWP
SARS-CoV 250 KSAAE---ASKKPRQKRTATKQ--YNYTQAFGRRGPEQTQGNFGDQDLI RQGTDYKHWP
MERS-CoV 241 KDAAA---AKNKMRRHKRTSTKS--FNIMQAFGLRGPGLQGNFGDLQLNKLGTEDPRWP
HCoV-HKU1 255 QNAKEI RHKI LTKPRQKRTPNKH--CNVQCQCFGKRGPS--QNFQNAEMLKLGTNDPQFP
HCoV-OC43 257 HTAKEVRQKI LNKPRQKRSNPKQ--CTVQCQCFGKRGN--QNFGGGEMKLGTSDPQFP
HCoV-NL63 225 ADKPS---QLKKPRWKRVP TRE--ENVI QCFGPRDFN--HNMGSDLVQNGVDAKGF P
HCoV-299E 238 KEQKH---EMQKPRWKRQPNDDVTSNVTQCFGPRDL D--HNFGSAGVVANGVKAKGYP

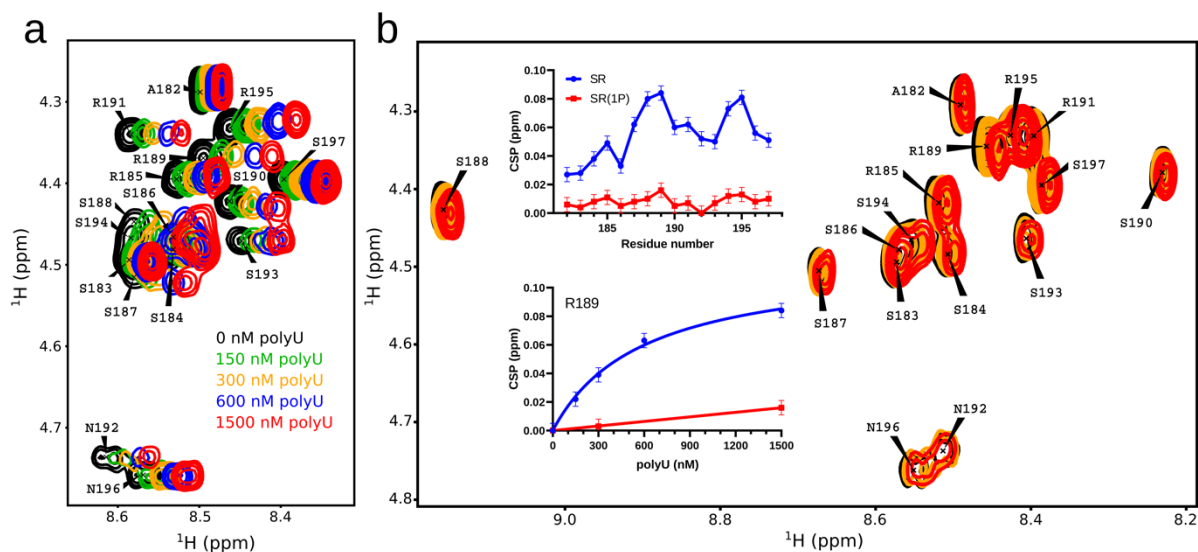
SARS-CoV-2 303 QI AQFAPSASAFFGMSRI GMEVT-----P-SGTWLYTGAI KLDDKDPNFKDQV
SARS-CoV 304 QI AQFAPSASAFFGMSRI GMEVT-----P-SGTWLYHGAI KLDDKDPQFKDNL
MERS-CoV 295 QI AELAPTASAFMGMSQFKLTHQN---ND-DH-GNPVYFLRYSGAI KLDPKNPNYKNVL
HCoV-HKU1 310 I LAELAPTGAFFFGSKLDL VKRDSEA--DSPVK-DVFE LHYSGSI RFDSTLPGFETI M
HCoV-OC43 312 I LAELAPTGAFFFGSRLELAKVQNLSGNPDEPQK-DVYELRYNGAI RFDSTLPGFETI M
HCoV-NL63 276 QLAELI PNQAALFFDSEVSTDEK-----G-DNVQI TYTYKMLVAKDNKLPKFI
HCoV-299E 291 QFAELVPSTAAMLFDSHI VSKES-----G-NTVVLTFTTRVTVPKDHHPHLGKFL

SARS-CoV-2 351 I LLNKHI DAYKTFPP-----TEPK--KDKKK-----KAD-ETQALPQRQKQQTVT-LL
SARS-CoV 352 I LLNKHI DAYKTFPP-----TEPK--KDKKK-----KTD-EAQPLPQRQKQPTVT-LL
MERS-CoV 349 ELLEQNI DAYKTFPK-----KEKKQKAPKEE-----STD-QMSEPPKEQRVCGSI T-QR
HCoV-HKU1 366 KVL EENL NAYVNSNQTDSDSLSSKPKQRKRGVKQLPEQFDSL NLSAGTQHI SN----DF
HCoV-OC43 371 KVL SENL NAYQQQDG---MMNMSPKPQRQRGHKNGQGNENI SVAVPKSRVQONKSI-EL
HCoV-NL63 324 EQI SAFTK-----PSSI KEMQS-----QSS-HV--AQNTVLNASI PESK
HCoV-299E 339 EEI NAFTREMQHPL-----LNPSALEFNPS-----QTS-PA-----TAE

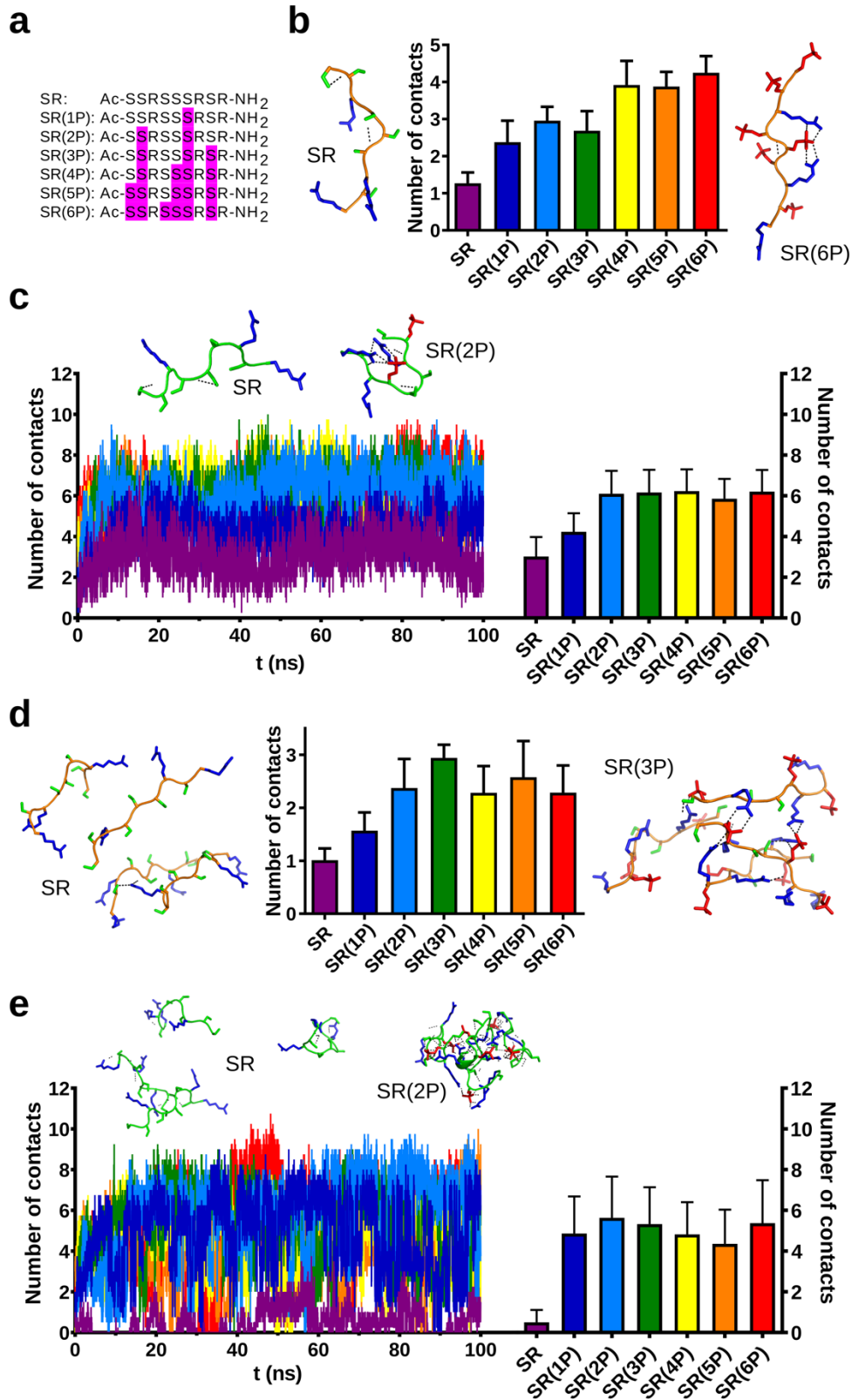
Supplementary Fig. 2. Sequence alignment of the seven human coronaviruses. The following sequences were used: SARS-CoV-2, *YP_009724397.2*; SARS-CoV, *NC_004718*; MERS-CoV, *NC_019843*; HCoV-HKU1, *NC_006577*; HCoV-OC43, *KF530099.1*; HCoV-299E, *NC_002645*; HCoV-NL63, *NC_005831*. Identical residues are marked in red, homologues residues in dark and light green.



Supplementary Fig. 3. Molecular dynamics simulations of RBD-SR (residues 48-204 of N^{SARS-CoV-2}) with RNA. a, Final structure at the end of one of the simulations (1 ns) highlighting the contact of R189 with polyU (zoom on the right). The orange spheres represent phosphate groups and the red color in the protein structure highlights residues, for which polar contacts with RNA were most frequent during the simulation. **b**, Averaged number of RNA contacts from three independent 1 ns simulations in the whole sequence (left) and in the SR-rich region (right). Error bars represent standard deviation.

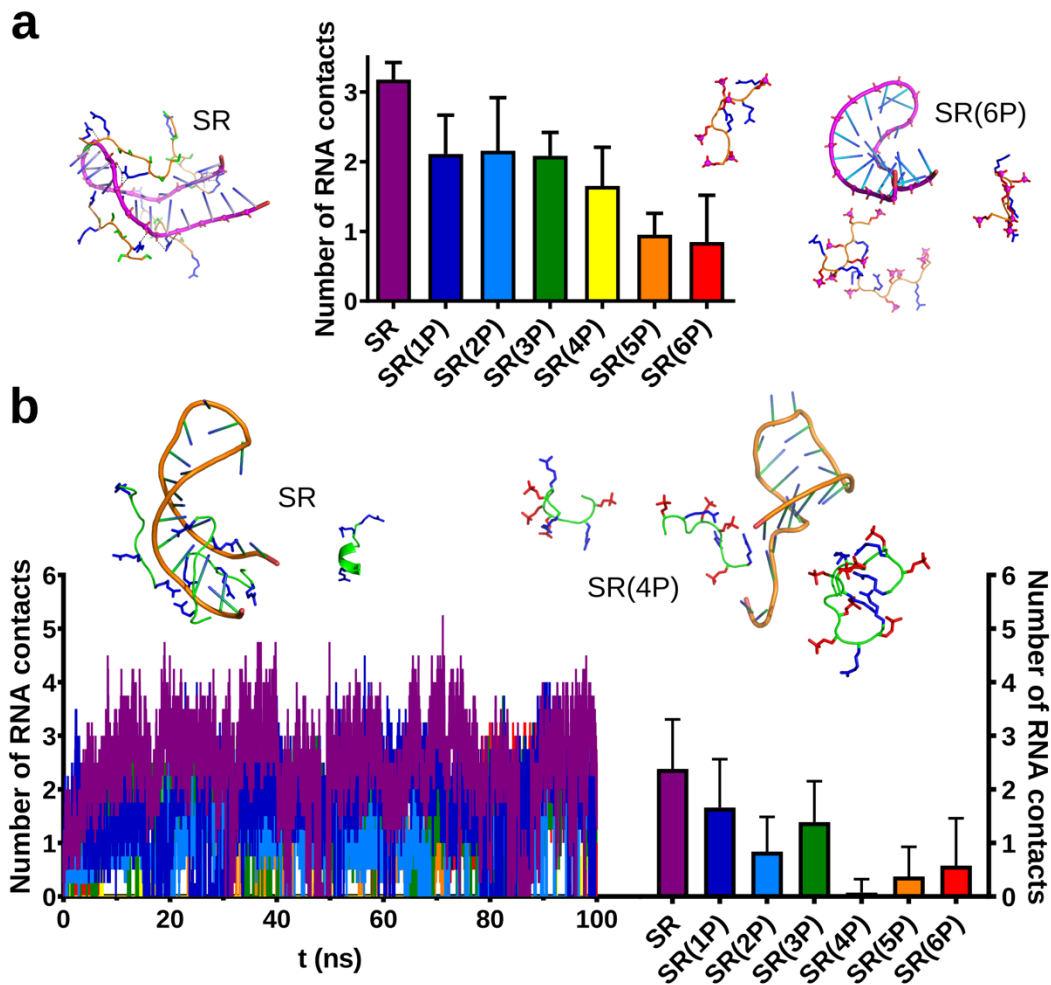


Supplementary Fig. 4. NMR spectroscopy of the interaction of polyU with the SR-peptide comprising residues 182-197 of N^{SARS-CoV-2}. **a**, Selected regions from 2D TOCSY experiments of the SR-peptide at five different concentrations of polyU (0, 150, 300, 600 and 1500 nM). Resonance assignments are indicated. **b**, TOCSY spectra of the SPRK1 single-phosphorylated peptide at three different concentrations of polyU (0, 300 and 1500 nM). The positive charges of the SR-peptide are compensated by the negative charges of polyU at around 300 nM polyU. Spectral color code as in (a). The residue-specific chemical shift perturbation (CSP) observed for each peptide at 1500 nM of polyU are shown in the top plot. The change in R189 CSP with increasing polyU concentration is shown in the bottom plot. Data for the non-phosphorylated SR-peptide are shown in blue, for the SPRK1 single-phosphorylated peptide in red. The CSP error is based on the resolution of the spectra.

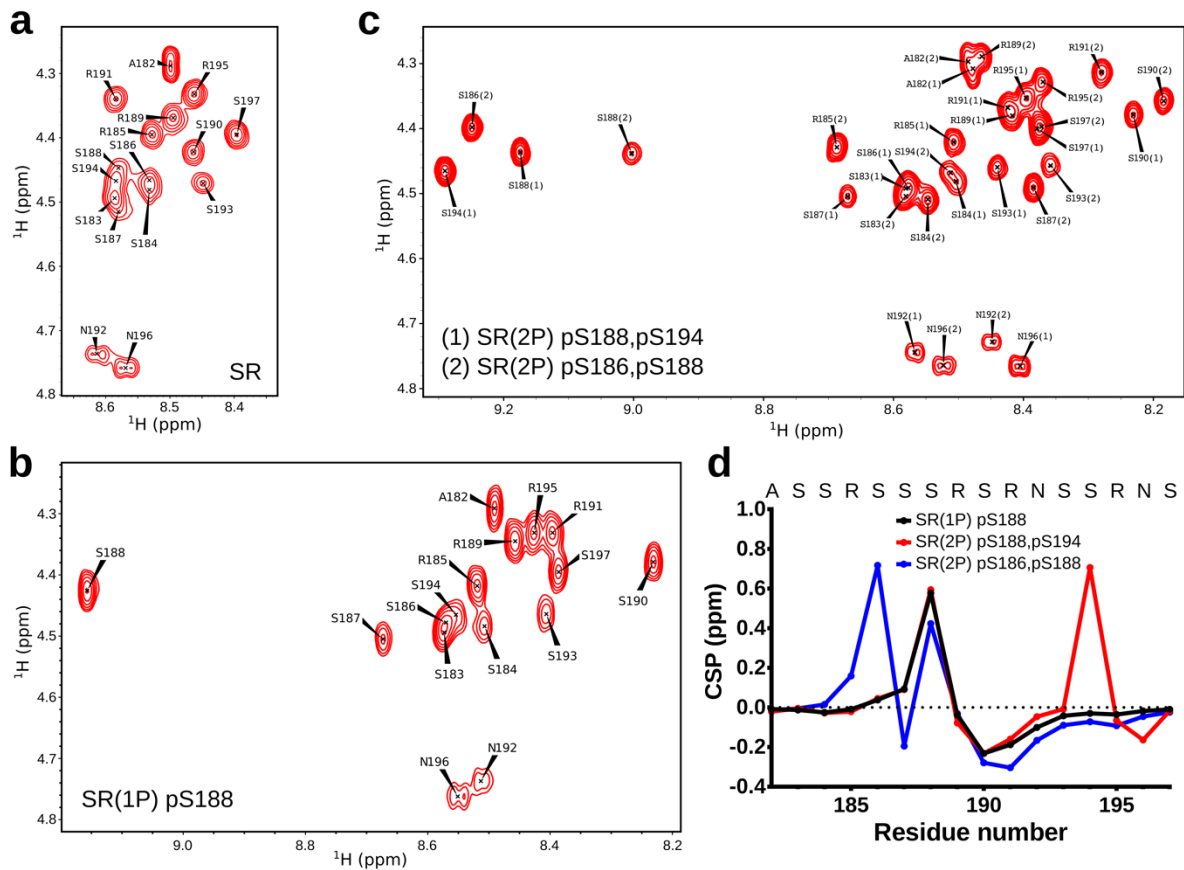


Supplementary Fig. 5. Intra- and intermolecular interactions observed in MD simulations of SR-peptides comprising residues 183-191 of N^{SARS-CoV-2}. a, Amino acid sequences of the SR-peptides with phosphorylated serine residues highlighted in magenta. b, Number of

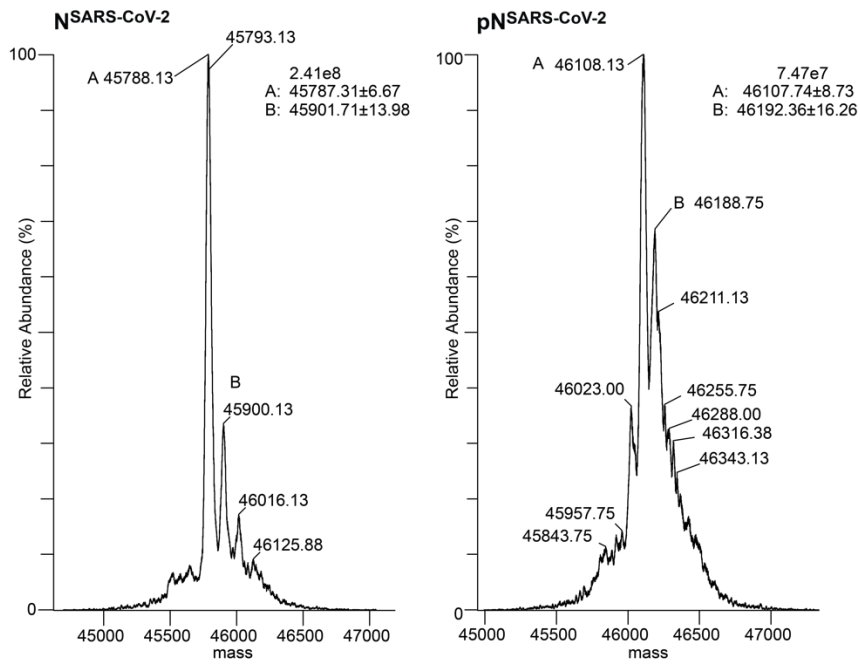
intramolecular polar contacts per peptide for each SR-peptide displayed as the average and standard deviation of five independent 1 ns simulations. Snapshots of the non-phosphorylated (left) and fully-phosphorylated SR-peptide (right) are shown. **c**, Number of intramolecular polar contacts over the trajectory of five 100 ns simulations for the same set of peptides and the averages with standard deviations are represented at the right. MD snapshots of the non-phosphorylated (left) and 2-times phosphorylated (right) peptide are shown above. **d**, Number of intermolecular polar contacts per peptide for each SR-peptide displayed as the average and standard deviation of five independent 1 ns simulations. Snapshots of non-phosphorylated (left) and 3-times phosphorylated SR-peptides (right). **e**, Number of intermolecular polar contacts over the trajectory of 100 ns simulations and the averages with standard deviations are represented to the right. MD snapshots of the non-phosphorylated (left) and 2-times phosphorylated (right) peptide are shown above. MD simulations were performed with four identical peptides in the water box. Salt bridges between the phosphate groups (red) and the arginine side-chains (blue) are marked by dashed lines.



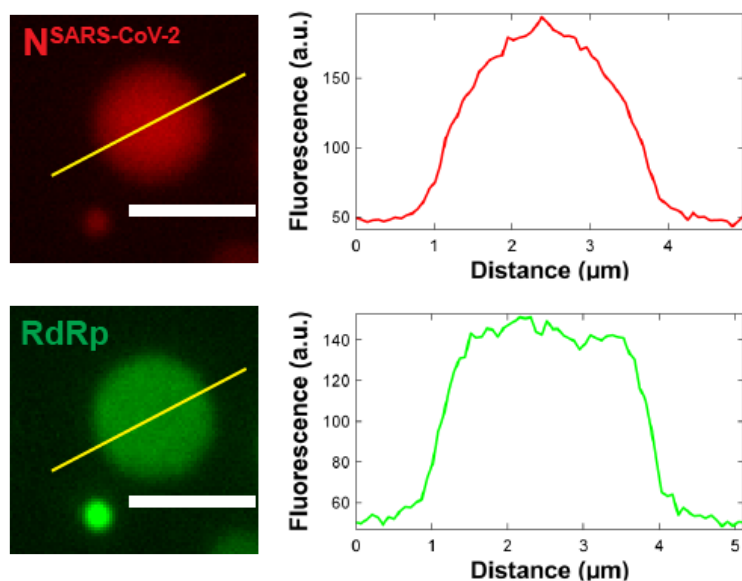
Supplementary Fig. 6. RNA-interactions observed in MD simulations of SR-peptides comprising residues 183-191 of N^{SARS-CoV-2}. **a**, The number of polar contacts of the different SR-peptides (see Supplementary Fig. 4a) with a structured RNA derived from the viral genome of SARS-CoV-2 are displayed as the average and standard deviation of five independent 1 ns simulations. MD snapshots of the RNA-interaction of the non-phosphorylated (left) and fully-phosphorylated (right) peptide are shown at both sides. **b**, Number of polar contacts with RNA over the trajectory of five 100 ns simulations for the same set of SR-peptides. Averages with standard deviations are represented at the right. MD snapshots of the non-phosphorylated (left) and 4-times phosphorylated (right) peptide are shown above. Arginine side chains in blue.



Supplementary Fig. 7. Assignment of TOCSY spectra of SR-peptide comprising residues 182-197 of N^{SARS-CoV-2} and different species of phosphorylated peptide. a, TOCSY spectrum of the non-phosphorylated peptide. **b**, TOCSY spectrum of the peptide phosphorylated only at S188. **c**, TOCSY spectrum of the twice phosphorylated peptides in a 1:1 mixture obtained from the phosphorylation reaction. **d**, Plot of the HN chemical shift perturbation (CSP) of each phosphorylated peptide (pS188 in black, pS188+pS194 in red, pS186+pS188 in blue) with respect to the non-phosphorylated one.



Supplementary Fig. 8. Phosphorylation of N^{SARS-CoV-2} using the kinase SRPK1. Left panel, mass spectrometry of unmodified N^{SARS-CoV-2}, right panel mass spectrometry of SRPK1-phosphorylated N^{SARS-CoV-2}. The identified masses indicate 4 to 5 sites being phosphorylated by SRPK1 in N^{SARS-CoV-2}.



Supplementary Fig. 9. Partitioning of the RdRp/RNA-complex into droplets formed by 50 μM $\text{N}^{\text{SARS-CoV-2}}$ and 1 μM polyU. Fluorescence intensity measurement across the yellow line was plotted against the distance in micrometers. Strong enrichments of Alexa Fluor 594 fluorescently labeled $\text{N}^{\text{SARS-CoV-2}}$ and the RdRp-complex with a fluorescein-labeled minimal RNA hairpin template were observed inside the droplets in 20 mM NaPi, pH 7.5. Scale bar 3 μm . Micrographs are representative of three independent biological replicates.

Supplementary Table 1 – Primer sequences

Primer	Sequence (5'-3')
nsp12 forward	TACTTCCAATCCAATGCATCTGCTGACGCTCAGTCCTTCCTG
nsp12 reverse	TTATCCACTTCCAATGTTATTATTGCAGCACGGTGTGAGGGG
nsp8 forward	TACTTCCAATCCAATGCAGCAATTGCAAGCGAATTTAGCAGCCTG
nsp8 reverse	TTATCCACTTCCAATGTTATTACTGCAGTTTAACTGCGCTATTTGCACG
nsp7 forward	TACTTCCAATCCAATGCAAGCAAATGTCCGATGTTAAATGCACCAGC
nsp7 reverse	TTATCCACTTCCAATGTTATTACTGCAGGGTTGCACGATTATCCAGC