

Phosphorylation modulates liquid-liquid phase separation of the SARS-CoV-2 N protein

Christopher R. Carlson^{1,3,4}, Jonathan B. Asfaha^{1,3,4}, Chloe M. Ghent^{1,4}, Conor J. Howard^{2,3}, Nairi Hartooni^{1,3}, and David O. Morgan^{1,3*}

¹Department of Physiology, University of California, San Francisco CA 94143

²Department of Biochemistry & Biophysics, University of California, San Francisco CA 94143

³Tetrad Graduate Program, University of California, San Francisco CA 94143

⁴Equal contribution

*Correspondence: David.morgan@ucsf.edu

The nucleocapsid (N) protein of coronaviruses serves two major functions: compaction of the RNA genome in the virion and regulation of viral gene transcription in the infected cell¹⁻³. The N protein contains two globular RNA-binding domains surrounded by regions of intrinsic disorder⁴. Phosphorylation of the central disordered region is required for normal viral genome transcription^{5,6}, which occurs in a cytoplasmic structure called the replication transcription complex (RTC)⁷⁻¹¹. It is not known how phosphorylation controls N protein function. Here we show that the N protein of SARS-CoV-2, together with viral RNA, forms biomolecular condensates¹²⁻¹⁵. Unmodified N protein forms partially ordered gel-like structures that depend on multivalent RNA-protein and protein-protein interactions. Phosphorylation reduces a subset of these interactions, generating a more liquid-like droplet. We speculate that distinct oligomeric states support the two functions of the N protein: unmodified protein forms a structured oligomer that is suited for nucleocapsid assembly, and phosphorylated protein forms a liquid-like compartment for viral genome processing. Inhibitors of N protein phosphorylation could therefore serve as antiviral therapy.

Introduction

Coronaviruses are enveloped viruses with a ~30 kb positive-sense single-stranded RNA genome, packed tightly inside the ~100 nm virion in a poorly-defined structure called the nucleocapsid¹⁶⁻¹⁹. Following viral entry and disassembly of the nucleocapsid, the genome is translated to produce RNA-dependent RNA polymerase and numerous non-structural proteins (Nsps)^{1,20-22}. These proteins rearrange membranes of the endoplasmic reticulum to form the RTC⁷⁻¹¹, which is thought to provide a scaffold for the viral proteins that perform genome replication and transcription, and which might also shield these processes from the host cell's innate immune response.

Using genomic (+) RNA as a template, the viral polymerase produces (-) RNA transcripts of subgenomic regions encoding the four major viral structural proteins (S, E, M, and N)^{1,20-22}. Subgenomic transcription involves a template-switching mechanism in which the polymerase completes transcription of a structural protein gene and then skips to a transcription-regulating sequence (TRS) at the 5' end of the genome, resulting in subgenomic (-) RNA fragments – which are then transcribed to produce (+) RNA for translation. N protein is encoded by the most abundant subgenomic RNA and is translated at high levels early in infection. The N protein is among the most abundant viral proteins in the infected cell^{2,3} and accumulates in dynamic

clusters at RTCs^{11,23-25}, where it is thought to help promote the RNA structural rearrangements required for subgenomic transcription^{5,26,27}.

The structural features of the N protein are well conserved among coronaviruses. The ~46 kDa N proteins of SARS-CoV and SARS-CoV-2 are ~90% identical. The N protein contains two globular domains, the N- and C-terminal domains (NTD and CTD), surrounded by intrinsically disordered regions⁴ (Figure 1a, Extended Data Fig. 1). N protein is highly basic (pI~10), and multiple RNA-binding sites are found throughout the protein²⁸. The NTD is an RNA-binding domain²⁹⁻³². The CTD forms a tightly-linked dimer with a large RNA-binding groove³³⁻³⁶, and the fundamental unit of N protein structure is a dimer^{37,38}. Under some conditions, the dimer self-associates to form oligomers that depend on multiple protein regions^{33,35,39-44}. The biochemical features and function of these oligomers are not known.

The central disordered linker contains a conserved serine-arginine (SR)-rich sequence that is likely to serve as a key regulatory hub in N protein function. Early in infection, the SR region is rapidly phosphorylated at multiple sites by cytoplasmic kinases^{5,6,45-50}. Phosphorylation leads to association with the RNA helicase DDX1, which promotes RNA structural changes required for transcription of long subgenomic RNAs in the RTC⁵. Later in infection, nucleocapsid formation and viral assembly do not seem to require N protein phosphorylation, which is substantially reduced in the nucleocapsid of MHV and SARS-CoV virions^{5,6}. We have little understanding of the molecular mechanisms by which phosphorylation influences N protein function.

N protein of SARS-CoV-2 forms RNA-dependent biomolecular condensates

To gain a better understanding of N protein structure and function, we explored the oligomerization of the N protein from SARS-CoV-2, the causative agent of the ongoing COVID-19 pandemic. Purified N protein produced in bacteria migrated on gel filtration as a dimer in high salt but as a large oligomer in physiological salt (Extended Data Fig. 2a). Light microscopy revealed the presence of liquid-like droplets in the presence and absence of added RNA (Extended Data Fig. 2b). We noted, however, that purified N protein contained nucleic acid, even after RNase treatment, raising the possibility that structures seen in the absence of added RNA were due to tightly-bound contaminating RNA. Following removal of RNA by protein denaturation and renaturation, the protein displayed few microscopic structures, but addition of viral RNA greatly enhanced the formation of structures similar to those in the native preparation

(Fig. 1b). We conclude that RNA is required for the formation of the higher-order oligomers seen in the microscope. All subsequent studies were performed with denatured and renatured proteins (Extended Data Fig. 3a).

We first analyzed the effects of three mid-sized viral RNA fragments: (1) 5'-400, containing the 400 nt at the 5' end of the SARS-CoV-2 genome, which is thought to include multiple secondary structure elements and the leader TRS⁵¹⁻⁵³; (2) PS-318, a sequence near the end of ORF1b in SARS-CoV, proposed as a packaging sequence^{54,55} but of unknown function^{51,56}; and (3) N-1260, containing the open reading frame of the N gene near the 3' end of the SARS-CoV-2 genome. RNA encoding firefly luciferase (Luc-1710) was a nonviral control. N protein structures were analyzed by microscopy 30 minutes after addition of RNA at room temperature (Fig. 1b). At 10 μ M N protein, the three viral RNAs rapidly generated branched networks of liquid-like beads. Higher N protein concentrations produced large liquid-like droplets several microns in diameter. Nonviral RNA led to amorphous filamentous aggregates with partial liquid-like appearance.

Incubation at higher temperature (30°C or 37°C) had little effect on droplet formation in a 30-minute incubation (Extended Data Fig. 3b), and longer incubations did not transform filamentous networks into droplets (Extended Data Fig. 3c). High salt dissolved N protein structures, indicating that they depend primarily on electrostatic interactions (Extended Data Fig. 3d).

N protein structures displayed different features at different ratios of N protein to 5'-400 RNA (Fig. 1c). At low RNA concentration (0.125 μ M), 10 μ M N protein formed small spherical droplets. Higher RNA concentrations led to filamentous structures. At RNA concentrations approaching that of N protein, no structures were formed. These results suggest that condensates depend on the crosslinking of multiple N proteins by a single RNA.

We tested the importance of multivalent RNA binding by measuring the effects of a 10-nucleotide RNA carrying the TRS sequence of SARS-CoV-2. The TRS sequence is thought to bind primarily to the NTD, with some contribution from the adjacent SR region^{57,58}. Surprisingly, addition of the TRS RNA triggered the rapid formation of droplets, without the filamentous structures seen with longer viral RNAs (Fig. 1b, c). Also in contrast to results with longer RNAs, droplet formation was greatly reduced when the N protein was in molar excess over the TRS RNA (Fig. 1b, c), indicating that RNA-free N protein exerts a dominant inhibitory effect on the

formation of TRS-bound oligomers. Droplets formed when TRS RNA was equimolar with or in excess of N protein (Fig. 1c), suggesting that these droplets depend on binding of a single TRS RNA to each N protein. Monovalent TRS binding appears to alter N protein structure to enhance low-affinity protein-protein interactions leading to droplet formation. In the more physiologically relevant context of long RNAs, these weak protein-protein interactions are presumably augmented by multivalent RNA-protein interactions.

We next explored the roles of N protein disordered regions. We analyzed mutant proteins lacking the following regions (Fig. 2a; Extended Data Fig. 1): (1) the 44-aa N-terminal extension (NTE), a poorly conserved prion-like sequence with a basic cluster that contributes to RNA binding²⁸; (2) the 31-aa SR region, a basic segment implicated in RNA binding, oligomerization⁴⁴, and phosphorylation^{5,6,45-50}; (3) the 55-aa C-terminal extension (CTE), implicated in oligomerization^{35,42-44}; and (4) the CTD basic patch (CBP), a 33-aa basic region that forms the RNA-binding groove on the surface of the CTD³³⁻³⁵, which can be deleted without affecting CTD dimer structure³⁶.

When combined with viral 5'-400 RNA, none of the deletions completely prevented droplet formation at high protein concentrations (Fig. 2a), indicating that no single disordered segment is essential for the interactions that mediate droplet formation. CTE deletion stimulated the formation of abundant filaments, suggesting that this region normally inhibits interactions. Deletion of the NTE or CBP abolished filaments. Droplets were also observed after deletion of both the NTE and CTE, showing that the central regions are sufficient for droplet formation. Turbidity analyses (Fig. 2b) showed that full-length N protein structures increased abruptly between 5 and 10 μ M, supporting a cooperative mechanism of oligomer assembly. CTE deletion reduced the saturating concentration and NTE deletion increased it, further supporting the negative and positive roles, respectively, of these regions. Deletion of both NTE and CTE resulted in an intermediate phenotype (Fig. 2b).

Further insights arose in studies of deletion mutants and the 10-nt TRS RNA (Fig. 2a). By minimizing the contribution of multivalent RNA binding, these studies illuminated critical protein-protein interactions that contribute to condensate formation. As in the experiments with long RNA, CTE deletion enhanced droplet formation, NTE deletion abolished it, and the double deletion had little effect, pointing to these regions as opposing but nonessential modulators of protein-protein interactions. As in the wild-type protein (Fig. 1b), a molar excess of N protein

suppressed droplet formation by TRS RNA in most mutants; only the CBP deletion caused abundant droplets when the protein was in excess of RNA, suggesting that the CBP is responsible for the reduced droplet formation seen with high N protein concentrations. In contrast to results with long RNA, deletion of the SR region abolished TRS-mediated condensates (Fig. 2a). TRS binding to the NTD is known to be enhanced by the basic SR region, but SR deletion has only a moderate impact on affinity⁵⁷. At the RNA concentration used in our experiments, it is unlikely that SR deletion abolished RNA binding. We therefore suspect that the SR region, perhaps in association with part of the RNA, is required for TRS-dependent droplet formation because it mediates a weak interaction with another N protein⁴⁴.

Phosphorylation promotes more liquid-like N protein condensates

N protein phosphorylation depends on a poorly understood collaboration between multiple kinases^{5,6,49}. A central player is the abundant cytoplasmic kinase GSK-3, which generally phosphorylates serines or threonines four residues upstream of pre-phosphorylated ‘priming’ sites⁵⁹. Studies of the N protein of SARS-CoV⁶ support the presence of two priming sites, P1 and P2 (Fig. 3a), where phosphorylation initiates a series of GSK-3-mediated phosphorylation events, each primed by the previous site, resulting in a high density of up to ten phosphates. The kinases responsible for priming phosphorylation are not known, but the P2 site (S206 in SARS-CoV-2) is a strong consensus sequence (S/T-P-x-K/R) for Cdk1, a major cell cycle kinase⁴⁹.

To produce phosphorylated N protein, we first tested the possibility that Cdk1 primes the protein for subsequent phosphorylation by GSK-3. We found that Cdk1-cyclin B1 phosphorylated N protein in the SR region, and mutation of S206 reduced Cdk1-dependent phosphorylation (Fig. 3b). Phosphorylation might also occur at T198, a nearby Cdk1 consensus site. A combination of Cdk1 and GSK-3 enhanced phosphorylation. Clear evidence for priming by Cdk1 was obtained by extensive unlabeled phosphorylation by Cdk1, followed by analysis of radiolabeled phosphorylation with GSK-3 (Fig. 3c).

Phosphorylation of N protein with a combination of Cdk1 and GSK-3 reduced filamentous structures and promoted the formation of liquid-like droplets (Fig. 3d). GSK-3 alone had no effect, whereas Cdk1 alone promoted droplets to a small extent. We conclude that

phosphorylation in the SR region shifts the behavior of N protein to promote the formation of liquid-like condensates.

We explored the role of phosphorylation in depth with studies of a phosphomimetic mutant in which the ten serines and threonines in the SR region were replaced with aspartate (the 10D mutant). When combined with the 5'-400 viral RNA, the 10D protein rapidly formed condensates with a spherical droplet morphology that was clearly distinct from the filamentous structures of the wild-type protein (Fig. 3e). All three large viral RNAs were effective in driving droplet formation, although N-1260 appeared to reduce the saturating concentration (Extended Data Fig. 4a). NTE deletion in the 10D protein reduced droplet formation, showing once again the positive role of this region (Fig. 3f). Importantly, the 10D mutation had a greater impact on condensate morphology than an SR deletion (Fig. 2a), suggesting that phosphorylation does not just block SR function but might also interfere with other interactions. One possibility, for example, is that the abundant negative charge of the phosphorylated SR region interacts intramolecularly with one or more of the positively-charged patches on the adjacent NTD or CTD, thereby interfering with multivalent RNA-protein interactions.

Droplet formation by TRS RNA was abolished in the 10D mutant (Fig. 3e), just as we observed with TRS RNA and the SR deletion (Fig. 2a). These results support the notion that phosphorylation blocks the weak protein-protein interactions mediated by the SR region.

Phosphorylated N protein is thought to be localized at the RTC⁴⁸. The N protein of mouse hepatitis virus (MHV) is known to interact directly with the N-terminal Ubl1 domain of Nsp3^{23,60}, a large transmembrane protein localized to RTC membranes^{23,60}. We found that a GFP-tagged Ubl1 domain of SARS-CoV-2 Nsp3 partitioned into N protein droplets and filamentous structures (Extended Data Fig. 4b), providing a potential mechanism for association of the RTC with N protein condensates.

To gain a better understanding of the properties of N protein structures, we analyzed the fusion dynamics of different structures over time. During a short (90 s) time course, the filamentous structures of 10 μ M wild-type protein remained immobile and did not fuse, while the spherical droplets of the 10D mutant were highly dynamic and fused rapidly (Fig. 4a). We also found that the droplet-like structures at higher concentrations of unmodified protein displayed relatively slow fusion activity compared to the more dynamic droplets of phosphorylated N protein (Fig.

4b). Thus, unmodified protein forms gel-like condensates that are relatively rigid, while phosphorylated N protein forms condensates that behave more like liquid droplets.

We used negative-stain electron microscopy to further analyze mixtures of 10 μ M N protein and 1 μ M RNA (Fig. 4c). These images represent the saturated protein-RNA solution surrounding condensates. Unmodified wild-type protein and viral RNA formed uniform particles of ~20 nm diameter, reminiscent of structures seen in previous studies of partially disrupted MHV nucleocapsids¹⁷. 2D classification of these particles revealed a uniform size and architecture (Extended Data Fig. 4c). Thus, the gel-like filamentous condensates of the unmodified protein – and possibly the nucleocapsid – are likely to be assembled on a foundation of discrete structural building blocks. In contrast, the mixture of 10D mutant and RNA formed nonuniform, diffuse chains (Fig. 4c), suggesting that phosphorylation disrupts the structural units of the unmodified protein to help create a more liquid-like condensate.

Discussion

We conclude that the N protein of SARS-CoV-2, together with viral RNA, assembles into multiple structural forms that depend on a complex blend of intra- and intermolecular interactions. The more rigid filamentous condensates of the unmodified protein likely depend on high-avidity interactions mediated by multivalent RNA-protein and protein-protein interactions. The latter might include prion-like interactions between NTEs, binding of the SR region to the CTD⁴⁴, or helical CTD polymers that depend on the CBP^{17,33}. Long RNAs augment these protein-protein interactions by interacting with the numerous RNA-binding sites on the protein. Our results also support a model in which phosphorylation of the SR region blocks SR-mediated protein-protein interactions and interferes intramolecularly with RNA binding at other sites, resulting in a loss of affinity that generates a more dynamic liquid droplet.

The different forms of N protein oligomers seem well suited for its two major functions. In the nucleocapsid, where extremely compact RNA packaging is the goal, the organized structures of the unmodified protein could represent an underlying structural framework that is supplemented by liquid-like condensation – much like chromosome packaging depends on underlying nucleosome structure and the liquid-like behavior of chromatin proteins^{61,62}.

The more liquid-like behavior of phosphorylated N protein droplets might be particularly important at the RTC. There is abundant evidence linking N protein phosphorylation with localization and function in the RTC. N protein is fully phosphorylated soon after synthesis and rapidly associates with large membrane structures that presumably represent the RTC⁴⁸. N protein is the only viral structural protein that localizes to the RTC⁶³, where it is seen in immunoelectron microscopy adjacent to the double-membrane vesicles and convoluted membranes of this organelle¹¹. GFP-tagged N protein forms large clusters at RTCs, and fluorescence recovery after photobleaching indicates that N protein is dynamically associated with these clusters²⁵. N protein helps control subgenomic transcription in the RTC, and inhibition of phosphorylation blocks this function^{5,26,27}. With these lines of evidence in mind, our work points to the possibility that a liquid-like matrix of phosphorylated N protein and loosely bound RNA, linked to RTC membranes by Nsp3, provides a compartment to concentrate and protect the viral replication and transcription machinery. Similar mechanisms are likely to exist in negative-sense RNA viruses, where replication is focused in dynamic biomolecular condensates⁶⁴⁻⁶⁶. In Measles virus, these condensates have also been implicated in nucleocapsid assembly⁶⁷. We also note that others have recently observed SARS-CoV-2 N protein condensates⁶⁸⁻⁷¹.

Small chemical inhibitors of GSK-3 kinase activity disrupt MHV genome processing and reduce the production of virions by MHV- or SARS-CoV-infected cells^{5,6}, consistent with the importance of N protein phosphorylation in genome replication. These inhibitors, perhaps together with inhibitors of priming kinases such as Cdk1, have the potential to serve as antiviral therapy in the early stages of COVID-19.

Methods

Plasmid construction and RNA preparation

All expression vectors were constructed by Gibson assembly and confirmed by sequencing. Codon-optimized DNAs encoding SARS-CoV-2 N protein (aa 1-419) and Nsp3 Ubl1 domain (aa 819-920 of polyprotein 1a) were ordered as gBlocks from Integrated DNA Technologies (IDT) and cloned into pET28a expression vectors. N protein mutants were constructed by site-directed mutagenesis and Gibson assembly. A 6xHis-SUMO tag was added to the N-terminus of all N protein constructs. Nsp3 Ubl1 domain was engineered with a C-terminal GFP-2xStrep tag. Codon-optimized DNAs encoding human Cdk1 (untagged) and Cyclin B1 (N-terminal 6xHis-4xMyc tag) were ordered as gBlocks from IDT and cloned individually into the baculovirus expression vector pLIB for construction of recombinant baculoviruses⁷². Human GSK-3 β was purchased from Promega (V1991).

Sequences of the three viral RNAs and firefly luciferase RNA are shown in Extended Data Fig. 5. Templates for transcription in vitro of 5'-400 RNA and PS-318 RNA were ordered as gBlocks from IDT and PCR-amplified with a 5' primer carrying a T7 promoter sequence. The template for N-1260 was PCR-amplified with a 5' primer containing a T7 promoter from a PET28a vector carrying PCR-amplified ORF DNA from the 2019-nCoV_N positive control plasmid from IDT. The N-1260 RNA includes 152 nucleotides from the pET28a plasmid backbone. All RNA synthesis was performed using the HiScribe T7 High Yield RNA synthesis kit (NEB E2040S) according to the manufacture's protocol. Luc-1710 RNA was included as a positive control template in the RNA synthesis kit. The SARS-CoV-2 TRS RNA (UCUAAACGAA, tagged with FAM fluorescent dye) was ordered from IDT.

Protein purification

N protein vectors were transformed into *E. coli* BL21 star (Thermo #C601003) for expression. Freshly transformed cells were grown in TB-Kanamycin (50 μ g/ml) to OD 0.4 at 37°C. The temperature was lowered to 16°C until the cells reached a density of 0.8. Protein expression was then induced with 0.4 mM IPTG for 16 h. Harvested cells were washed with PBS, snap frozen in LN₂, and stored at -80°C until lysis.

To remove contaminating nucleic acid, N protein was purified under denaturing conditions as previously described^{47,73}. Frozen cell pellets were thawed and resuspended in buffer A (50 mM

Hepes pH 7.5, 500 mM NaCl, 10% glycerol, 20 mM imidazole, 6 M urea) and lysed by sonication on ice. Lysed cells were centrifuged at 30,000 rpm for 30 min at 4°C to remove cell debris. Clarified lysate was added to Ni-NTA agarose beads (Qiagen) and incubated for 45 min at 4°C. Ni-NTA beads were then washed with 3 x 10 bed volumes of buffer A, and N protein was eluted with 3 bed volumes buffer B (50 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol, 250 mM imidazole, 6 M urea). The eluate was concentrated to ~1 ml using centrifugal concentrators (30 kDa cutoff, Amicon) and renatured by overnight dialysis against 2 l buffer C (50 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol) at 4°C. 100 µg recombinant Ulp1 catalytic domain (expressed in and purified from bacteria) was added to renatured protein for 10 min on ice to cleave the 6xHis-SUMO tag. Mutants lacking the NTE were not cleaved as efficiently by Ulp1, and complete cleavage required incubation with Ulp1 at 25°C for 4 h. Cleaved protein was then centrifuged at 15,000 rpm for 10 min and injected onto a Superdex 200 10/300 size exclusion column equilibrated in buffer C. Peak fractions were pooled, frozen in LN₂, and stored at -80°C.

In early experiments, N protein was purified under native conditions. Frozen cell pellets were thawed and resuspended in buffer D (50 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol, 20 mM imidazole) supplemented with benzonase, cOmplete EDTA-free protease inhibitor cocktail (Roche), and 1 mM phenylmethylsulfonyl fluoride (PMSF). Cells were lysed by sonication on ice and centrifuged at 30,000 rpm for 30 min at 4°C. Clarified lysate was added to Ni-NTA agarose beads and incubated for 45 min at 4°C. Ni-NTA beads were then washed with 3 x 10 bed volumes buffer D, and N protein was eluted with 3 bed volumes buffer D + 500 mM imidazole. The eluate was concentrated to ~1 ml using centrifugal concentrators (30 kDa cutoff, Amicon) and injected onto a Superdex 200 10/300 size exclusion column. Peak fractions were pooled, snap frozen in LN₂, and stored at -80°C. Protein concentration was measured by nanodrop, and a major A260 peak was observed. The A260 peak was insensitive to treatment with DNase I, RNase A, RNase H, or benzonase. Additionally, small RNA species were observed on native TBE gels stained with Sybr gold.

Nsp3 Ubl1-GFP was expressed in *E. coli* as described above. Frozen cell pellets were thawed and resuspended in buffer E (50 mM Hepes pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM DTT) supplemented with benzonase, cOmplete EDTA-free protease inhibitor cocktail (Roche), and 1 mM PMSF. Cells were lysed by sonication on ice and centrifuged at 30,000 rpm for 30 min at 4°C. Clarified lysate was then added to a 5 ml StrepTrap HP prepacked column (GE). The column was washed with 10 column volumes buffer E and eluted with 4 column volumes buffer

E supplemented with 2.5 mM desthiobiotin. Peak fractions were pooled, snap frozen in LN₂, and stored at -80°C.

Cdk1-cyclin B1 complexes were prepared as follows. Two 666 ml cultures of SF-900 cells (1.6×10^6 cells/ml) were infected separately with Cdk1 or cyclin B1 baculovirus and harvested after 48 h. The two cell pellets were frozen in LN₂. Frozen pellets were thawed and each resuspended in 20 ml lysis buffer (50 mM Hepes pH 7.5, 300 mM NaCl, 10% glycerol, 20 mM imidazole, benzonase, and cOmplete EDTA-free protease inhibitor cocktail). Cells were lysed by sonication. To generate an active Cdk1-cyclin B1 complex phosphorylated by Cdk-activating kinase in the lysate⁷⁴, the two lysates were combined, brought to 5 mM ATP and 10 mM MgCl₂, and incubated at room temperature for 20 min. The combined lysates were centrifuged at 55,000 rpm for 45 min at 4°C. The supernatant was filtered and passed over a HisTRAP nickel affinity column, washed with wash buffer (50 mM Hepes pH 7.5, 300 mM NaCl, 10% glycerol) and eluted with the same buffer plus 200 mM imidazole. Peak fractions were pooled, concentrated to 0.75 ml, and injected on an S200 size exclusion column in wash buffer. Peak fractions containing the Cdk1-cyclin B1 complex were pooled, concentrated to 1.5 mg/ml, and snap frozen in LN₂.

Light microscopy

Glass was prepared as described previously⁷⁵. Individual wells in a 384-well glass bottom plate (Greiner #781892) were incubated with 2% Hellmanex detergent for 1 h. Wells were then washed 3 times with 100 µl ddH₂O. 1 M NaOH was added to the glass for 30 min, followed by washing 3 times with 100 µl ddH₂O. The glass was dried, and 20 mg/ml PEG silane dissolved in 95% EtOH was added to individual wells and incubated overnight (~16 h). The glass was then washed 3 times with 100 µl ddH₂O and dried before sample addition.

The day prior to imaging, protein was thawed and dialyzed against droplet buffer (25 mM Hepes pH 7.5, 70 mM KCl) overnight at 4°C. Protein concentration was then quantified by nanodrop. Reactions containing protein and RNA were combined and mixed immediately before adding to individual wells in the PEG-treated 384-well plate. All reactions were incubated for 30 min at room temperature, unless otherwise indicated, before imaging on a Zeiss Axiovert 200M microscope with a 40x oil objective.

Turbidity analysis

Protein was dialyzed into droplet buffer the night before performing turbidity measurements, as stated above. Protein was then serially diluted into room temperature droplet buffer. After dilution, RNA was added and mixtures were incubated at room temperature for 15 min before measuring the absorbance at 340 nm on a Spectramax M5 plate reader.

Kinase reactions

1 μM N protein was incubated with either 0.14 μM GSK3- β , 0.91 μM Cdk1-Cyclin B1, or both, in a 20 μl reaction mixture containing 10 mM MgCl_2 , 1 mM DTT, 500 μM ATP, 25 mM Hepes pH 7.8, 50 mM NaCl, and 0.001 mCi/ml ^{32}P - γ -ATP. After incubation at 30°C for 30 min, reactions were quenched with 10 μl SDS loading buffer for analysis by SDS-PAGE and visualization with a Phosphorimager. For the priming experiment in Fig. 3c, 19 μM of each N protein variant was incubated with 2.5 μM Cdk1-Cyclin B1 overnight at 25°C in kinase buffer A (20 mM Hepes pH 7.4, 300 mM NaCl, 20 mM MgCl_2 , 1 mM DTT, 1 mM ATP). Reactions were desalted into kinase buffer B (20 mM Hepes pH 7.4, 20 mM MgCl_2 , 1 mM DTT, 500 μM ATP). 10 μM of desalted N protein was incubated with 0.17 μM GSK3- β and 0.05 μCi ^{32}P - γ -ATP. After various times, reactions were quenched with SDS loading buffer for analysis by SDS-PAGE and visualization with a Phosphorimager.

Phosphorylated protein for droplet analysis was prepared in reactions containing 25 μM N protein, 1 μM Cdk1-cyclin B1, 0.2 μM GSK-3 β , 2 mM ATP, 20 mM MgCl_2 , 15 mM Hepes pH 7.5, 300 mM NaCl, 5% glycerol, 1 mM DTT, 8 mM phosphocreatine (Sigma 10621714001), and 0.016 mg/ml creatine kinase (Sigma C3755). The reaction was incubated for 2 h at 30°C and then dialyzed overnight into droplet buffer at 4°C. RNA addition and droplet visualization were carried out as described above.

Electron microscopy

N protein (10 μM) and RNA (1 μM PS-318) were mixed in 10 μl droplet buffer and incubated 15 min at room temperature. 3.5 μl of this solution was adsorbed onto glow-discharged (PELCO EasiGlow, 15 mA, 0.39 mBar, 30 s) carbon-coated grids (200-400 mesh copper, Quantifoil) for 1 min at room temperature. Sample was blotted off, stained and blotted 5 times with 0.75% uranyl formate, and allowed to air dry. Negative stain images were collected with a Tecnai T12 microscope (FEI) with a LaB6 filament, operated at 120 kV, and a Gatan Ultrascan CCD camera

(final pixel size 2.21 Å). Contrast Transfer Function (CTF) estimation was performed with CTFFIND4⁷⁶. Automated particle picking, CTF-correction, and two-dimensional averaging and classification were performed in RELION3⁷⁷.

Data Availability

The data that support the findings of this study are available from the corresponding author upon request.

Acknowledgements

We thank Geeta Narlikar, Sy Redding, Alan Frankel, and Adam Frost for discussions, and Madeline Keenen and Emily Wong for reagents and technical advice. This work was supported by the National Institute of General Medical Sciences (R35-GM118053) and the UCSF Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation.

Author contributions

C.R.C., J.B.A., and C.M.G. contributed to conceptualization, experimental design, and generation of results; N.H. contributed to conceptualization and experimental design; C.J.H. performed EM analysis; D.O.M. provided guidance and wrote the paper with contributions from all authors.

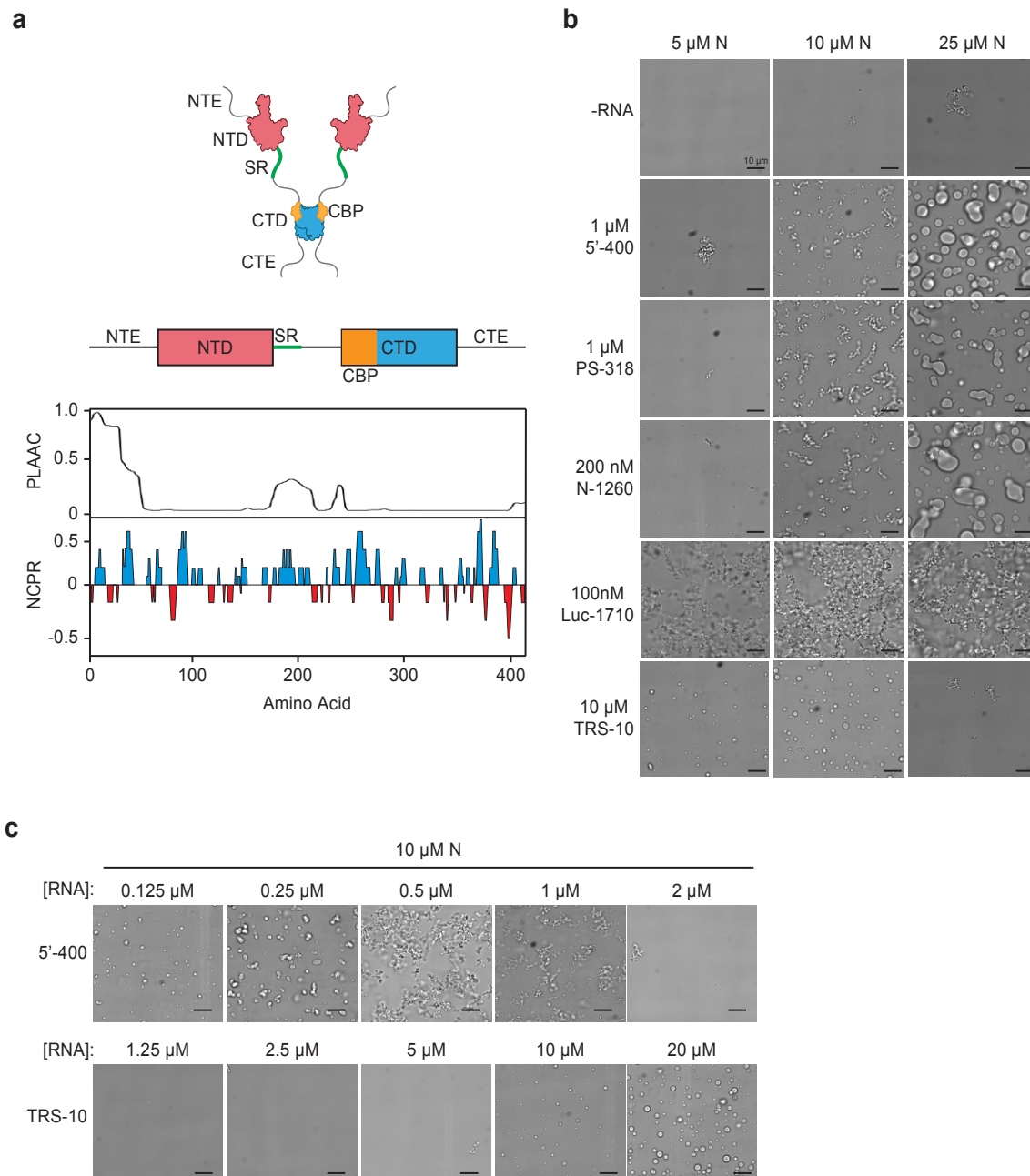


Figure 1. SARS-CoV-2 N protein forms biomolecular condensates in the presence of RNA. **a**, Top, schematic of N protein domain architecture. NTE, N-terminal extension; NTD, N-terminal domain; SR, SR region; CTD, C-terminal domain; CTE, C-terminal extension; CBP, CTD basic patch. Bottom, features of amino acid sequence. PLAAC, prion-like amino acid composition⁷⁸, NCPR, net charge per residue. See Extended Data Fig. 1 for sequence. **b**, Light microscopy images of N protein condensates after 30 min incubation at room temperature with various RNA molecules. **c**, Condensate formation by N protein (10 μ M) over a range of 5'-400 and TRS-10 RNA concentrations. All images are representative of multiple independent experiments; scale bar, 10 μ m (**b**, **c**).

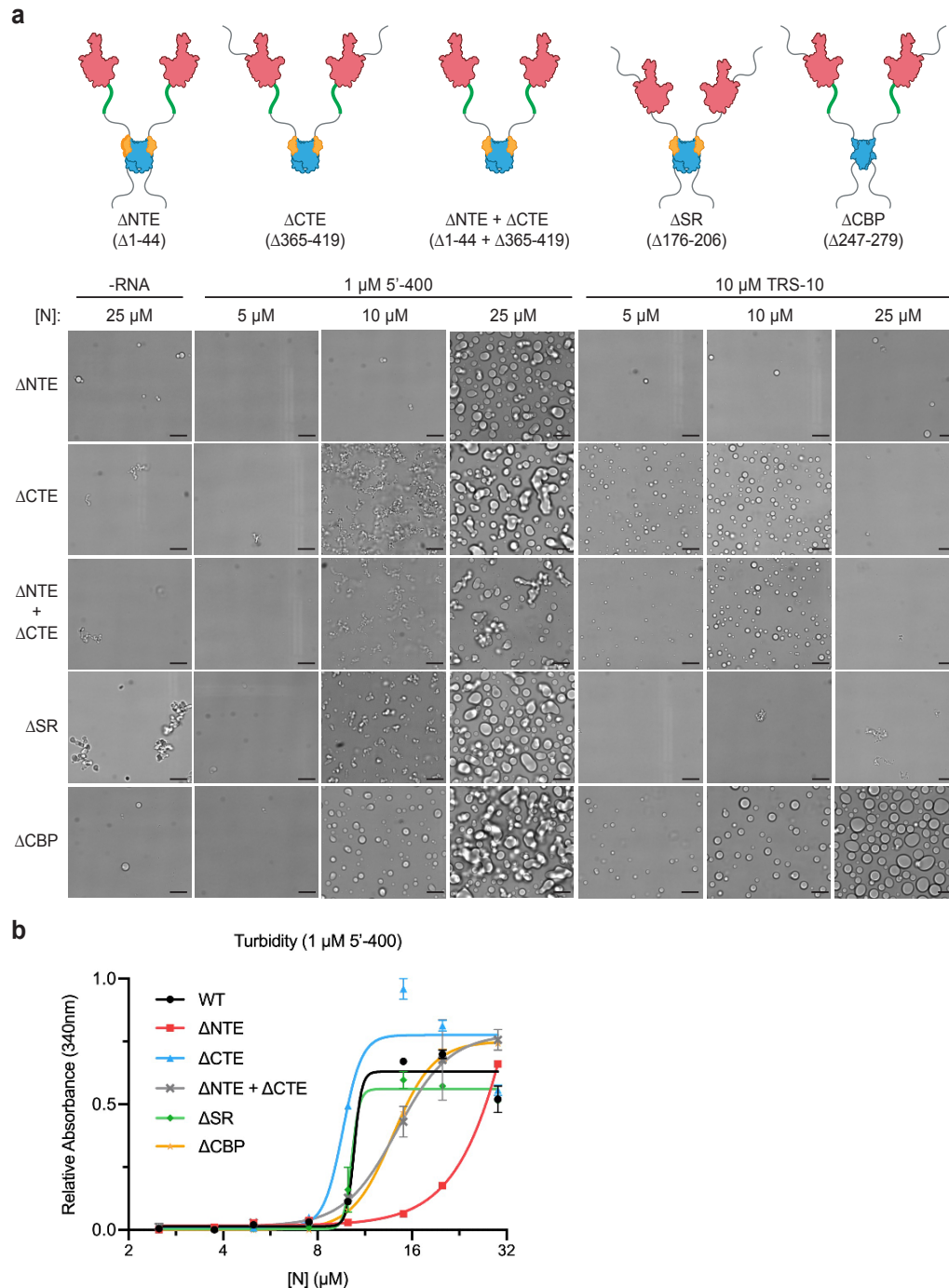


Figure 2. Disordered regions modulate N protein condensate formation. **a**, Top, schematics of N protein deletion mutants. Bottom, N protein condensates observed in the presence of 5'-400 RNA or TRS-10 RNA after a 30 min incubation at room temperature. Images are representative of multiple independent experiments. Scale bar, 10 μ m. **b**, Absorbance at 340 nm was used to quantify the turbidity of N protein mixtures after a 15 min incubation at room temperature with 1 μ M 5'-400. Data points indicate mean $\pm$ s.e.m. of duplicates; representative of two independent experiments.

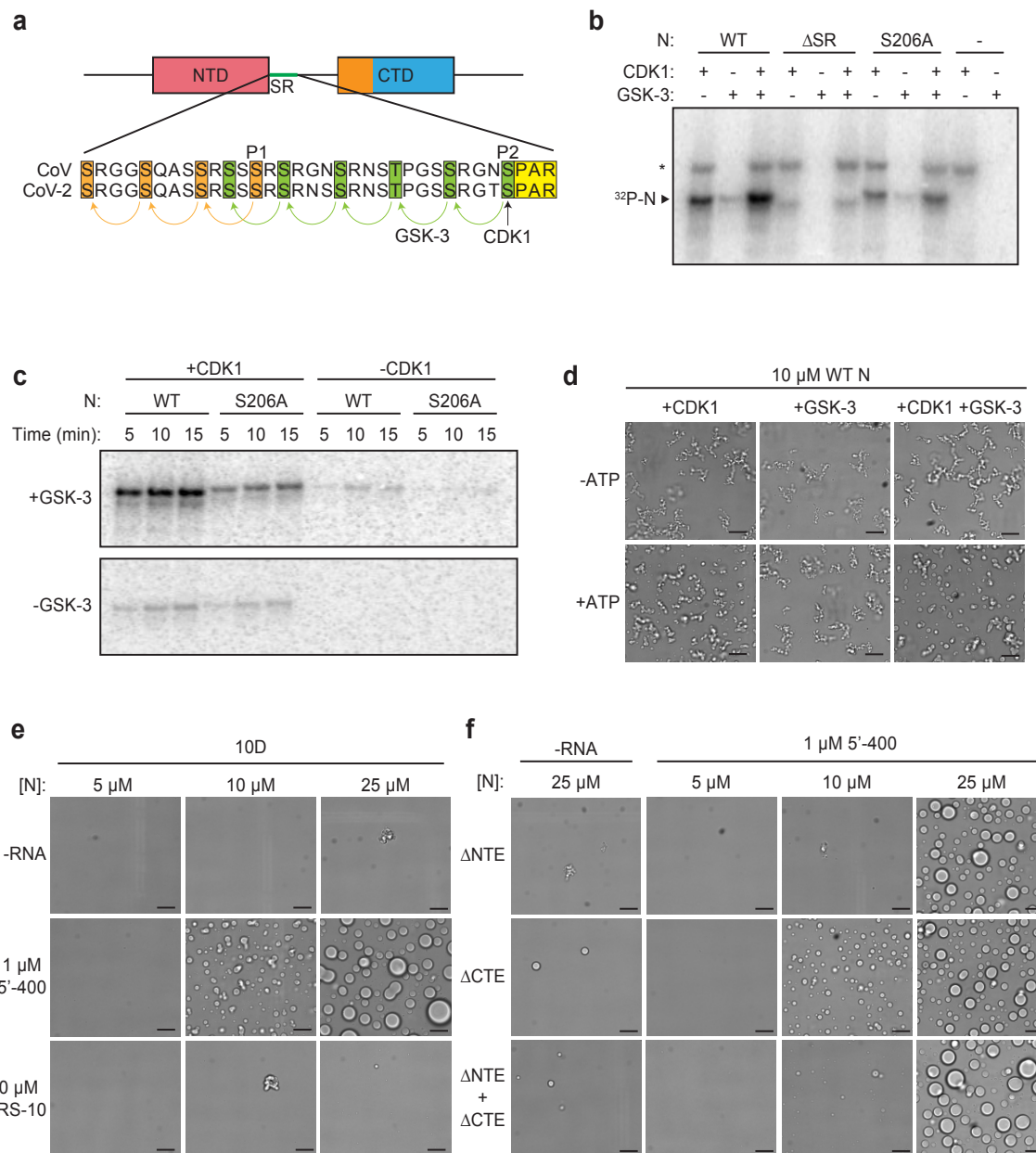


Figure 3. Phosphorylation modulates N protein condensate properties. **a**, Sequences of the SR regions of SARS-CoV N protein (aa 177-210) and SARS-CoV-2 N protein (aa 176-209). Proposed priming sites (P1 and P2) for GSK-3 are indicated⁶. P2 (S206 in CoV-2) is a Cdk consensus site (yellow), where phosphorylation is thought to prime sequential phosphorylation (arrows) of 5 upstream sites (green) by GSK-3. P1 phosphorylation by an unknown kinase primes phosphorylation at 3 upstream sites (orange). **b**, The indicated N protein variants were incubated 30 min with Cdk1-cyclin B1 and/or GSK-3 and radiolabeled ATP, and reaction products were analyzed by SDS-PAGE and autoradiography. Radiolabeled N protein is indicated. Asterisk indicates cyclin B1 autophosphorylation. **c**, N protein was incubated overnight with unlabeled ATP and Cdk1-cyclin B1 (left lanes) or no kinase (right lanes),

desalted, and incubated with or without GSK-3 and radiolabeled ATP. Reaction products were analyzed by SDS-PAGE and autoradiography. **d**, 10 μ M N protein was incubated 2 h with Cdk1-cyclin B1 and GSK-3 in the presence or absence of ATP, dialyzed into droplet buffer overnight, and then mixed with 1 μ M 5'-400 RNA. After 30 min, N protein condensates were analyzed by light microscopy. **e**, Images of N protein 10D mutant following 30 min incubation with or without 1 μ M 5'-400 or 10 μ M TRS-10 RNA. **f**, Images of 10D mutants with the indicated deletions, incubated with or without 1 μ M 5'-400 RNA. All results are representative of multiple independent experiments; scale bar, 10 μ m (**d-f**).

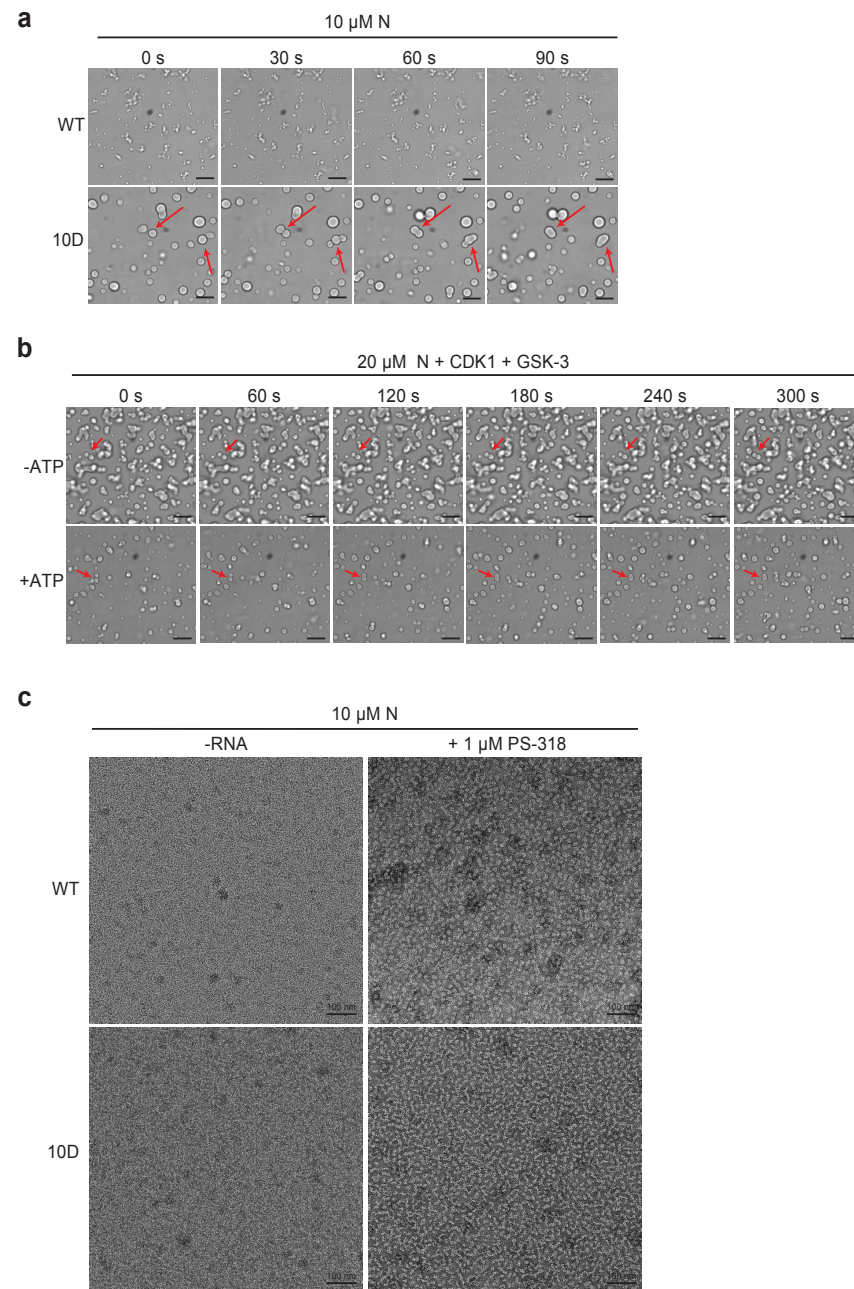
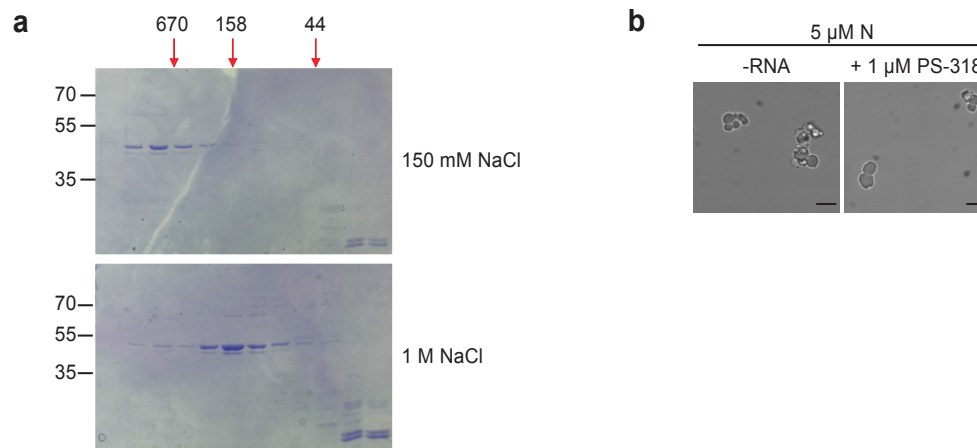


Figure 4. Phosphorylation of N protein promotes liquid-like behavior. **a**, N protein (10 μ M wild-type [WT] or 10D) was mixed with 1 μ M 5'-400 RNA for 20 min, and images were taken at 30 s intervals. Arrows indicate droplet fusion events in the 10D mutant. No fusion events were observed in WT structures. **b**, 20 μ M N protein was phosphorylated with Cdk1-cyclin B1 and GSK-3 as in Fig. 3d, incubated with 1 μ M 5'-400 RNA for 20 min, and imaged every 60 s. Arrows indicate droplet fusion events. Images are representative of multiple independent experiments; scale bar, 10 μ m. **c**, 10 μ M N protein (WT or 10D) was incubated without or with 1 μ M PS-318 RNA for 15 min prior to analysis by negative-stain electron microscopy. Images are representative of three independent experiments. Scale bar, 100 nm.

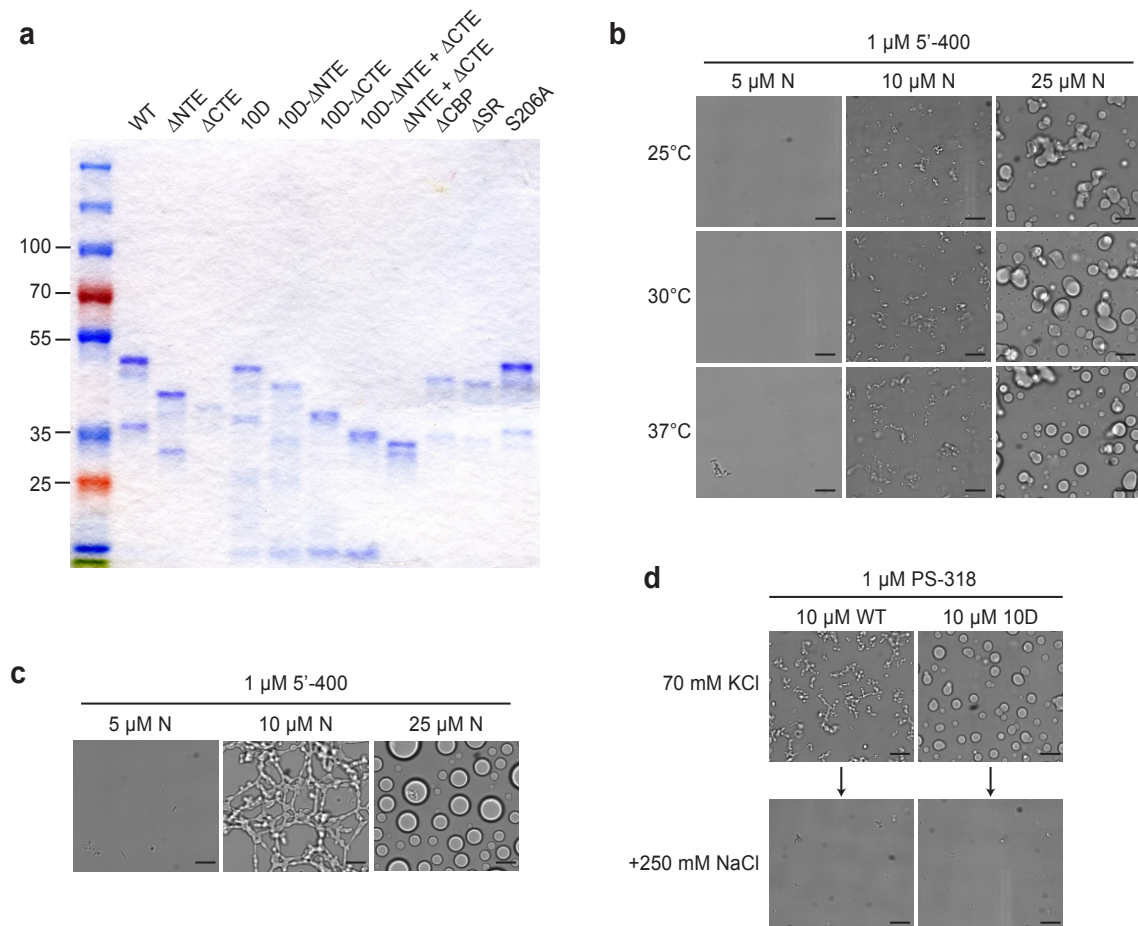
CoV	1	MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQH	60
CoV-2	1	MSDNGPQ-NQRNAPRITFGGPDSTGSGNQNGERSGARSKQRRPQGLPNNTASWFTALTQH	59
NTE			
CoV	61	GKEELRFPRGQGVPIINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYLLGTGPEA	120
CoV-2	60	GKEDLKFRPGQGVPIINTNSSPDQIGYYRRATRRIRGGDGKMKDLSPRWYFYLLGTGPEA	119
CoV	121	SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLPKGFYAEGSRGG	180
CoV-2	120	GLPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIIVLQLPQGTTLPKGFYAEGSRGG	179
CoV	181	SQASSRSSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSQKQ	240
CoV-2	180	SQASSRSSSSRSRNSSRNSTPGSSRGTS ^{SR} PARMAGNGGDAALALLLLDRLNQLESKMSGKQ	239
SR			
CoV	241	QQQGQTVTTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTQDYK	300
CoV-2	240	QQQGQTVTTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTQDYK	299
CBP			
CoV	301	HWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA	360
CoV-2	300	HWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDA	359
CoV	361	YKTFPPTPEPKKDKKKKTDEAQPLPQROKKQPTVTLLPAADMDDFSRQLQNSMSGASADST	420
CoV-2	360	YKTFPPTPEPKKDKKKKADETQALPQROKKQPTVTLLPAADLDDFSKQLQQSMS--SADST	417
CTE			
CoV	421	QA	422
CoV-2	418	QA	419

Extended Data Fig. 1. Amino acid sequences of N proteins from SARS-CoV and SARS-CoV-2. The two globular domains, NTD and CTD, are highlighted in gray. Underlining indicates the four regions analyzed by deletion mutants. Acidic residues highlighted in pink, basic residues highlighted in blue, and phosphorylation sites of the SR region highlighted in yellow.

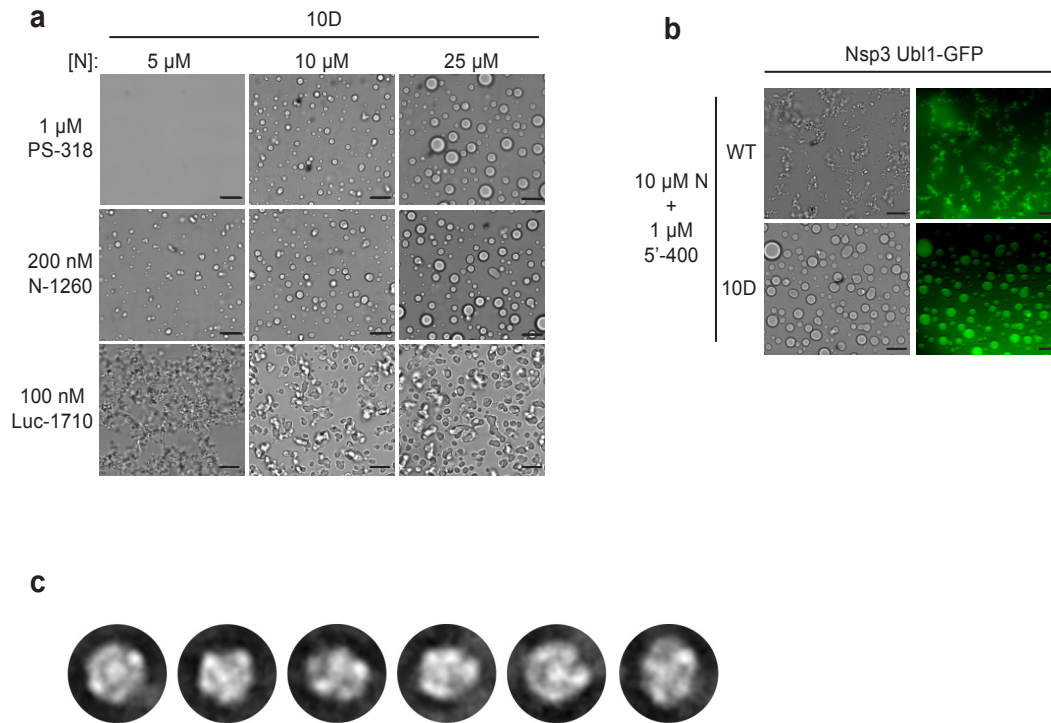


Extended Data Fig. 2. Native N protein preparations form condensates without added

RNA. **a**, Superdex 200 gel filtration analysis of native N protein at 150 mM or 1 M NaCl, purified under non-denaturing conditions. Arrows at top indicate migration of molecular weight standards (kDa). Fractions were analyzed by SDS-PAGE and Coomassie Blue staining. **b**, Representative images of 5 μ M native N protein after incubation in the presence or absence of 1 μ M PS-318 RNA. Scale bar, 10 μ m.



Extended Data Fig. 3. Characterization of N protein condensates. **a**, SDS-PAGE analysis of all N protein mutants used in this study, stained with Coomassie Blue. **b**, N protein was incubated at the indicated temperature for 30 min in the presence of 1 μM 5'-400 RNA. Scale bar, 10 μm. **c**, N protein was incubated with 1 μM 5'-400 for 16 h at room temperature. Scale bar, 10 μm. **d**, Condensates of 10 μM WT or 10D N protein were formed in droplet buffer (70 mM KCl) by incubation with 1 μM PS-318 RNA for 30 min and imaged. NaCl was then added to a final concentration of 250 mM for 15 min before imaging again.



Extended Data Fig. 4. Characterization of N protein condensates. **a**, Images of N protein 10D mutant following 30 min incubation with the indicated RNAs. Scale bar, 10 μ m. **b**, 10 μ M wild-type (WT) or 10D N protein was incubated with 1 μ M 5'-400 RNA for 10 min. Nsp3 Ubl1-GFP was then added to a concentration of 1 μ M and incubated for an additional 15 min before imaging in brightfield (left) or fluorescence (right). **c**, 2D class averages of particles from the EM analysis of wild-type N protein and PS-318 RNA shown in Fig. 4c.

a. 5'-400

ATTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCTGTTCTCTAA
ACGAACCTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCACGCAGTATAATTAAT
AACTAATTACTGTCTGTTGACAGGACACGAGTAACCTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGT
CCGTGTTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAG
AGCCTTGTCCCTGGTTTCAACGAGAAAACACACGTCCAACCTCAGTTTGCCTGTTTTACAGGTTTCGCG
ACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACAT

b. PS-318

TGAGCTTTGGGCTAAGCGTAACATTAACCAGTGCCAGAGATTAAGATACTCAATAATTTGGGTGTTG
ATATCGCTGCTAATACTGTACATCTGGGACTACAAAAGAGAAGCCCCAGCACATGTATCTACAATAG
GTGTCTGCACAATGACTGACATTGCCAAGAAACCTACTGAGAGTGCTTGTCTTCACTTACTGTCTTG
TTTGATGGTAGAGTGGAAGGACAGGTAGACCTTTTTAGAAACGCCCCGTAATGGTGTTTTAATAACAGA
AGGTTTCAGTCAAAGGTCTAACACCTTCAAAGGGACCAGCACAAGCTA

c. N-1260

GGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATAC
CATGGGCAGCAGCCATCATCATCATCACAGCAGCGCCTGGTGCCGCGCGGTACCACGGAAAA
CCTGTATTTTTCAGGGATCCATGTCTGATAATGGACCCCAAATCAGCGAAATGCACCCCGCATTACG
TTTGGTGGACCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAA
CAACGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTACCGCTCTCACTCAACATGGCA
AGGAAGACCTTAAATTCCCTCGAGGACAAGGCGTTCCAATTAACACCAATAGCAGTCCAGATGACCA
AATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATGAAAGATCTCAGT
CCAAGATGGTATTTCTACTACCTAGGAATGGGCCAGAAGCTGGACTTCCCTATGGTGCTAACAAAG
ACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACCAAAAAGATCACATTGGCACCCGCAA
TCCTGCTAACAAATGCTGCAATCGTGCTACAACCTTCTCAAGGAACAACATTGCCAAAAGGCTTCTACG
CAGAAGGGAGCAGAGGCGGCAGTCAAGCCTCTTCTCGTTCTCATCACGTAGTCGCAACAGTTCAA
GAAATTCAACTCCAGGCAGCAGTAGGGGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATG
CTGCTCTTGTCTTGTCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCA
ACAACAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGGCAAAAA
CGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTTCGGCAGACGTGGTCCAGAACAACCCAAAG
GAAATTTTGGGGACCAGGAACCTAATCAGACAAGGAACCTGATTACAAACATTGGCCGCAAAATTGCACA
ATTTGCCCCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTGGCATGGAAGTCACACCTTCGGG
AACGTGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTCAAAGATCAAGTCA
TTTTGCTGAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAAAAAG
AAGAAGGCTGATGAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACCTGTGACTCTTCTTC
CTGCTGCAGATTTGGATGATTTCTCCAACAATTGCAACAATCCATGAGCAGTGCTGACTCAACTCAG
GCCTAAGAATTTCGAGCTCCGTCGACA

d. Luc-1710

GGTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATAACCATGAAAATCGAAGAAGGTAAAGGT
CACCATCACCATCACCACGGATCCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCCGCTTC
TATCCTCTAGAGGATGGAACCGCTGGAGAGCAACTGCATAAGGCTATGAAGAGATACGCCCTGGTT
CCTGGAACAATTGCTTTTACAGATGCACATATCGAGGTGAACATCACGTACGCGGAATACTTCGAAAT
GTCCGTTTCGGTTGGCAGAAGCTATGAAACGATATGGGCTGAATACAAATCACAGAATCGTCGTATGC
AGTGAAAACCTCTCTTCAATTCTTTATGCCGGTGTGGGGCGCGTTATTTATCGGAGTTGCAGTTGCGCC
CGCGAACGACATTTATAATGAACGTGAATTGCTCAACAGTATGAACATTTTCGCAGCCTACCGTAGTGT
TTGTTTCCAAAAAGGGTTGCAAAAAATTTGAACGTGCAAAAAAATTACCAATAATCCAGAAAAATTA
TTATCATGGATTCTAAAACGGATTACCAGGGATTTTCAGTCGATGTACACGTTTCGTCACATCTCATCTA
CCTCCCGGTTTTAATGAATACGATTTTGTACCAGAGTCCTTTGATCGTGACAAAACAATTGCACTGAT
AATGAATTCCTCTGGATCTACTGGGTTACCTAAGGGTGTGGCCCTTCCGCATAGAACTGCCTGCGTC
AGATTCTCGCATGCCAGAGATCCTATTTTGGCAATCAAATCATTCCGGATACTGCGATTTTAAGTGT
TGTTCCATTCCATCACGTTTTTGGAAATGTTTACTACACTCGGATATTTGATATGTGGATTTTCGAGTCGT
CTTAATGTATAGATTTGAAGAAGAGCTGTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTGCGTT
GCTAGTACCAACCCTATTTTCACTTTCGCCAAAAGCACTCTGATTGACAAATACGATTTATCTAATTT
ACACGAAATTGCTTCTGGGGGCGCACCTCTTTCGAAAGAAGTCGGGGAAGCGGTTGCAAAACGCTT

```
CCATCTTCCAGGGATACGACAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCC
GAGGGGGATGATAAACCGGGCGCGGTCCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGAT
CTGGATACCGGGAAAACGCTGGGCGTTAATCAGAGAGGCGAATTATGTGTCAGAGGACCTATGATTA
TGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAAGGATGGATGGCTACATTC
TGGAGACATAGCTTACTGGGACGAAGACGAACACTTCTTCATAGTTGACCGCTTGAAGTCTTTAATTA
AATACAAAGGATATCAGGTGGCCCCCGCTGAATTGGAATCGATATTGTTACAACACCCCAACATCTTC
GACGCGGGCGTGGCAGGTCTTCCCGACGATGACGCCGGTGAACCTCCCGCCGCCGTTGTTGTTTTG
GAGCACGGAAAGACGATGACGGAAAAAGAGATCGTGGATTACGTGCGCCAGTCAAGTAACAACCGCG
AAAAAGTTGCGCGGAGGAGTTGTGTTTGTGGACGAAGTACCGAAAGGTCTTACCGGAAAACCTCGAC
GCAAGAAAAATCAGAGAGATCCTCATAAAGGCCAAGAAGGGCGGAAAGTCCAAACTCGAGTAAGGT
TAACCTGCAGGAGG
```

Extended Data Fig. 5. RNA sequences used in this study. **a**, 5'-400 RNA from SARS-CoV-2 (Wuhan Hu-1 strain; nt 1-400). **b**, PS-318 RNA from SARS CoV (Tor2 strain; nt 19715-20031), with an extra C (red) after A19802 as in Woo et al⁵⁵. **c**, N-1260 RNA containing the open reading frame (gray highlight) of the N gene from SARS-CoV-2 (Wuhan Hu-1 strain; nt 28274-29533), plus flanking plasmid sequence. **d**, Luc-1710 RNA containing the firefly luciferase open reading frame (gray highlight) plus flanking plasmid sequence.

References

- 1 Fung, T. S. & Liu, D. X. Human Coronavirus: Host-Pathogen Interaction. *Annu Rev Microbiol* **73**, 529-557, doi:10.1146/annurev-micro-020518-115759 (2019).
- 2 Laude, H. & Masters, P. S. The Coronavirus Nucleocapsid Protein; in *The Coronaviridae* (ed S. G. Siddell) pp 141-163 (Plenum Press, 1995).
- 3 Masters, P. S. The molecular biology of coronaviruses. *Adv Virus Res* **66**, 193-292, doi:10.1016/S0065-3527(06)66005-3 (2006).
- 4 Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D. & Huang, T. H. The SARS coronavirus nucleocapsid protein--forms and functions. *Antiviral Res* **103**, 39-50, doi:10.1016/j.antiviral.2013.12.009 (2014).
- 5 Wu, C. H., Chen, P. J. & Yeh, S. H. Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* **16**, 462-472, doi:10.1016/j.chom.2014.09.009 (2014).
- 6 Wu, C. H. *et al.* Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. *J Biol Chem* **284**, 5229-5239, doi:10.1074/jbc.M805747200 (2009).
- 7 den Boon, J. A. & Ahlquist, P. Organelle-like membrane compartmentalization of positive-strand RNA virus replication factories. *Annu Rev Microbiol* **64**, 241-256, doi:10.1146/annurev.micro.112408.134012 (2010).
- 8 Hagemeijer, M. C., Rottier, P. J. & de Haan, C. A. Biogenesis and dynamics of the coronavirus replicative structures. *Viruses* **4**, 3245-3269, doi:10.3390/v4113245 (2012).
- 9 Knoops, K. *et al.* SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol* **6**, e226, doi:10.1371/journal.pbio.0060226 (2008).
- 10 Neuman, B. W., Angelini, M. M. & Buchmeier, M. J. Does form meet function in the coronavirus replicative organelle? *Trends Microbiol* **22**, 642-647, doi:10.1016/j.tim.2014.06.003 (2014).
- 11 Ulasli, M., Verheije, M. H., de Haan, C. A. & Reggiori, F. Qualitative and quantitative ultrastructural analysis of the membrane rearrangements induced by coronavirus. *Cell Microbiol* **12**, 844-861, doi:10.1111/j.1462-5822.2010.01437.x (2010).
- 12 Alberti, S., Gladfelter, A. & Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434, doi:10.1016/j.cell.2018.12.035 (2019).
- 13 Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298, doi:10.1038/nrm.2017.7 (2017).

- 14 Owen, I. & Shewmaker, F. The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins. *Int J Mol Sci* **20**, doi:10.3390/ijms20215501 (2019).
- 15 Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, doi:10.1126/science.aaf4382 (2017).
- 16 Barcena, M. *et al.* Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirus. *Proc Natl Acad Sci U S A* **106**, 582-587, doi:10.1073/pnas.0805270106 (2009).
- 17 Gui, M. *et al.* Electron microscopy studies of the coronavirus ribonucleoprotein complex. *Protein Cell* **8**, 219-224, doi:10.1007/s13238-016-0352-8 (2017).
- 18 Neuman, B. W. *et al.* Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J Virol* **80**, 7918-7928, doi:10.1128/JVI.00645-06 (2006).
- 19 Neuman, B. W. & Buchmeier, M. J. Supramolecular Architecture of the Coronavirus Particle. *Adv Virus Res* **96**, 1-27, doi:10.1016/bs.aivir.2016.08.005 (2016).
- 20 Kim, D. *et al.* The Architecture of SARS-CoV-2 Transcriptome. *Cell* **181**, 914-921 e910, doi:10.1016/j.cell.2020.04.011 (2020).
- 21 Snijder, E. J., Decroly, E. & Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res* **96**, 59-126, doi:10.1016/bs.aivir.2016.08.008 (2016).
- 22 Sola, I., Almazan, F., Zuniga, S. & Enjuanes, L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* **2**, 265-288, doi:10.1146/annurev-virology-100114-055218 (2015).
- 23 Cong, Y. *et al.* Nucleocapsid Protein Recruitment to Replication-Transcription Complexes Plays a Crucial Role in Coronaviral Life Cycle. *J Virol* **94**, doi:10.1128/JVI.01925-19 (2020).
- 24 Hagemeijer, M. C. *et al.* Dynamics of coronavirus replication-transcription complexes. *J Virol* **84**, 2134-2149, doi:10.1128/JVI.01716-09 (2010).
- 25 Verheije, M. H. *et al.* The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes. *J Virol* **84**, 11575-11579, doi:10.1128/JVI.00569-10 (2010).
- 26 Almazan, F., Galan, C. & Enjuanes, L. The nucleoprotein is required for efficient coronavirus genome replication. *J Virol* **78**, 12683-12688, doi:10.1128/JVI.78.22.12683-12688.2004 (2004).
- 27 Zuniga, S. *et al.* Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J Virol* **84**, 2169-2175, doi:10.1128/JVI.02011-09 (2010).
- 28 Chang, C. K. *et al.* Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. *J Virol* **83**, 2255-2264, doi:10.1128/JVI.02001-08 (2009).

- 29 Fan, H. *et al.* The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. *Structure* **13**, 1859-1868, doi:10.1016/j.str.2005.08.021 (2005).
- 30 Huang, Q. *et al.* Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. *Biochemistry* **43**, 6059-6063, doi:10.1021/bi036155b (2004).
- 31 Jayaram, H. *et al.* X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J Virol* **80**, 6612-6620, doi:10.1128/JVI.00157-06 (2006).
- 32 Kang, S. *et al.* Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *BioRxiv*, doi:<https://doi.org/10.1101/2020.03.06.977876> (2020).
- 33 Chen, C. Y. *et al.* Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA. *J Mol Biol* **368**, 1075-1086, doi:10.1016/j.jmb.2007.02.069 (2007).
- 34 Takeda, M. *et al.* Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J Mol Biol* **380**, 608-622, doi:10.1016/j.jmb.2007.11.093 (2008).
- 35 Ye, Q., West, A. M. V., Silletti, S. & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *bioRxiv*, doi:10.1101/2020.05.17.100685 (2020).
- 36 Yu, I. M., Oldham, M. L., Zhang, J. & Chen, J. Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arteriviridae. *J Biol Chem* **281**, 17134-17139, doi:10.1074/jbc.M602107200 (2006).
- 37 Luo, H. *et al.* In vitro biochemical and thermodynamic characterization of nucleocapsid protein of SARS. *Biophys Chem* **112**, 15-25, doi:10.1016/j.bpc.2004.06.008 (2004).
- 38 Yu, I. M. *et al.* Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. *J Biol Chem* **280**, 23280-23286, doi:10.1074/jbc.M501015200 (2005).
- 39 Chang, C. K., Chen, C. M., Chiang, M. H., Hsu, Y. L. & Huang, T. H. Transient oligomerization of the SARS-CoV N protein—implication for virus ribonucleoprotein packaging. *PLoS One* **8**, e65045, doi:10.1371/journal.pone.0065045 (2013).
- 40 Cong, Y., Kriegenburg, F., de Haan, C. A. M. & Reggiori, F. Coronavirus nucleocapsid proteins assemble constitutively in high molecular oligomers. *Sci Rep* **7**, 5740, doi:10.1038/s41598-017-06062-w (2017).
- 41 He, R. *et al.* Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem Biophys Res Commun* **316**, 476-483, doi:10.1016/j.bbrc.2004.02.074 (2004).

- 42 Hurst, K. R., Koetzner, C. A. & Masters, P. S. Identification of in vivo-interacting domains of the murine coronavirus nucleocapsid protein. *J Virol* **83**, 7221-7234, doi:10.1128/JVI.00440-09 (2009).
- 43 Luo, H., Chen, J., Chen, K., Shen, X. & Jiang, H. Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* **45**, 11827-11835, doi:10.1021/bi0609319 (2006).
- 44 Luo, H., Ye, F., Chen, K., Shen, X. & Jiang, H. SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization. *Biochemistry* **44**, 15351-15358, doi:10.1021/bi051122c (2005).
- 45 Davidson, A. D. *et al.* Characterisation for the transcriptome and proteome of SARS-CoV-2 using direct RNA sequencing and tandem mass spectrometry reveals evidence for a cell passage induced in-frame deletion in the spike glycoprotein that removes the furin-like cleavage site. *BioRxiv*, doi:<https://doi.org/10.1101/2020.03.22.002204> (2020).
- 46 Fung, T. S. & Liu, D. X. Post-translational modifications of coronavirus proteins: roles and function. *Future Virol* **13**, 405-430, doi:10.2217/fvl-2018-0008 (2018).
- 47 Peng, T. Y., Lee, K. R. & Tarn, W. Y. Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization. *FEBS J* **275**, 4152-4163, doi:10.1111/j.1742-4658.2008.06564.x (2008).
- 48 Stohlman, S. A., Fleming, J. O., Patton, C. D. & Lai, M. M. Synthesis and subcellular localization of the murine coronavirus nucleocapsid protein. *Virology* **130**, 527-532, doi:10.1016/0042-6822(83)90106-x (1983).
- 49 Surjit, M. *et al.* The severe acute respiratory syndrome coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation. *J Virol* **79**, 11476-11486, doi:10.1128/JVI.79.17.11476-11486.2005 (2005).
- 50 White, T. C., Yi, Z. & Hogue, B. G. Identification of mouse hepatitis coronavirus A59 nucleocapsid protein phosphorylation sites. *Virus Res* **126**, 139-148, doi:10.1016/j.virusres.2007.02.008 (2007).
- 51 Chen, S. C. & Olsthoorn, R. C. Group-specific structural features of the 5'-proximal sequences of coronavirus genomic RNAs. *Virology* **401**, 29-41, doi:10.1016/j.virol.2010.02.007 (2010).
- 52 Yang, D. & Leibowitz, J. L. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res* **206**, 120-133, doi:10.1016/j.virusres.2015.02.025 (2015).
- 53 Yang, D., Liu, P., Wudeck, E. V., Giedroc, D. P. & Leibowitz, J. L. SHAPE analysis of the RNA secondary structure of the Mouse Hepatitis Virus 5' untranslated region and N-terminal nsp1 coding sequences. *Virology* **475**, 15-27, doi:10.1016/j.virol.2014.11.001 (2015).

- 54 Hsieh, P. K. *et al.* Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J Virol* **79**, 13848-13855, doi:10.1128/JVI.79.22.13848-13855.2005 (2005).
- 55 Woo, J., Lee, E. Y., Lee, M., Kim, T. & Cho, Y. E. An in vivo cell-based assay for investigating the specific interaction between the SARS-CoV N-protein and its viral RNA packaging sequence. *Biochem Biophys Res Commun* **520**, 499-506, doi:10.1016/j.bbrc.2019.09.115 (2019).
- 56 Masters, P. S. Coronavirus genomic RNA packaging. *Virology* **537**, 198-207, doi:10.1016/j.virol.2019.08.031 (2019).
- 57 Grosseohme, N. E. *et al.* Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. *J Mol Biol* **394**, 544-557, doi:10.1016/j.jmb.2009.09.040 (2009).
- 58 Keane, S. C., Liu, P., Leibowitz, J. L. & Giedroc, D. P. Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucleocapsid (N) protein. *J Biol Chem* **287**, 7063-7073, doi:10.1074/jbc.M111.287763 (2012).
- 59 Beurel, E., Grieco, S. F. & Jope, R. S. Glycogen synthase kinase-3 (GSK3): regulation, actions, and diseases. *Pharmacol Ther* **148**, 114-131, doi:10.1016/j.pharmthera.2014.11.016 (2015).
- 60 Keane, S. C. & Giedroc, D. P. Solution structure of mouse hepatitis virus (MHV) nsp3a and determinants of the interaction with MHV nucleocapsid (N) protein. *J Virol* **87**, 3502-3515, doi:10.1128/JVI.03112-12 (2013).
- 61 Larson, A. G. *et al.* Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240, doi:10.1038/nature22822 (2017).
- 62 Larson, A. G. & Narlikar, G. J. The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation. *Biochemistry* **57**, 2540-2548, doi:10.1021/acs.biochem.8b00401 (2018).
- 63 V'Kovski, P. *et al.* Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8**, doi:10.7554/eLife.42037 (2019).
- 64 Heinrich, B. S., Maliga, Z., Stein, D. A., Hyman, A. A. & Whelan, S. P. J. Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. *mBio* **9**, doi:10.1128/mBio.02290-17 (2018).
- 65 Nikolic, J. *et al.* Negri bodies are viral factories with properties of liquid organelles. *Nat Commun* **8**, 58, doi:10.1038/s41467-017-00102-9 (2017).
- 66 Zhou, Y., Su, J. M., Samuel, C. E. & Ma, D. Measles Virus Forms Inclusion Bodies with Properties of Liquid Organelles. *J Virol* **93**, doi:10.1128/JVI.00948-19 (2019).

- 67 Guseva, S. *et al.* Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci Adv* **6**, eaaz7095, doi:10.1126/sciadv.aaz7095 (2020).
- 68 Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid is dynamic, disordered, and phase separates with RNA. *BioRxiv*, doi:<https://doi.org/10.1101/2020.06.17.158121> (2020).
- 69 Iserman, C. *et al.* Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate. *BioRxiv*, doi:<https://doi.org/10.1101/2020.06.11.147199> (2020).
- 70 Perdikari, T. M. *et al.* SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins. *BioRxiv*, doi:<https://doi.org/10.1101/2020.06.09.141101> (2020).
- 71 Savastano, A., Ibanez de Opakua, A., Rankovic, M. & Zweckstetter, M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *BioRxiv*, doi:<https://doi.org/10.1101/2020.06.18.160648> (2020).
- 72 Weissmann, F. *et al.* biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proc Natl Acad Sci U S A* **113**, E2564-2569, doi:10.1073/pnas.1604935113 (2016).
- 73 Wang, Y. *et al.* Low stability of nucleocapsid protein in SARS virus. *Biochemistry* **43**, 11103-11108, doi:10.1021/bi049194b (2004).
- 74 Desai, D., Gu, Y. & Morgan, D. O. Activation of human cyclin-dependent kinases in vitro. *Mol Biol Cell* **3**, 571-582, doi:10.1091/mbc.3.5.571 (1992).
- 75 Keenen, M. M., Larson, A. G. & Narlikar, G. J. Visualization and Quantitation of Phase-Separated Droplet Formation by Human HP1alpha. *Methods Enzymol* **611**, 51-66, doi:10.1016/bs.mie.2018.09.034 (2018).
- 76 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
- 77 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).
- 78 Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501-2502, doi:10.1093/bioinformatics/btu310 (2014).