Title
Mechanisms of RNA packaging in Coronaviruses

Author
Carlson, Christopher

Publication Date
2022

Peer reviewed|Thesis/dissertation

Mechanisms of RNA packaging in Coronaviruses

by
Christopher Carlson

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Alexander Johnson
E2BBE6A63C2745B...

Alexander Johnson
Chair

DocuSigned by:
David Morgan
DocuSigned by:

David Morgan

David A. Agard
D6BB224E0EB04B9...

David A. Agard

Committee Members

Dedicated to my cat, Atlas

Acknowledgements

Any attempt to properly acknowledge those who have made my journey into what it was would be a disservice. I'll go ahead and try anyway.

Only two have been by my side nearly every day for the entirety of my PhD: my advisor, Dave Morgan, and my cat, Atlas. I'll start with Dave. I first met Dave back in February of 2017 during my interview at UCSF; we had a great discussion, and he is a large part of the reason I chose to come to UCSF. Dave embodies UCSF culture to his core – on the surface, he is laid back and chill, but if you zoom in, you realize he is focused, rigorous, and deeply driven by his curiosity. I resonate with these qualities and feel lucky to have worked so closely with Dave over the last few years. Who would have thought what we would stumble upon. Anyway, thank you Dave for making my graduate school experience a blast.

Next up, my cat Atlas. Since, to my knowledge, he cannot read, I will keep this brief. Atlas has been by my side for the last five years and has been through it all with me. While I can easily imagine better, nicer cats than Atlas, I cannot imagine a better cat for me. Thanks Atlas, for reminding me what's really important – a strict, food-focused routine, plenty of naps, and an abundance of play time.

It is easy enough to quantify the contributions that Dave and Atlas have made to my graduate school experience. It is much harder to dissect what my friends have contributed to my life over the last five years because of how instrumental my community has been in shaping the person I am today. Varun and Elise are the day-ones. My time at UCSF would be certifiably worse without them in my life. Thank you both. Henry also deserves a major shoutout; it has been fun working with you over the

last couple of years (and fun being friends throughout graduate school), and I will very much miss our coffee dates. I want to thank my class as well, and particularly Eric, Ady, Francesca, and Luke, for keeping me sane and enlightening me to new perspectives. And then of course, I need to acknowledge Tyler and Eddie. You have both shaped me fundamentally as a person, and I carry pieces of you both into the rest of my life.

Lastly, and most importantly, we have the family. My parents, Rich and Sally, have given me opportunities to succeed at every step of my life. This is worth of reiteration: I would not be where I am or who I am without your love and support – no one has had a more positive impact on my life than either of you. The only person who comes close is my sister Stephanie. For most of my life, you paved the trail, and I followed in tow. I wouldn't have made it here without you, and I'm looking forward to a future with you all in it.

Contributions

This thesis is a reproduction of material published in Molecular Cell and in review for publication at the Journal of Biological Chemistry. Citations are presented below.

Carlson CR*, Asfaha JB*, Ghent CM*, Howard CJ, Hartooni N, Safari M, Frankel AD, Morgan DO. (2020) Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Molecular Cell* 80: 1092-1103. *equal contribution

C.R.C., J.B.A., C.M.G., C.J.H., and N.H. contributed to conceptualization, experimental design, and generation of results; C.R.C. prepared proteins and performed microscopy analysis of condensates; J.B.A. and C.M.G. prepared proteins and performed kinase experiments; C.J.H. performed EM analysis; N.H. performed FRAP analysis; M.S. assisted with protein expression and microscopy with guidance from A.D.F.; and D.O.M. provided guidance and wrote the paper with contributions from all authors.

Carlson CR, Adly A, Bi M, Howard C, Frost A, Cheng Y, Morgan DO. (*In Review*) Reconstitution of the SARS-CoV-2 ribonucleosome provides insights into genomic RNA packaging and regulation by phosphorylation. *Journal of Biological Chemistry*

C.R.C. and A.N.A. conceived the project and performed most experiments with guidance from D.O.M.; M.B. performed negative stain analysis with guidance from Y.C.; C.R.C. and D.O.M. wrote the paper with contributions from all authors.

Abstract

Mechanisms of RNA packaging in Coronaviruses

Christopher R. Carlson

The nucleocapsid (N) protein of coronaviruses serves two major functions: compaction of the RNA genome in the virion and regulation of viral gene transcription. It is not clear how the N protein mediates such distinct functions. The N protein contains two RNA-binding domains surrounded by regions of intrinsic disorder. Phosphorylation of the central disordered region promotes the protein's transcriptional function, but the underlying mechanism is not known. Here we show that the N protein of SARS-CoV-2, together with viral RNA, forms biomolecular condensates and ordered RNA packaging units that are both regulated by phosphorylation. Unmodified N protein forms partially ordered gel-like condensates and discrete 15-nm viral ribonucleoprotein (vRNP) complexes based on multivalent RNA-protein and protein-protein interactions. Phosphorylation reduces these interactions, generating a more liquid-like droplet and a less compact vRNP. We propose 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.

Table of Contents

List of Figures

Chapter 1

Chapter 2

List of Tables

Chapter 1

Chapter 2

Chapter 1

Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions

Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions

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], Maliheh Safari[2], Alan D. Frankel[2], and David O. Morgan[1,3]

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

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

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

[4]These authors contributed equally

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 nucleocapsid. Condensation of the genome into the compact nucleocapsid structure depends primarily on the nucleocapsid (N) protein, one of the four major structural proteins encoded in the viral genome [1-3].

Following viral entry and disassembly of the nucleocapsid, the genome is translated to produce RNA-dependent RNA polymerase and numerous other non-structural proteins (Nsps) [1,4-6]. These proteins rearrange membranes of the endoplasmic reticulum to form the replication transcription complex (RTC) [7-14], 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 viral structural proteins (S, E, M, and N) [1,4-6]. 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 the most abundant viral protein in the infected cell [2,3,15] and accumulates in dynamic clusters at RTCs [11-13,16-20],

where it is thought to help promote the RNA structural rearrangements required for subgenomic transcription [21-23].

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 regions [24] (Figure 1-1A, Figure S1-1). N protein is highly basic (pI~10), and multiple RNA-binding sites are found throughout the protein [25]. The NTD is an RNA-binding domain [26-29]. The CTD forms a tightly-linked dimer with a large RNA-binding groove [30-33], and the fundamental unit of N protein structure is a dimer [34,35]. Under some conditions, the dimer self-associates to form oligomers that depend on multiple protein regions [30,32,36-41]. The biochemical features and function of these oligomers are not clear.

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 [15,22,42-48]. Phosphorylation leads to association with the RNA helicase DDX1, which promotes RNA structural changes required for transcription of long subgenomic RNAs in the RTC [22]. 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 [22,48]. We have little understanding of the molecular mechanisms by which phosphorylation influences N protein function.

To gain a better understanding of N protein function and its control by phosphorylation, we explored the biochemical properties of the N protein from SARS-CoV-2, the causative agent of the ongoing COVID-19 pandemic. We found that N protein and viral RNA form biomolecular condensates [49-53]. The ordered, gel-like condensates of the unmodified protein are changed to more liquid-like droplets by phosphorylation, providing a potential structural foundation for the distinct functions of the N protein in nucleocapsid assembly and genome processing.

Results

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

Purified N protein produced in bacteria migrated on gel filtration as a dimer in high salt but as a large oligomer in physiological salt (Fig. S1-2A). Light microscopy revealed the presence of liquid-like droplets in the presence and absence of added RNA (Fig. S1-2B), suggesting that N protein forms biomolecular condensates. We noted, however, that purified N protein contained nucleic acid, even after nuclease 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 (Fig. S1-2C), 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. 1-1B). 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 (Fig. S1-2D).

We 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 TRS [54-56]; (2) PS-318, a 318-nt sequence near the end of ORF1b in SARS-CoV, proposed as a packaging sequence [57,58] but of unknown function [54,59]; and (3) N-1260, a 1260-nt sequence 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. 1-1B). At 10 μM 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. For most subsequent studies, we used 5'-400 RNA as a representative viral RNA.

Incubation at higher temperature (30°C or 37°C) had little effect on droplet formation in a 30-minute incubation (Fig. S1-2E), and longer incubations did not transform filamentous networks into droplets (Fig. S1-2F). High salt dissolved N protein structures, indicating that they depend primarily on electrostatic interactions (Fig. S1-2G).

We analyzed the formation and morphology of condensates over a wide range of N protein and 5'-400 RNA concentrations (Fig. S1-3, 1-1C, top row). At low RNA concentrations, small spherical droplets were seen at multiple N protein concentrations.

Higher RNA and protein concentrations led to filamentous structures. Interestingly, at RNA concentrations approaching those 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 [60,61]. Surprisingly, addition of the TRS RNA triggered the rapid formation of droplets, without the filamentous structures seen with longer viral RNAs (Fig. 1-1B, 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. 1-1B, 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. 1-1C), suggesting that these droplets depend on binding of a single TRS RNA to each N protein. A 10-nt RNA with a random sequence displayed similar behavior but at higher concentrations, suggesting that specific binding is not entirely responsible for the effects of TRS RNA (Fig. 1-1C). We speculate that the binding of a short RNA molecule alters N protein structure to promote 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. 1-2A; Fig. S1-1): (1) the 44-aa N-terminal extension (NTE), a poorly conserved prion-like sequence with a basic cluster that contributes to RNA binding [25]; (2) the 31-aa SR region, a basic segment implicated in RNA binding, oligomerization [41], and phosphorylation [15,22,42-48]; (3) the 55-aa C-terminal extension (CTE), implicated in oligomerization [32,39-41]; and (4) the CTD basic patch (CBP), a 33-aa basic region that forms the RNA-binding groove on the surface of the CTD [30-32], which can be deleted without affecting CTD dimer structure [33].

When combined with viral 5'-400 RNA, none of the deletions completely prevented droplet formation at high protein concentrations (Fig. 1-2A), 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 certain 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. 1-2B) showed that full-length N protein structures increased abruptly between 5 and 10 μM, supporting a cooperative mechanism of oligomer assembly. Consistent with the microscopy analyses, 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. 1-2B).

Further insights arose in studies of deletion mutants and the 10-nt TRS RNA (Fig. 1-2A). 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 inhibited 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. 1-1B), 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. 1-2A). TRS binding to the NTD is known to be enhanced by the basic SR region, but SR deletion has only a moderate impact on affinity [60]. 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 protein [41].

Phosphorylation promotes more liquid-like N protein condensates

N protein phosphorylation depends on a poorly understood collaboration between multiple kinases [22,46,48]. A central player is the abundant cytoplasmic kinase GSK-3, which generally phosphorylates serines or threonines four residues upstream of pre-phosphorylated ‘priming’ sites [62]. Studies of the N protein of SARS-CoV [48] support the

presence of two priming sites, P1 and P2 (Fig. 1-3A), 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 kinase [46,63].

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. 1-3B). 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. 1-3C).

Phosphorylation of N protein with a combination of Cdk1 and GSK-3 reduced filamentous structures and promoted the formation of liquid-like droplets (Fig. 1-3D). GSK-3 alone had no effect, whereas Cdk1 alone promoted droplets to a small extent. Phosphorylation in the SR region therefore appears to promote the formation of more spherical, 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. 1-3E). The droplets seen with the 10D mutant were more spherical and uniform than those observed following phosphorylation in vitro (Fig. 1-3D), presumably because phosphorylation at all sites is unlikely to be achieved in vitro. All three viral RNAs were effective in driving droplet formation, although N-1260 appeared to reduce the saturating concentration (Fig. S4A). NTE deletion in the 10D protein reduced droplet formation, showing once again the positive role of this region (Fig. 1-3F).

Importantly, the 10D mutation had a greater impact on condensate morphology (Fig. 3E) than an SR deletion (Fig. 1-2A), indicating that phosphorylation does not simply block SR function but might also interfere with other interactions. One plausible 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. 1-3E), just as we observed with TRS RNA and the SR deletion (Fig. 1-2A). 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 RTC [45]. The N protein of mouse hepatitis virus (MHV) is known to interact directly with the N-terminal Ubl1 domain of Nsp3, a large transmembrane protein localized to RTC membranes [17,64-66]. We found that a GFP-tagged Ubl1 domain of SARS-CoV-2 Nsp3 partitioned into N protein droplets and filamentous structures (Fig. S1-4B), 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 μM wild-type protein remained immobile and did not fuse, while the spherical droplets of the 10D mutant were highly dynamic and fused rapidly (Fig. 1-4A). We also found that the droplet-like structures at higher concentrations of unmodified protein displayed relatively slow fusion activity compared to the droplets of phosphorylated N protein (Fig. 1-4B). These results suggest that unmodified protein forms gel-like condensates that are relatively rigid, while phosphorylated N protein forms condensates that behave more like liquid droplets.

We analyzed condensate properties further by labeling N protein with a fluorescent dye, incubating 20 μM N protein with 5'-400 RNA to form droplets, and measuring fluorescence recovery after photobleaching (FRAP). Wild-type N protein droplets recovered slowly (about 7% over 4 minutes), and recovery was most apparent at the edge of the droplets (Fig. 1-4C, S1-4C). 10D mutant droplets recovered more uniformly and about 2-fold more rapidly (12% over 4 min), suggesting that phosphorylation

generates a more dynamic condensate but also showing that both wild-type and 10D condensates are less dynamic in vitro than many previously described condensates.

We used negative-stain electron microscopy to further analyze mixtures of 10 μM N protein and 1 μM viral RNA (Fig. 1-4D). Unmodified wild-type protein and PS-318 RNA formed particles of ~15 nm diameter, and 2D classification of these particles revealed a uniform size and architecture (Fig. 1-4E). Similar particles were observed with 5'-400 RNA (Fig. S1-4D). 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, poorly-defined chains (Fig. 1-4D), suggesting that phosphorylation disrupts the structural units of the unmodified protein to create a more diffuse, 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, RNA-RNA, and protein-protein interactions. The latter might include prion-like interactions between NTEs, binding of the SR region to the CTD [41], or helical CTD polymers that depend on the CBP [30,67]. 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 condensate behavior of chromatin proteins [68,69]. Consistent with this possibility, the ribonucleoprotein particles we observed with unmodified N protein and viral RNA are remarkably similar in size and shape to nucleocapsid particles seen in electron microscopy and tomography studies of partially disrupted MHV nucleocapsids [67] and SARS-CoV-2 virions [70,71].

The more liquid-like behavior of phosphorylated N protein condensates 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 RTC [45]. N protein is the only viral structural protein that localizes to the RTC [11,18], and immunoelectron microscopy reveals high concentrations of N protein in electron-dense regions adjacent to the double-membrane vesicles and convoluted membranes of the RTC [11-13]. At these sites, N protein co-localizes with the transmembrane viral protein Nsp3, which is known to bind N protein

and might therefore anchor N protein oligomers to RTC membranes [14,17,64-66] (Fig. S1-4B). In light microscopy of infected cells, GFP-tagged N protein forms large cytoplasmic granules that co-localize with Nsp3 at RTCs [12,16,17,20], and FRAP indicates that N protein is dynamically associated with these granules [19]. N protein helps control subgenomic transcription in the RTC, and inhibition of phosphorylation blocks this function [21-23]. With these many lines of evidence in mind, our work points to the possibility that a condensate 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 condensates [72-74]. In Measles virus, these condensates have also been implicated in nucleocapsid assembly [75].

Others have also observed recently that the N protein of SARS-CoV-2 forms RNA-dependent biomolecular condensates in vitro [76-82]. When overexpressed at high levels in uninfected cells, the N proteins of MHV [20], SARS-CoV [44], and SARS-CoV-2 [78,80] form cytoplasmic granules. In the case of SARS-CoV-2 N protein, these granules display the fusion and rapid turnover behaviors of dynamic condensates [78,80]. Most importantly, as mentioned above, the granules of N protein that form at the RTC in infected cells also display rapid turnover [19].

FRAP reveals that N protein granules in the cell turn over much more rapidly than N protein condensates in vitro (Fig. 1-4C) [19,78,80]. Phosphorylation in vivo does not entirely

explain this discrepancy, as we found that the 10D mutant still displays relatively slow turnover. We speculate that N protein dynamics in the cell, and particularly in the infected cell, are further governed by cellular and viral proteins and RNA.

There is some evidence to suggest a link between N protein and stress granules, the cytoplasmic mRNA-protein condensates that form in response to translation arrest and contribute to the response to some viral infections [83-86]. In uninfected cells, overexpressed N protein of SARS-CoV [44] or SARS-CoV-2 [82] associates with stress granules, and N protein associates in vitro with condensates of stress granule proteins [81]. Recent proteomic analyses indicate that the N protein of SARS-CoV-2, when expressed in uninfected cells, associates with proteins found in stress granules [87-90] and appears to suppress stress granule formation [89]. Studies of infected cells will be required, however, to determine if stress granules form during SARS-CoV-2 infection and whether N protein has any role in their regulation.

Consistent with our results, N protein granules in MHV-infected cells display dramatic changes in morphology following chemical inhibition of GSK-3 kinase activity [22]. Most importantly, GSK-3 inhibition disrupts MHV genome processing and reduces the production of virions by MHV- or SARS-CoV-infected cells [22,48]. Cdk inhibitors inhibit virus production by SARS-CoV-2-infected cells [15]. Thus, GSK-3 inhibitors, together with inhibitors of priming kinases such as Cdks, 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 baculoviruses [91]. Human GSK-3□ was purchased from Promega (V1991).

Sequences of the three viral RNAs and firefly luciferase RNA are provided in Table S1-1. 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 μg/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 LN_2, and stored at -80°C until lysis. To remove contaminating nucleic acid, N protein was purified under denaturing conditions as previously described [44,92]. 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 3 times with 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 LN_2, 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 3 times with 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 LN_2, 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 LN_2, 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 LN_2. 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 lysate [93], the two lysates were combined, brought to 5 mM ATP and 10 mM $MgCl_2$, 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 LN_2.

Light microscopy

Glass was prepared as described previously [94]. 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 ddH_2O. 1 M NaOH was added to the glass for 30 min, followed by washing 3 times with 100 μl ddH_2O. 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 ddH_2O 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 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 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. 1-3C, 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, 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.

Fluorescence recovery after photobleaching (FRAP)

Purified N protein (wild type or 10D) was labeled by incubation with 3:1 Alexa Fluor™ 488 NHS Ester (Invitrogen A20000) at 30°C for 45 min. N protein was then purified by size exclusion chromatography, concentrated, and stored at -80°C until use. For FRAP experiments, glass coverslips and slides were cleaned and functionalized with

polyethylene glycol as previously described [95], and chambers were constructed with double-sided tape. 20 μM N protein, diluted to 5% labeled N with unlabeled protein, was mixed with 1 μM 5'-400 RNA and immediately added to preassembled glass chambers. Fluorescent droplets were visualized by Total Internal Reflection Fluorescence on a DeltaVision OMX SR microscope with a 60x objective (immersion oil refractive index: 1.516). Images were acquired at a 1-s frame rate for 3 frames, bleached 30 s at 5% absolute transmission at 405 nm, and imaged immediately thereafter. All experiments were performed on droplets of similar size (2.3-2.6 μm), 15-30 min after RNA addition.

Electron microscopy

N protein (10 μM) and RNA (1 μM PS-318 or 5'-400) 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 CTFFIND4 [96]. Automated particle picking, CTF-correction, and two-dimensional averaging and classification were performed in RELION-3 [97].

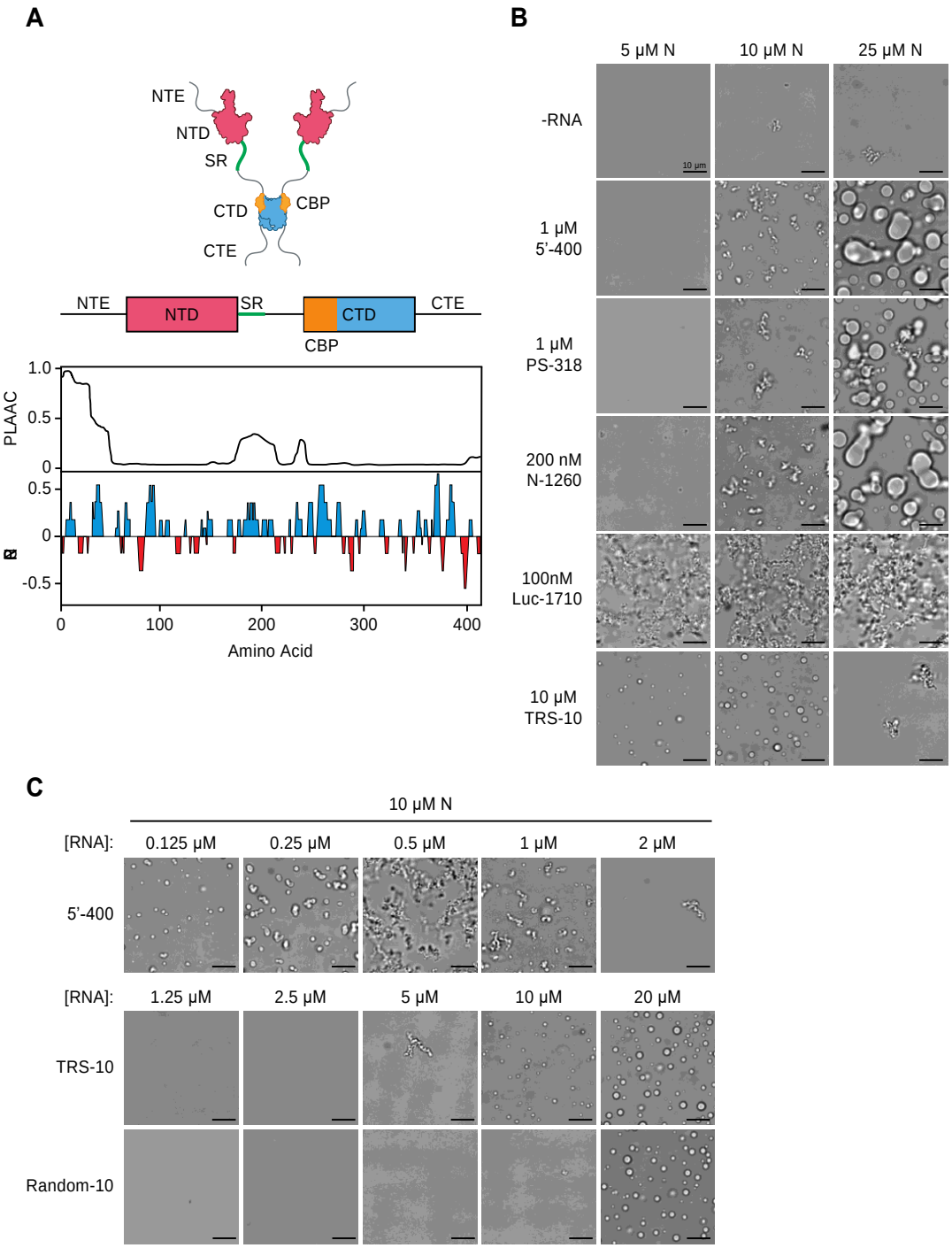

Fig 1-1. SARS-CoV-2 N protein forms biomolecular condensates in the presence of RNA.

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 composition [98], NCPR, net charge per residue. See S1-1 for sequence. **b**

Light microscopy images of N protein condensates after 30 min incubation at room temperature with the indicated RNA molecules. All images are representative of multiple independent experiments; scale bar, 10 μm. **c** Condensate formation by N protein (10 μM) over a range of 5'-400, TRS-10, and Random-10 RNA concentrations. All images are representative of multiple independent experiments; scale bar, 10 μm. Random-10 RNA is a mixed population of 10-nucleotide random sequences. See S1-3 for condensate formation at other N protein and 5'-400 RNA concentrations.

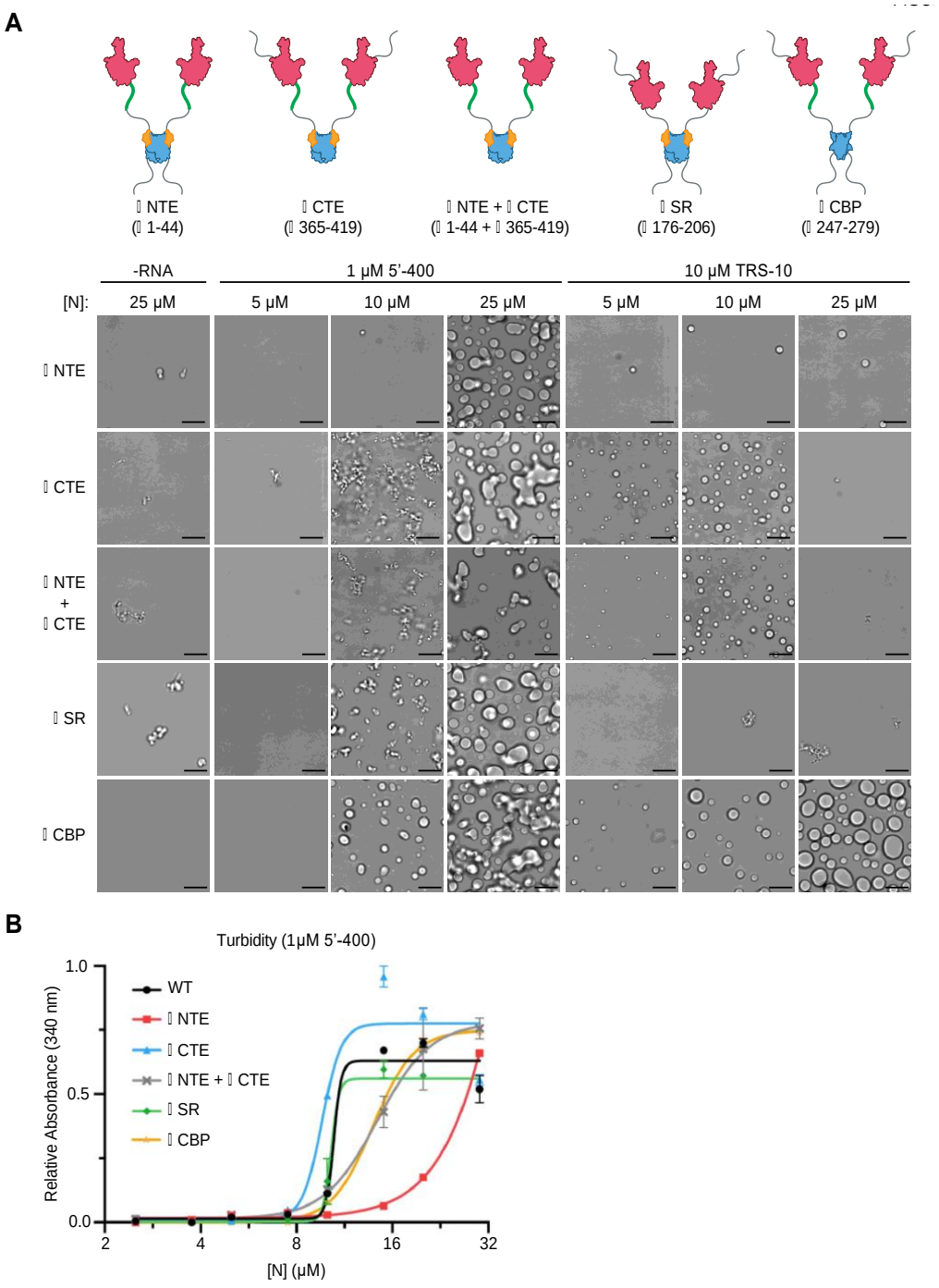

Fig 1-2. Disordered regions modulate N protein condensate formation.

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 μm. **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 μM 5'-400. Data points indicate mean +/- s.e.m. of duplicates; representative of two independent experiments.

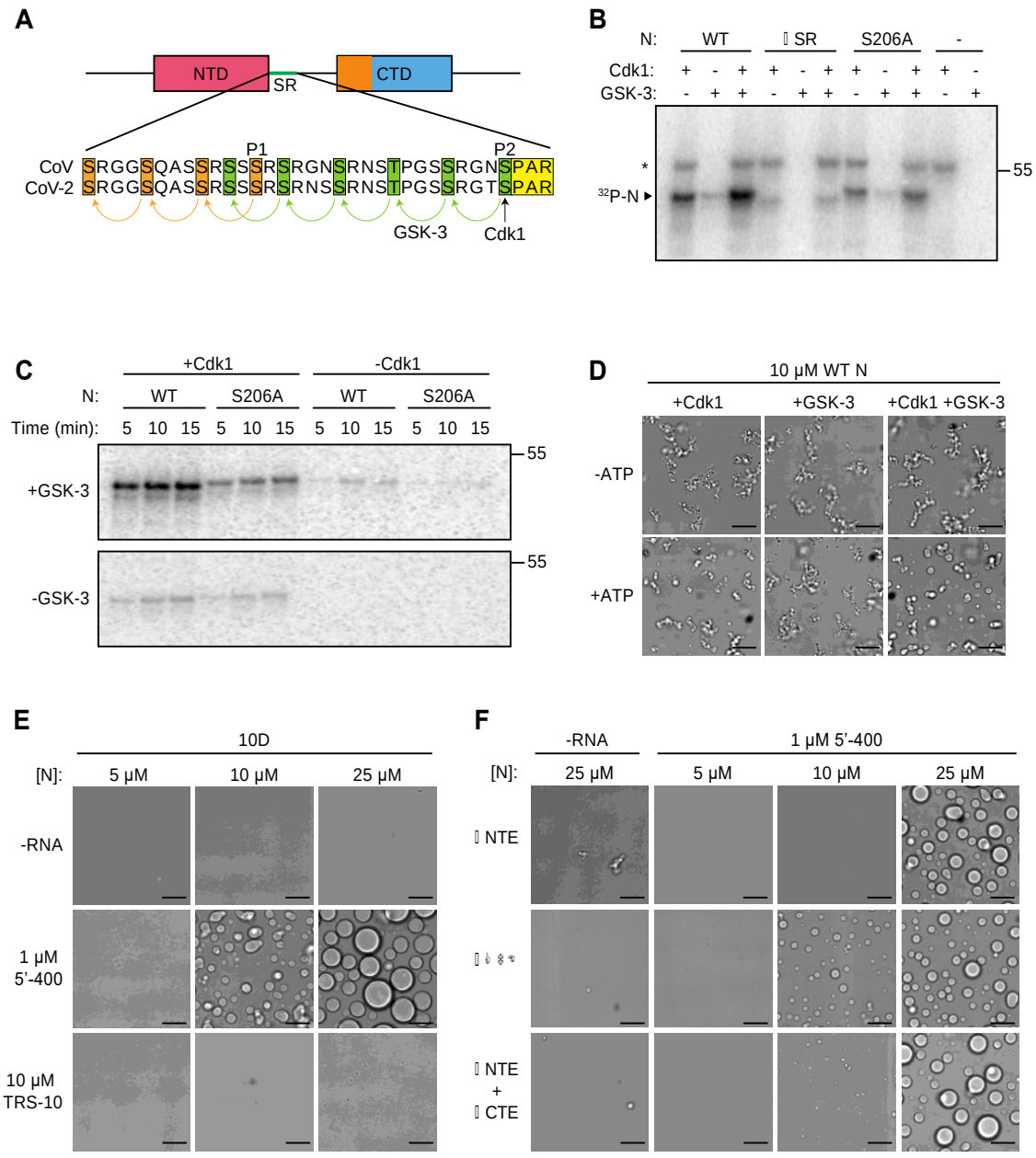

Fig 1-3. Phosphorylation modulates N protein condensate properties.

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 indicated [48]. 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. Molecular weight marker shown on right (kDa). **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. Molecular weight marker shown on right (kDa). **d** 10 μM 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 mixed with 1 μM 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 μM 5'-400 or 10 μM TRS-10 RNA. **f** Images of 10D mutants with the indicated deletions, incubated with or without 1 μM 5'-400 RNA. **b-f** All results are representative of multiple independent experiments; scale bar, 10 μm.

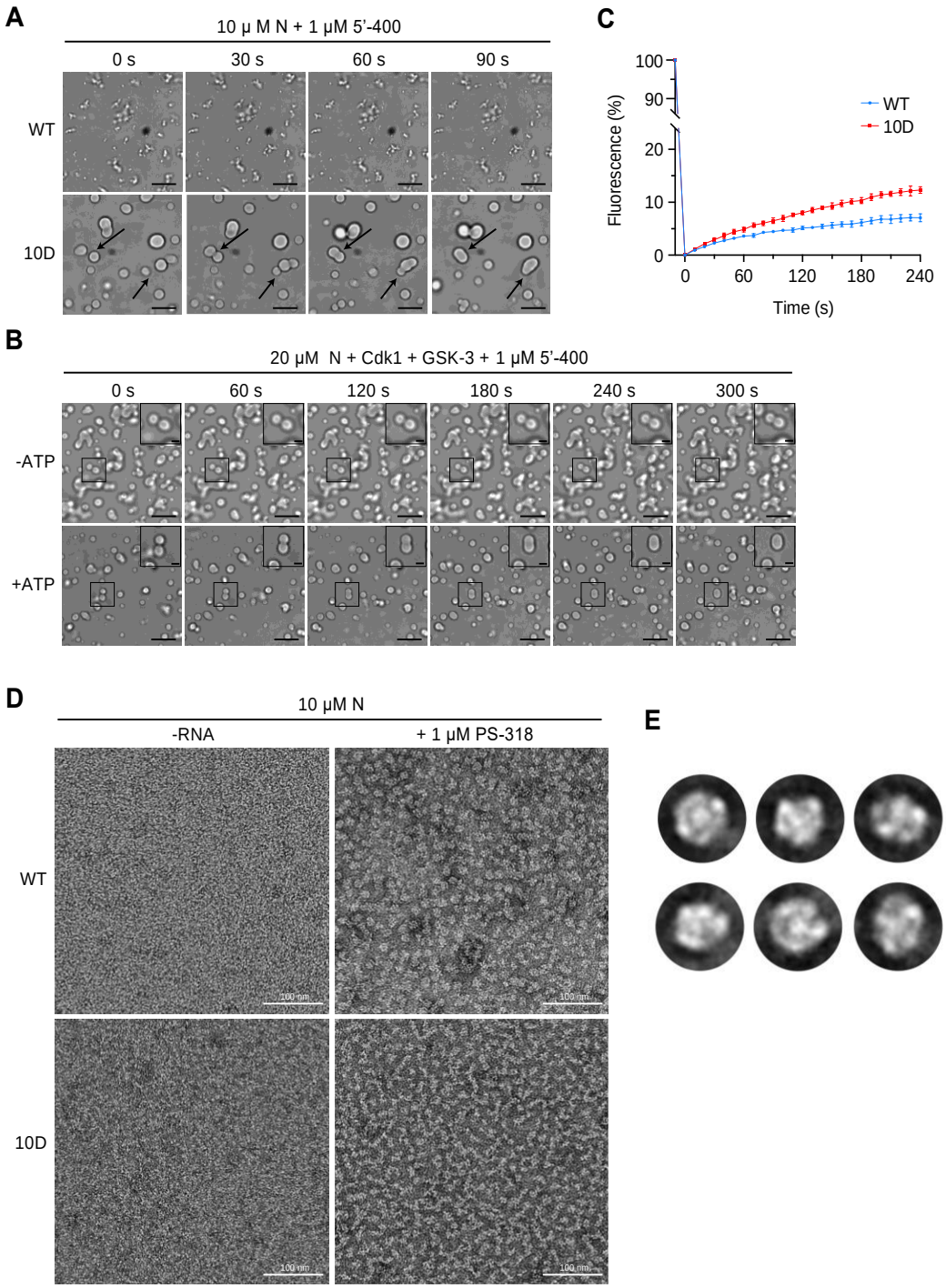

Fig 1-4. Phosphorylation of N protein promotes liquid-like behavior.

a N protein (10 μM wild-type [WT] or 10D) was mixed with 1 μM 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 μM N protein was phosphorylated with Cdk1-cyclin B1 and GSK-3 as in 1-3D, incubated with 1 μM 5'-400 RNA for 20 min, and imaged every 60 s. Representative droplet fusion events are

boxed, with higher magnification images in insets at upper right. Images are representative of multiple independent experiments; scale bar, 10 μm (2 μm for insets). **c** FRAP analysis of droplets formed with 20 μM dye-labeled N protein (WT or 10D) and 1 μM 5'-400 RNA. Following 30 s of bleaching, droplet fluorescence was measured starting at time zero. Data points indicate mean fluorescence intensity as a percentage of pre-bleaching intensity (mean +/- s.e.m., n=2 for WT and n=3 for 10D). Note break in y axis to allow better viewing of recovery data. Results are representative of two independent experiments. See S1-4C for images.**d** 10 μM N protein (WT or 10D) was incubated without or with 1 μM 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. **e** 2D class averages of particles from the EM analysis of WT N protein and RNA in panel **d**. Particle selection was not possible with the nonuniform structures formed by the 10D mutant.

Supplemental Information

```
CoV     1    MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALTQH  60
CoV-2   1    MSDNGPQ-NQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQH  59
                                  NTE

CoV     61   GKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTGPEA  120
CoV-2   60   GKEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEA  119

CoV     121  SLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLPKGFYAEGSRGG  180
CoV-2   120  GLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGG  179

CoV     181  SQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESKVSGKGQ  240
CoV-2   180  SQASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQ  239
                       SR

CoV     241  QQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYK  300
CoV-2   240  QQQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYK  299
                                  CBP

CoV     301  HWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDA  360
CoV-2   300  HWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDA  359

CoV     361  YKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADST  420
CoV-2   360  YKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMS--SADST  417
                                                  CTE

CoV     421  QA  422
CoV-2   418  QA  419
```

Fig S1-1. 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.

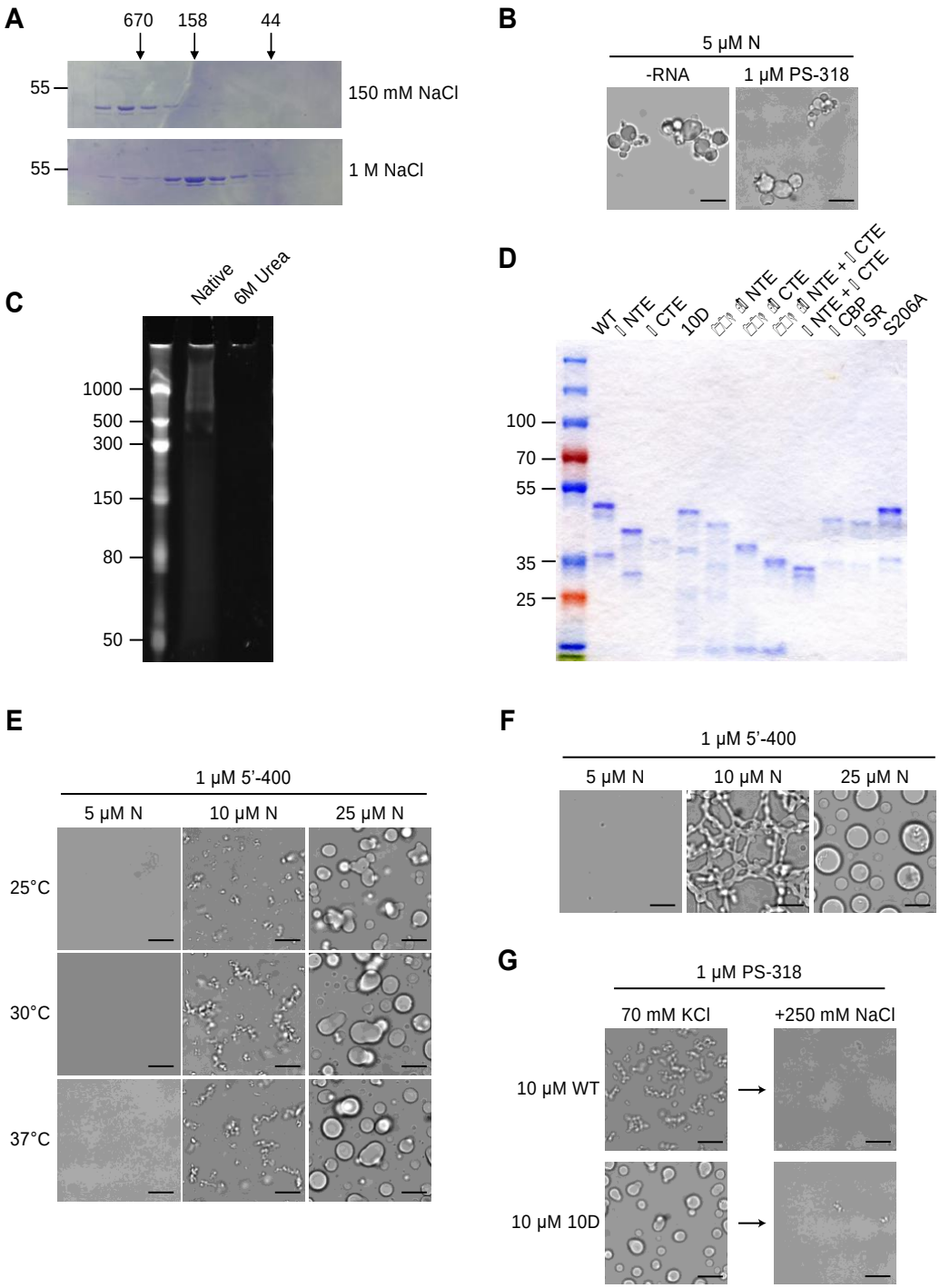

Fig S1-2. Characterization of N protein behavior.

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 μM native N protein after incubation in the presence or absence of 1 μM PS-318 RNA. Scale bar, 10 μm. **c** 5 μl of 10 μM N protein

(purified under native or denaturing [6M Urea] conditions) was mixed with urea loading buffer, incubated at 65°C for 5 min, and separated on a 10% polyacrylamide TBE-Urea gel, followed by staining with SYBR Gold. RNA length standards (nt) in left lane. **d** SDS-PAGE analysis of all N protein mutants used in this study, stained with Coomassie Blue. **e** N protein was incubated at the indicated temperature for 30 min in the presence of 1 μM 5'-400 RNA. Scale bar, 10 μm. **f** N protein was incubated with 1 μM 5'-400 for 16 h at room temperature. Scale bar, 10 μm. **g** 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.

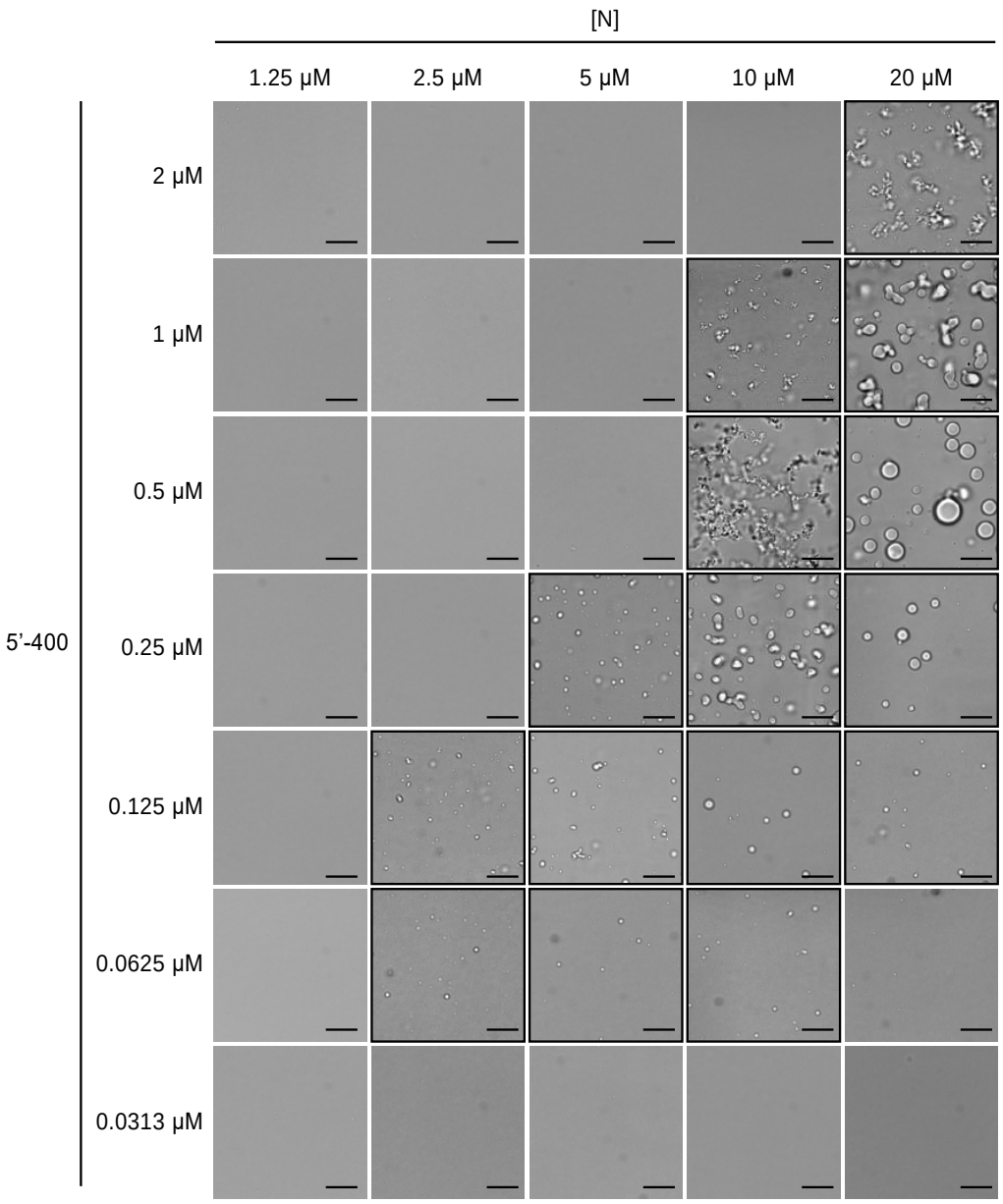

Fig S1-3. Condensate formation at a range of N protein and RNA concentrations.

Light microscopy images of N protein condensates after 30 min incubation at room temperature with the indicated concentrations of 5'-400 RNA. Images that contain visible condensates are highlighted with a black border. Scale bar, 10 μm.

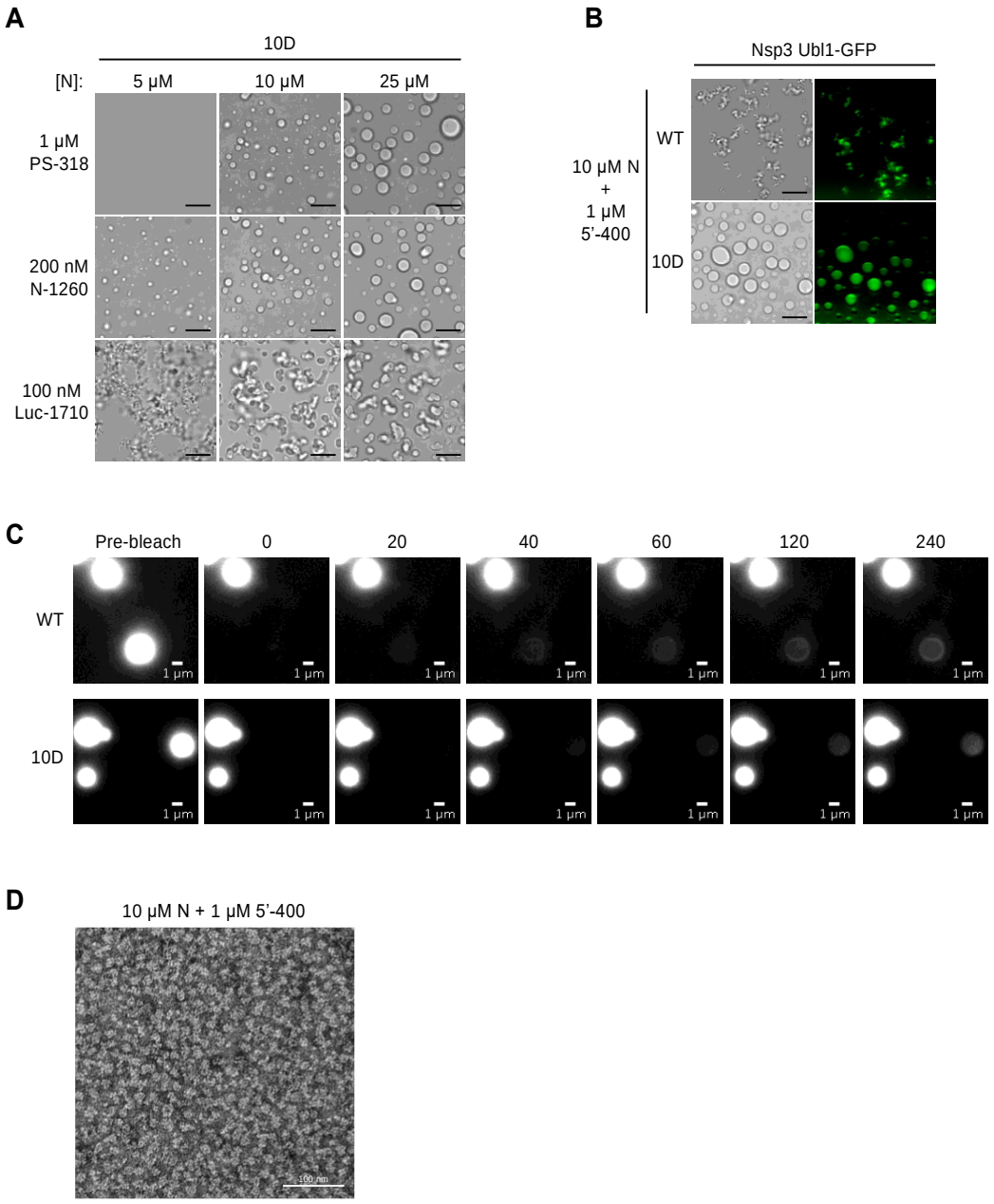

Fig S1-4. Properties of wild-type N protein and 10D mutant.

a Images of N protein 10D mutant following 30 min incubation with the indicated RNAs. Scale bar, 10 μm. b 10 μM wild-type (WT) or 10D N protein was incubated with 1 μM 5'-400 RNA for 10 min. Nsp3 Ubl1-GFP was then added to a concentration of 1 μM and incubated for an additional 15 min before imaging in brightfield (left) or fluorescence (right). **c** Images of droplets formed with 20 μM fluorescent N protein (WT or 10D) and 1 μM 5'-400 RNA, before bleaching and at the indicated times after bleaching. Scale bar, 1 μm. **d** 10 μM wild-type N protein was incubated with 1 μM 5'-400 RNA for 15 min prior to analysis by negative-stain electron microscopy. Scale bar, 100 nm.

Table S1-1. RNA sequences used in this study.

5'-400 5' 400 nt from SARS-CoV-2 (Wuhan Hu-1 strain; nt 1-400).	AUUAAAGGUUUAUACCUUCCCAGGUAACAAACCAACCAACUUUCGAU CUCUUGUAGAUCUGUUCUCUAAACGAACUUUAAAAUCUGUGUGGCU GUCACUCGGCUGCAUGCUUAGUGCACUCACGCAGUAUAAUUAAUAA CUAAUUACUGUCGUUGACAGGACACGAGUAACUCGUCUAUCUUCUG CAGGCUGCUUACGGUUUCGUCCGUGUUGCAGCCGAUCAUCAGCACA UCUAGGUUUCGUCCGGGUGUGACCGAAAGGUAAGAUGGAGAGCCUU GUCCCUGGUUUCAACGAGAAAACACACGUCCAACUCAGUUUGCCUG UUUUACAGGUUCGCGACGUGCUCGUACGUGGCUUUGGAGACUCCG UGGAGGAGGUCUUAUCAGAGGCACGUCAACAU
PS-318 318 nt from SARS CoV (Tor2 strain; nt 19715-20031), with an extra C (red) after A19802 as in: [58].	UGAGCUUUGGGCUAAGCGUAACAUUAAACCAGUGCCAGAGAUUAAG AUACUCAAUAAUUUGGGUGUUGAUAUCGCUGCUAAUACUGUA**C**AUC UGGGACUACAAAAGAGAAGCCCCAGCACAUGUAUCUACAAUAGGUG UCUGCACAAUGACUGACAUUGCCAAGAAACCUACUGAGAGUGCUUG UUCUUCACUUACUGUCUUGUUUGAUGGUAGAGUGGAAGGACAGGUA GACCUUUUUAGAAACGCCCGUAAUGGUGUUUUAAUAACAGAAGGUU CAGUCAAAGGUCUAACACCUUCAAAGGGACCAGCACAAGCUA
N-1260 RNA containing the 1260 nt open reading frame (gray highlight) of the N gene from SARS-CoV-2 (Wuhan Hu-1 strain; nt 28274-29533), plus flanking plasmid sequence.	GGGAAUUGUGAGCGGAUAACAAUUCCCCUCUAGAAAUAAUUUUGUU UAACUUUAAGAAGGAGAUAUACCAUGGGCAGCAGCCAUCAUCAUCAU CAUCACAGCAGCGGCCUGGUGCCGCGCGGUACCACGGAAAACCUGU AUUUUCAGGGAUCCAUGUCUGAUAAUGGACCCCAAAAUCAGCGAAAU GCACCCCGCAUUACGUUUGGUGGACCCUCAGAUUCAACUGGCAGUA ACCAGAAUGGAGAACGCAGUGGGGCGCGAUCAAAACAACGUCGGCC CCAAGGUUUACCCAAUAAUACUGCGUCUUGGUUCACCGCUCUCACU CAACAUGGCAAGGAAGACCUUAAAUUCCCUCGAGGACAAGGCGUUC CAAUUAACACCAAUAGCAGUCCAGAUGACCAAAUUGGCUACUACCGA AGAGCUACCAGACGAAUUCGUGGUGGUGACGGUAAAAUGAAAGAUC UCAGUCCAAGAUGGUAUUUCUACUACCUAGGAACUGGGCCAGAAGC UGGACUUCCCUAUGGUGCUAACAAAGACGGCAUCAUAUGGGUUGCA ACUGAGGGAGCCUUGAAUACACCAAAAGAUCACAUUGGCACCCGCAA UCCUGCUAACAAUGCUGCAAUCGUGCUACAACUUCCUCAAGGAACAA CAUUGCCAAAAGGCUUCUACGCAGAAGGGAGCAGAGGCGGCAGUCA AGCCUCUUCUCGUUCCUCAUCACGUAGUCGCAACAGUUCAAGAAAU UCAACUCCAGGCAGCAGUAGGGGAACUUCUCCUGCUAGAAUGGCUG GCAAUGGCGGUGAUGCUGCUCUUGCUUUGCUGCUGCUUGACAGAU UGAACCAGCUUGAGAGCAAAAUGUCUGGUAAAGGCCAACAACAACAA GGCCAAACUGUCACUAAGAAAUCUGCUGCUGAGGCUUCUAAGAAGC CUCGGCAAAAACGUACUGCCACUAAAGCAUACAAUGUAACACAAGCU UUCGGCAGACGUGGUCCAGAACAAACCCAAGGAAAUUUUGGGGACC AGGAACUAAUCAGACAAGGAACUGAUUACAAACAUUGGCCGCAAAUU GCACAAUUUGCCCCCAGCGCUUCAGCGUUCUUCGGAAUGUCGCGCA UUGGCAUGGAAGUCACACCUUCGGGAACGUGGUUGACCUACACAGG UGCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUCAAAGAUCAAGUCA UUUUGCUGAAUAAGCAUAUUGACGCAUACAAAACAUUCCCACCAACA GAGCCUAAAAAGGACAAAAAGAAGAAGGCUGAUGAAACUCAAGCCUU ACCGCAGAGACAGAAGAAACAGCAAACUGUGACUCUUCUUCCUGCU GCAGAUUUGGAUGAUUUCUCCAAACAAUUGCAACAAUCCAUGAGCAG UGCUGACUCAACUCAGGCCUAAGAAUUCGAGCUCCGUCGACA

Luc-1710 RNA containing the 1710 nt firefly luciferase open reading frame (gray highlight) plus flanking plasmid sequence.	GGUCUAGAAAUAAUUUUGUUUAACUUUAAGAAGGAGAUAUAACCAUG AAAAUCGAAGAAGGUAAAGGUCACCAUCACCAUCACCACGGAUCCAU GGAAGACGCCAAAAACAUAAAGAAAGGCCCGGCGCCAUUCUAUCCUC UAGAGGAUGGAACCGCUGGAGAGCAACUGCAUAAGGCUAUGAAGAG AUACGCCCUGGUUCCUGGAACAAUUGCUUUUACAGAUGCACAUAUC GAGGUGAACAUCACGUACGCGGAAUACUUCGAAAUGUCCGUUCGGU UGGCAGAAGCUAUGAAACGAUAUGGGCUGAAUACAAAUCACAGAAUC GUCGUAUGCAGUGAAAACUCUCUUCAAUUCUUUAUGCCGGUGUUGG GCGCGUUAUUUAUCGGAGUUGCAGUUGCGCCCGCGAACGACAUUUA UAAUGAACGUGAAUUGCUCAACAGUAUGAACAUUUCGCAGCCUACC GUAGUGUUUGUUUCCAAAAAGGGGUUGCAAAAAAUUUUGAACGUGC AAAAAAAAUUACCAAUAAUCCAGAAAAUUAUUAUCAUGGAUUCUAAAA CGGAUUACCAGGGAUUUCAGUCGAUGUACACGUUCGUCACAUCUCA UCUACCUCCCGGUUUUAAUGAAUACGAUUUUGUACCAGAGUCCUUU GAUCGUGACAAAACAAUUGCACUGAUAAUGAAUUCCUCUGGAUCUAC UGGGUUACCUAAGGGUGUGGCCCUUCCGCAUAGAACUGCCUGCGU CAGAUUCUCGCAUGCCAGAGAUCCUAUUUUUGGCAAUCAAAUCAUU CCGGAUACUGCGAUUUUAAGUGUUGUUCCAUUCCAUCACGGUUUUG GAAUGUUUACUACACUCGGAUAUUUGAUAUGUGGAUUUCGAGUCGU CUUAAUGUAUAGAUUUGAAGAAGAGCUGUUUUUACGAUCCCUUCAG GAUUACAAAAUUCAAAGUGCGUUGCUAGUACCAACCCUAUUUUCAUU CUUCGCCAAAAGCACUCUGAUUGACAAAUACGAUUUAUCUAAUUUAC ACGAAAUUGCUUCUGGGGGCGCACCUCUUUCGAAAGAAGUCGGGGA AGCGGUUGCAAAACGCUUCCAUCUUCCAGGGAUACGACAAGGAUAU GGGCUCACUGAGACUACAUCAGCUAUUCUGAUUACACCCGAGGGGG AUGAUAAACCGGGCGCGGUCGGUAAAGUUGUUCCAUUUUUUGAAGC GAAGGUUGUGGAUCUGGAUACCGGGAAAACGCUGGGCGUUAAUCAG AGAGGCGAAUUAUGUGUCAGAGGACCUAUGAUUAUGUCCGGUUAUG UAAACAAUCCGGAAGCGACCAACGCCUUGAUUGACAAGGAUGGAUG GCUACAUUCUGGAGACAUAGCUUACUGGGACGAAGACGAACACUUC UUCAUAGUUGACCGCUUGAAGUCUUUAAUUAAAUACAAAGGAUAUCA GGUGGCCCCCGCUGAAUUGGAAUCGAUAUUGUUACAACACCCCAAC AUCUUCGACGCGGGCGUGGCAGGUCUUCCCGACGAUGACGCCGGU GAACUUCCCGCCGCCGUUGUUGUUUUGGAGCACGGAAAGACGAUGA CGGAAAAAGAGAUCGUGGAUUACGUCGCCAGUCAAGUAACAACCGC GAAAAAGUUGCGCGGAGGAGUUGUGUUUGUGGACGAAGUACCGAAA GGUCUUACCGGAAAACUCGACGCAAGAAAAAUCAGAGAGAUCCUCAU AAAGGCCAAGAAGGGCGGAAAGUCCAAACUCGAGUAAGGUUAACCU GCAGGAGG

Chapter 2

Reconstitution of the SARS-CoV-2 ribonucleosome provides insights into genomic RNA packaging and regulation by phosphorylation

Reconstitution of the SARS-CoV-2 ribonucleosome provides insights into genomic RNA packaging and regulation by phosphorylation

Christopher R. Carlson[1], Armin N. Adly[1], Maxine Bi[2], Conor J. Howard[2], Adam Frost[2], Yifan Cheng[2], and David O. Morgan[1]

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

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

Introduction

At different stages of the viral life cycle, viral genomes switch between two distinct structural states: a tightly-packaged state inside the virion and a decondensed state that serves as a substrate for translation, transcription, or other processes in the infected cell. The mechanisms that govern the switch between these states are not well understood.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the COVID-19 pandemic, is a highly contagious betacoronavirus[99]. The ~30-kb single-stranded RNA genome is packed inside the virus in a structure called the nucleocapsid[1,3]. Following infection and genomic RNA unpackaging, the first two-thirds of the genome is translated to produce numerous non-structural proteins (Nsps) that rearrange host cell membranes to establish the replication-transcription complex (RTC), a network of double-membrane vesicles that scaffolds viral genome replication and transcription[11,100-102]. The final third of the genome then serves as a template for generation of the four structural proteins that form the mature virus[4,6,103].

Transcription of structural protein genes by the viral RNA-dependent RNA polymerase generates negative-sense subgenomic RNAs through a template switching mechanism. These RNAs are then transcribed to positive-sense RNAs, which are translated to produce the Spike (S), Membrane (M), Envelope (E), and Nucleocapsid (N) proteins[6]. The S, M, and E proteins contain transmembrane domains that insert into the ER membrane, while the N protein localizes at high concentrations in the cytosol at the

RTC and at nearby sites of viral assembly[11,12,16-18,100]. N protein is the most abundant viral protein in an infected cell[15] and serves two major functions in the coronavirus life cycle. First, it is critical for compaction of the viral RNA genome into the nucleocapsid structure within the virion. Second, the N protein has a poorly understood role in viral transcription at the RTC[21-23,104].

The 46 kDa N protein contains two globular domains flanked by three regions of intrinsic disorder (Fig. 2-1a)[24]. The N-terminal domain (NTD) and the C-terminal domain (CTD) bind RNA and are highly conserved among coronaviruses[26,28,30,34,105,106]. In solution, N protein exists predominantly as a dimer due to a high-affinity dimerization interface on the CTD but also forms tetramers and higher-order oligomers that are modulated by the disordered N-terminal extension (NTE) and C-terminal extension (CTE, Fig. 2-1a)[30,34,36-38,105]. The central disordered region contains a conserved serine/arginine (SR)-rich sequence, which is extensively phosphorylated in the cytoplasm of infected cells but not inside the virion[15,22,44,48,107]. The central disordered region also contains sequences that interact with Nsp3, a transmembrane protein that spans both membranes of double-membrane vesicles at the RTC[14,17,108,109].

Purified N protein and viral RNA form biomolecular condensates in vitro[110-113]. Unphosphorylated N protein forms gel-like condensates containing discrete substructures, perhaps reflecting N protein function in the nucleocapsid[110,112,113]. Phosphorylated N forms more liquid-like condensates that are reminiscent of dynamic N protein foci seen at the RTC[19,110,113]. Phosphorylated N protein condensates might

therefore provide a local compartment that facilitates viral RNA processing at the RTC, but this possibility remains untested.

During viral assembly, hypo-phosphorylated N protein binds genomic RNA to form the compact nucleocapsid structure, which is then engulfed by ER membranes containing the S, E, and M proteins to form a mature virus[22,48,100,101]. Early electron microscopy studies of coronavirus nucleocapsids demonstrated the existence of viral ribonucleoprotein (vRNP) complexes aligned helically along an RNA strand[114-116]. Recent cryo-electron tomography studies of intact SARS-CoV-2 virions revealed that each virus contains 35-40 discrete, cylindrical nucleosome-like vRNP complexes[117,118]. These vRNPs, or ribonucleosomes, are ~15 nm in diameter and, through low resolution modeling efforts, are speculated to contain twelve N proteins in complex with up to 800 nt of RNA (30,000 nt ÷ 38 vRNPs = 800 nt). A ‘beads-on-a-string’ model has been proposed as a general mechanism of coronavirus packaging: vRNPs (the beads) locally compact RNA within the long genomic RNA strand (the string).

Unlike string, however, the SARS-CoV-2 genomic RNA is highly structured, containing an elaborate array of heterogeneous secondary and tertiary structural elements that are present in both infected cells and in the virion[119,120]. Thus, N protein must accommodate a variety of RNA structural elements to form the compact vRNPs of the nucleocapsid. Mechanistic insight into this model and overall vRNP architecture is lacking.

In our previous work, we observed that purified, unphosphorylated N protein and a 400-nt viral RNA fragment assemble into vRNP particles similar to those seen inside the intact virus, suggesting that N protein and RNA alone are sufficient to form the vRNP[113]. Here, we explore the biochemical properties, composition, and regulation of these particles. We find that vRNPs form in the presence of stem-loop-containing RNA though a multitude of protein-protein and protein-RNA interactions. Phosphorylation of N protein weakens these interactions to reshape vRNP structure, providing insights into the mechanisms by which N protein switches between its two major functions.

Results

Stem-loop-containing RNA promotes ribonucleosome formation

We previously observed vRNP complexes in vitro when unphosphorylated N protein was mixed with a 400-nt viral RNA from the highly structured 5’ end of the genome, while cryo-electron tomography studies of intact viruses suggest that the vRNP packages up to 800 nt of RNA[113,117,118]. To further investigate the impact of RNA length on vRNP assembly, we mixed N protein with 400-, 600-, and 800-nt RNA fragments from the 5’ end of the genome (5’-400, 5’-600, and 5’-800, respectively) and analyzed the resulting complexes by electrophoresis on a native TBE gel. All three RNAs shifted to a larger species in the presence of N (Fig. 2-1b), indicating that N protein bound the RNAs and retarded their electrophoretic mobility. N protein in complex with 5’-600 RNA resulted in a particularly discrete, intense band, suggesting that it forms a stable RNA-N protein complex.

We used mass photometry to better characterize these RNA-N protein complexes. Mass photometry uses light scattering to measure the mass of single molecules in solution, resulting in a histogram of mass measurements centered around the average molecular mass of the protein complex. N protein in complex with 5'-400 RNA resulted in two mass peaks that were smaller than the single broad peak of N protein bound to 5'-600 RNA, suggesting the 5'-400 vRNP was not fully assembled and contained subcomplexes (Fig. 2-1c). N protein mixed with 5'-800 RNA formed two broad peaks: one smaller peak that appears similar in size to the 5'-600 species (both ~750-800 kDa) and a second larger peak roughly twice as large as the first (~1400 kDa). This suggests that one (~750 kDa) or two vRNPs (~1400 kDa) can form on a single 5'-800 RNA molecule (Fig. 2-1c). The 5'-600 RNA was therefore chosen as a representative viral RNA to further study the ribonucleosome.

To purify the vRNP complex for more detailed analysis, N protein was mixed with 5'-600 RNA and separated by centrifugation on a 10-40% glycerol gradient. Individual fractions were analyzed by native gel electrophoresis (Fig. 2-1d, top), after which fractions 7 and 8 were combined for analysis by mass photometry. We observed three major peaks centered at 97 ± 2 kDa, 207 ± 6 kDa, and 766 ± 6 kDa (Fig. 2-1e, top and Table S2-1). These peaks likely correspond to free N protein dimer (predicted mass 91.2 kDa; see Fig. 2-2d, top), unbound 5'-600 RNA (predicted mass 192.5 kDa), and the vRNP complex, respectively. The presence of free N protein dimer and unbound RNA suggests that the vRNP complex dissociated upon dilution for mass photometry analysis.

To stabilize the complex, a crosslinker (0.1% glutaraldehyde) was added to the 40% glycerol buffer, creating a gradient of glutaraldehyde throughout the glycerol to crosslink the protein complex during centrifugation (a technique known as gradient fixation, or GraFix)[121]. Analysis of the GraFix-purified fractions by native gel electrophoresis revealed sharper, more discrete bands compared to the non-crosslinked sample (Fig. 2-1d, bottom). The distribution of vRNP complexes across the gradient was similar in the two conditions. The GraFix purified sample (fractions 7 + 8) was analyzed by mass photometry, revealing one sharp peak with an approximate mass of 719 ± 7 kDa (Fig. 2-1e, bottom and Table S2-1). This is consistent with the idea that the non-crosslinked sample dissociates upon dilution for mass photometry and suggests a likely stoichiometry of 12 N proteins (547.5 kDa) bound to one 5'-600 RNA (192.5 kDa; total predicted mass: 740 kDa). Alternatively, a complex of 8 N proteins (365 kDa) with two 5'-600 RNAs (385 kDa; total predicted mass: 750 kDa) is consistent with these results.

Negative stain electron microscopy (EM) of the GraFix-purified sample revealed discrete 15-nm particles with an electron-dense center surrounded by an outer ring (Fig. 2-1f). Two-dimensional classification revealed particles with variable composition and conformation, suggesting inherent structural heterogeneity in the vRNP complex that may result from the diverse RNA secondary structures in the 600 nt RNA strand (see Fig. 2-2a). While these averages are heterogeneous, they are similar in size and shape to vRNP complexes previously observed within SARS-CoV-2 virions by cryo-electron tomography[117,118].

We next tested if specific RNA sequences or regions of the genome promote formation of the vRNP. Four 600-nt genomic regions were transcribed in vitro, individually mixed with N protein, and analyzed by native gel electrophoresis: (1) 5'-600 (nucleotides 1–600), (2) Nsp3 (nucleotides 7,800–8,400), (3) Nsp8/9 (nucleotides 12,250–12,850), (4) Nsp10/12 (nucleotides 13,200–13,800). All RNAs appeared to form vRNPs, although the Nsp3 RNA appeared less effective (Fig. 2-1g). Each of these regions is known to contain a diverse array of secondary structure in infected cells and in the mature virion[119,120]. Thus, these results suggest that the vRNP can accommodate a variety of viral RNA and does not require specific sequences to form, although certain sequences or secondary structures might form more stable ribonucleosomes.

We next dissected the structures required for vRNP assembly within the 5'-600 RNA. This highly structured 600 nt genomic region contains several well-characterized stem-loops varying in size from 20 to ~150 nt (Fig. 2-2a)[55,119]. Three stem-loop RNAs (SL4a, 56 nt; SL7, 46 nt; SL8, 72 nt) were individually mixed with N protein, crosslinked with 0.1% glutaraldehyde to stabilize the resulting complexes and assessed for vRNP formation by native gel electrophoresis. vRNP complexes formed in the presence of all three stem-loops (Fig. 2-2b and Fig. S2-1a). Each crosslinked complex was analyzed by mass photometry. SL4a mixtures contained three broad peaks of 515 ± 19 kDa, 739 ± 5 kDa, and 876 ± 1 kDa (Fig. S2-1b, top, and Table S2-1). SL7 generated three broad mass peaks at 502 ± 26 kDa, 615 ± 22 kDa, and 713 ± 13 kDa (Fig. S2-1b, middle, and Table S2-1). SL8 generated two broad peaks at 737 ± 11 kDa and 840 ± 12 kDa (Fig. 2-

1b, bottom, and Table S2-1) and was chosen for further analysis due to less heterogeneity in the composition of the complex.

SL8-containing vRNPs were purified by GraFix (Fig. S2-1c). Analysis of peak fractions (7 + 8) by mass photometry (Fig. S2-1d) indicated that SL8 vRNPs were similar in mass to vRNPs assembled with 5'-600 RNA (Fig. 2-1e). Negative stain EM and two-dimensional class averages revealed ring structures that resemble vRNPs assembled with the 5'-600 RNA (Fig. 2-2c, compare to Fig. 2-1f).

These data suggest that ribonucleosome formation does not require 600 continuous bases of RNA but can be achieved with multiple copies of a relatively short and simple stem-loop structure. Unlike the 5'-600 RNA, the short stem-loop RNA is unlikely to serve as a platform to recruit multiple copies of N protein to assemble a vRNP. We speculate that the binding of a short RNA to N protein induces a conformational change that promotes protein-protein interactions required for vRNP formation. In the more physiologically relevant context of long RNAs, these weak protein-protein interactions are likely stabilized by multivalent interactions with an RNA molecule.

To test the requirement for secondary structure in vRNP formation, we generated a mutant SL8 (mSL8) carrying 12 mutations predicted to abolish the SL8 stem-loop structure (depicted in Fig. S2-2a, bottom). The mSL8 RNA promoted vRNP formation across a range of N protein concentrations (Fig. S2-2a). Analysis by GraFix and mass photometry confirmed that mSL8 forms a heterogeneous vRNP that is similar in size to

vRNPs containing SL8 (Fig. S2-2b, c). mSL8 vRNPs included some higher molecular mass species in the native gel and GraFix experiments. Thus, vRNPs can form in the presence of structured and largely unstructured RNA molecules, but vRNPs assembled with different RNAs are likely to exhibit small differences in overall structure or composition. We therefore speculate that the structural heterogeneity of the viral genome results in variations in the structure of the vRNPs of the viral nucleocapsid.

Analysis of non-crosslinked SL8-N protein complexes provided important clues about vRNP assembly. Mass photometry of the SL8-N sample revealed a major species at ~110 kDa, with five evenly-spaced complexes every 120-130 kDa thereafter up to ~755 kDa (108 ± 4 kDa, 225 ± 10 kDa, 360 ± 2 kDa, 468 ± 28 kDa, 600 ± 31 kDa, 736 ± 27 kDa) (Fig. 2-2d, middle, and Table S2-1). N protein alone exists primarily as a ~96 kDa dimer (97 ± 1 kDa) at the low concentration used for mass photometry (Fig. 2-2d, top, and Table S2-1; predicted mass 91.2 kDa), so the ~110 kDa peak likely represents one N protein dimer bound to one SL8 RNA (predicted mass of RNA: 23.1 kDa; predicted mass of complex: 114.4 kDa). The stepwise ~120-130 kDa increases in molecular mass are consistent with the addition of an N dimer bound to either one or two SL8 RNA molecules (predicted mass: 114.4 kDa or 137.5 kDa, respectively). These results support a potential assembly mechanism in which N protein dimers, bound to one or two stem-loops, iteratively assemble to form a full ribonucleosome containing twelve N proteins and six to twelve stem-loop RNAs (Fig. 2-2e). These data support the possibility that the vRNP assembled with 5’-600 RNA (Fig. 2-1e) contains 12 N proteins bound to one RNA.

In some crosslinked vRNP preparations, we observed an additional large peak in mass photometry that is likely to contain more than 12 N proteins. As mentioned above, crosslinked SL8 vRNPs contain a broad peak of 840 ± 12 kDa in addition to the 737 ± 11 kDa peak (Fig. S2-1b, bottom; also shown in Fig. 2-2d, bottom). Based on the similar molecular mass of the smaller peak in the crosslinked sample (737 ± 11 kDa) to the non-crosslinked sample (736 ± 27 kDa), we suspect that the larger crosslinked complex of 840 ± 12 kDa contains 14 N proteins. These results suggest that the ribonucleosome defaults to a stable complex of 12 N proteins bound to a variable number of RNA stem-loops but can adapt to accommodate fewer or more N protein dimers bound to additional RNA.

Multiple N protein regions promote formation of the ribonucleosome

Next, we sought to explore the regions of the N protein required for vRNP formation. We analyzed mutant proteins lacking the following regions: (1) the 44-aa N-terminal extension (NTE), a poorly-conserved disordered sequence[25,113]; (2) the highly conserved 31-aa serine/arginine (SR) region that has been implicated in RNA binding, oligomerization, and phosphorylation[15,22,39,41,44,48,60,107]; (3) the 20-aa leucine helix (LH), an alpha helix downstream of the SR region that interacts with Nsp3[109]; (4) the 33-aa CTD basic patch (CBP), a highly basic region containing a short disordered segment followed by an RNA-binding groove on the CTD; these structures have been implicated in Nsp3 binding[109] and helical stacking of N protein[30,31]; and (5) the 55-aa C-terminal

extension (CTE), which has been implicated in tetramerization and oligomerization of $N^{39\text{-}41,105}$ (Fig. 2-3a and Fig. S2-3a).

Mutant N proteins were mixed with 5'-600 RNA and analyzed by native gel electrophoresis (Fig. 2-3b). All mutant N proteins, with the exception of the CTE deletion, appeared to form fully assembled vRNPs. Most mutants contained a small population of lower bands beneath the fully shifted vRNP. These lower bands might represent sub-complexes in which the 5'-600 RNA is bound to fewer N proteins, presumably due to defects in vRNP assembly or stability. Deletion of the CTE resulted in a small shift that was considerably lower than the fully shifted vRNP. These results suggest that the ΔCTE N protein binds RNA but fails to form the fully assembled vRNP, hinting at an important role for the CTE in vRNP formation.

Studies of deletion mutants in complex with SL8 RNA, which minimizes the contribution of multivalent RNA binding, allowed us to investigate the critical protein-protein interactions that contribute to ribonucleosome formation. Mutant N proteins were mixed with SL8 RNA, crosslinked, and analyzed by native gel electrophoresis and mass photometry (Fig. 2-3c, d, and Table S2-1). Deletion of the NTE had little effect, other than to decrease the size of the vRNP complexes by ~30-40 kDa, suggesting that the NTE is not required for vRNP formation. All other deletion mutants had major defects in vRNP assembly.

Deletion of the CTE and LH resulted in almost complete disappearance of the vRNP when analyzed by native gel electrophoresis (Fig. 2-3c). These mutants appeared cloudy, and turbidity analysis revealed a higher absorbance at 340 nm compared to wild-type, suggesting formation of biomolecular condensates (Fig. S2-3b)[113]. Mass photometry analysis of the LH deletion showed a dominant peak at ~110 kDa, with two minor peaks at ~230 kDa and ~345 kDa (Fig. 2-3d and Table S2-1). The smallest peak represents an N dimer bound to one SL8 RNA, with the next two representing stepwise additions of one or two N dimers bound to an RNA. Thus, protein-protein interactions mediated by the LH are required for vRNP formation. Deletion of the CTE resulted in no discernable peaks above background on the mass photometer (Fig. 2-3d), further confirming the essential role of the CTE in vRNP formation and suggesting that multimerization driven by the CTE is required for ribonucleosome formation or stability.

Deletion of the SR and CBP regions also resulted in defects in vRNP assembly; both mutants exhibited a laddering of ribonucleoprotein subcomplexes when analyzed by native gel electrophoresis, as well as stepwise 120-130 kDa increases in molecular mass revealed by mass photometry (Fig. 2-3c, d, and Table S2-1). These data suggest the SR and CBP regions are required for complete assembly of the ribonucleosome.

LH and CBP deletions resulted in a minimal ribonucleoprotein complex of ~110 kDa, consistent with one N protein dimer bound to one SL8 RNA molecule. Interestingly, the SR deletion resulted in a minimal ribonucleoprotein complex of ~230 kDa, consistent with one N protein tetramer bound to two SL8 RNAs.

Native gel analysis revealed an increase in free SL8 RNA in mutant N protein samples compared to wild-type (Fig. 2-3c), suggesting reduced RNA binding in these mutants. Many of the deleted regions (SR, CBP, CTE) have been implicated in protein-protein and protein-RNA interactions. RNA binding defects might result simply from partial loss of RNA-binding sites, or they could occur because cooperative RNA binding is associated with vRNP assembly.

Phosphorylation inhibits formation of the ribonucleosome

The SR region of N protein is heavily phosphorylated in cells infected by SARS-CoV-2, and this modification promotes the protein's role in viral transcription[15,22,44,48,107,122]. In contrast, N protein in the virion is thought to be poorly phosphorylated[22,48]. We previously observed defects in vRNP formation when the 5'-400 RNA was mixed with a phosphomimetic N protein (the 10D mutant, in which 10 serines and threonines in the SR region are replaced with aspartic acid)[113], and here we sought to further explore phosphoregulation of the ribonucleosome. We mixed 5'-600 RNA with the 10D mutant and analyzed vRNP formation by native gel electrophoresis (Fig. 2-4a, left). The mutant formed an appropriately sized vRNP, apart from minor subcomplexes formed below the fully assembled vRNP. GraFix purification of the 10D mutant in complex with 5'-600 RNA revealed a range of ribonucleoprotein complexes similar to that seen with wild-type N protein (Fig. 2-4b and Fig. 2-1d, bottom). Mass photometry of fractions 7 + 8 confirmed a similar mass of the 10D and wild-type vRNPs, apart from a minor ~830 kDa peak observed with the 10D mutant (Fig. S2-4a and Fig. 2-1e, bottom).

Negative stain EM and two-dimensional class average analysis of the GraFix-purified 10D ribonucleoprotein complex, however, revealed a markedly different structure compared to the wild-type vRNPs (Fig. 2-4c and Fig. 2-1f). The 10D complex appears extended and heterogeneous, unlike the compact structure of the wild-type vRNP, and does not average into discrete, recognizable two-dimensional classifications. We therefore speculate that the 600 nt RNA provides sufficient binding sites for twelve 10D N proteins, but the 10D mutant is unable to condense into the compact ring structure observed with the wild-type N protein.

vRNP formation with the SL8 RNA was severely reduced in the 10D mutant when analyzed by native gel electrophoresis and mass photometry (Fig. 2-4a, right, and Fig. 2-4d). Both assays revealed a laddering of vRNP complexes, consistent with an inability of the 10D mutant to form a stable, fully assembled vRNP. Purification of the 10D + SL8 complex by GraFix revealed a clear shift toward lower molecular mass species when compared to wild-type N (Fig. 2-4e compared to Fig. S2-1c). This result was confirmed by mass photometry analysis of fractions 19 + 20 (Fig. S2-4b). Interestingly, the minimal unit of vRNP complex assembly with the 10D mutant (like the SR deletion) is ~230 kDa, which is consistent with an N protein tetramer bound to two SL8 RNAs. Negative stain EM and two-dimensional class averages of the GraFix-purified complex (fractions 19 + 20) revealed a smaller overall structure with an electron density distribution clearly distinct from vRNP complexes formed by wild-type N (Fig. 2-4f compared to Fig. 2-2c).

We next tested vRNP assembly with N protein that had been phosphorylated in vitro. In recent work, Yaron et al.[107] elegantly demonstrated a multi-kinase cascade that results in maximally phosphorylated N protein: SRPK phosphorylates S188 and S206, which primes the protein for subsequent phosphorylation of 8 more sites within the SR by GSK3, which then primes a final 4 sites for phosphorylation by CK1 (Fig. 2-5a). Consistent with this model, we observed maximal phosphorylation of N in the presence of all three kinases (Fig. 2-5b). Phosphorylation was greatly reduced when both SRPK priming sites were mutated to alanine (S188A + S206A mutant) (Fig. 2-5b). We mixed kinase-treated wild-type or S188A + S206A N proteins with SL8 RNA and purified the resulting vRNP complexes by GraFix (Fig. 2-5c). Wild-type phosphorylated N protein migrated as a low molecular weight ribonucleoprotein complex across the gradient, similar to the 10D mutant. The poorly phosphorylated S188A + S206A mutant, however, formed a vRNP across the gradient, similar to wild type unphosphorylated N protein (Fig. 2-5c). Mass photometry of the GraFix-purified samples further substantiated the defect in wild-type phospho-N vRNP assembly (Fig. 2-5d, top), which is rescued by mutation of the two priming phosphorylation sites (the S188A + S206A mutant) (Fig. 2-5d, bottom).

Discussion

The ‘beads-on-a-string’ model for coronavirus genome packaging lacks mechanistic detail. Here, we demonstrate that the N protein of SARS-CoV-2 assembles with viral RNA in vitro to form ribonucleosomes. These structures, which have been observed

previously in intact SARS-CoV-2 virions by cryo-electron tomography[117,118], appear to contain twelve N proteins (6 dimers) and a variable number of RNA segments.

Short RNAs appear to induce conformational changes in N protein that promote protein-protein interactions necessary for ribonucleosome assembly. These interactions might involve SR binding to the CTD[41], LH binding to other regions of the N protein, helical stacking of the CBP[30], and tetramerization driven by the CTE[39,40,105,123]. All of these binding interfaces contribute to the stability of the vRNP, but the CTE seems particularly critical for ribonucleosome formation.

vRNPs formed with long viral RNA (600 nt) do not fall apart as readily when diluted for mass photometry and do not require crosslinking for visualization by native gel electrophoresis. These results suggest that vRNPs assembled with 600-nt RNAs are more stable than those formed with multiple copies of a single short RNA, potentially because a single long RNA provides binding sites for all twelve N proteins in the vRNP. This multivalent RNA scaffold stabilizes low-affinity protein-protein interactions within the vRNP and reflects the more physiologically relevant state of RNA compaction by N protein in the virion.

Phosphorylation of N protein in its disordered SR region results in a less compact vRNP structure, providing further mechanistic insight into the functions of N protein phosphorylation and dephosphorylation during coronavirus infection. Given the high level of N protein phosphorylation in the infected cell, mechanisms must exist to generate a poorly phosphorylated N protein population at sites of viral assembly.

Coronavirus genomic RNA is structurally heterogeneous[119,120], and it remains unclear how ribonucleosomes accommodate variable RNA sequence and structure to package RNA in the virion. We find that the vRNP assembles in the presence of 600-nt RNA fragments from multiple genomic regions, suggesting that no specific sequences or secondary structures are required for vRNP formation. Furthermore, the ability of short RNAs to trigger vRNP formation suggests that ribonucleosome formation does not require 600 continuous bases of RNA. Inside the virion, it is not known whether each ribonucleosome forms on a continuous stretch of RNA in a nucleosome-like fashion or instead acts as a hub that binds stem-loops distributed across the genome, creating a web of condensed, interlinked protein-RNA interactions with 'nodes' at the ~38 vRNPs.

Our studies of ribonucleosome assembly with a small stem-loop RNA demonstrate that the vRNP is compositionally adaptive – that is, it can contain a variable number of N protein dimers bound to a variable number of stem-loop RNAs and assembles by iterative additions of N protein dimers bound to stem-loop RNAs. Our data suggest that the most stable form of the vRNP is 12 N proteins in complex with ~600 nt of RNA, but we also observed complexes that contain fewer or more N protein dimers. Given the iterative assembly of the vRNP, the multitude of protein-protein and protein-RNA interactions, and the high concentrations of N protein and RNA in the nucleocapsid, it seems reasonable to expect that the vRNP can expand to expose binding sites that allow additional N protein dimers to insert themselves into, or dissociate from, the cylindrical vRNP complex.

Our results, together with data from other studies, provide insights into the general architecture of coronavirus RNA packaging. There are ~38 vRNPs per virus[117,118], with each vRNP likely containing ~12 N proteins in complex with ~600 bases of viral RNA. This suggests that within a virus, the vRNPs contain ~500 N proteins bound to ~23,000 nt of RNA. It has been estimated that there are ~1000 N proteins per virus[124], while the viral genome is 30,000 nt in length, suggesting that some N proteins and RNA in the virion are not incorporated into vRNPs. Cryo-electron tomography studies indicate that most vRNPs are associated with the inner face of the membrane envelope, with a structure-free center in every virus[114,117,125]. Based on previous studies from our lab and others, this central region in the virion might contain a gel-like condensate of N protein bound heterogeneously to viral RNA[110,112,113].

During viral assembly, one copy of the ~30 kb viral genome is packaged in each virus, while cellular and subgenomic viral RNA are excluded[59]. In murine hepatitis virus (MHV), a 94 nt stem-loop in genomic RNA is necessary for exclusion of subgenomic RNA from the virus, suggesting that an analogous sequence or structure exists in SARS-CoV-2[126]. Our results demonstrate that the vRNP of SARS-CoV-2 does not appear to possess strict sequence or structure specificities, suggesting that another mechanism ensures specific incorporation of genomic RNA into the mature virus. The Membrane (M) protein likely functions in this capacity.

M protein is a 25 kDa structural protein containing three transmembrane helices followed by a ~100 aa C-terminal domain that faces the interior of the virion and is thought to interact with the C-terminus of N protein[59,127-132]. This interaction is required for maintaining packaging specificity in MHV[133,134]. The soluble CTD of M protein triggers RNA-independent phase separation when mixed with N protein[110], suggesting that M protein binding promotes a conformational change in N protein that leads to multivalent protein-protein interactions. Additionally, vRNPs in coronavirus virions appear to interact directly with the inner face of the virus membrane, with the circular 'base' of each vRNP cylinder proximal to the membrane[117,118,125]. With these lines of evidence in mind, it seems likely that M protein binds ribonucleosomes through the CTD and CTE of N, and tethers them to the viral membrane. Studies in MHV-infected cells have shown that N protein interacts with all coronavirus subgenomic RNAs, while M protein interacts only with full length genomic RNA[130]. The interaction of M with the vRNP might therefore promote binding of specific sequences or structures in the SARS-CoV-2 genomic RNA that allows for exclusive packaging of the coronavirus genome. Further biochemical and genetic studies will be necessary to clarify the precise role of each protein in this process and to see if any specific sequences promote packaging specificity in SARS-CoV-2.

N protein is highly phosphorylated in the cytoplasm of infected cells, and numerous kinases have been implicated in this process[15,22,48,107,113,122]. Yaron et al.[107] recently provided evidence for sequential phosphorylation of N by SRPK, GSK3 and CK1 (Fig. 2-5a). Our results are consistent with their model. We also show that phosphorylated N

protein cannot form compact ribonucleosomes and instead forms elongated, heterogeneous vRNP structures when mixed with longer viral RNA. Perhaps these heterogeneous vRNPs help maintain RNA in an uncompacted state that facilitates RNA processing at the RTC.

Chemical inhibition of SRPK with the FDA-approved drug Alectinib severely reduces replication of SARS-CoV-2 in multiple cell types[107]. Additionally, inhibition of GSK3 with lithium reduces coronavirus replication in cultured cells, and analysis of clinical data of patients taking lithium revealed a ~50% reduction in COVID-19 infection compared to those not on lithium[135]. Thus, inhibition of N protein phosphorylation represents a promising target for therapeutic intervention that has the potential to reduce mortality in individuals infected with SARS-CoV-2.

Methods

N protein preparation

Wild-type and mutant N proteins were produced as described previously[113]. Briefly, a codon-optimized synthetic DNA (Integrated DNA Technologies, IDT) was inserted into a pET28 expression vector by Gibson assembly, fused to DNA encoding an N-terminal 6xHis-SUMO tag. Mutant N proteins were generated by site-directed mutagenesis. N proteins were expressed in *E. coli* BL21 Star (Thermo #C601003), grown in TB-Kanamycin to OD 0.6, and induced with 0.4 mM IPTG. Cells were harvested, washed with PBS, snap frozen in LN_2 and stored at -80ºC until use. Thawed cells were resuspended in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 6 M urea)

and lysed by sonication. The lysate was clarified by centrifugation and bound to Ni-NTA agarose beads (QIAGEN #30230) for 45 min at 4ºC. Ni-NTA beads were washed 3 times with 10 bed volumes of buffer A and eluted with buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 250 mM imidazole, 6M urea). The eluate was concentrated in centrifugal concentrators (Millipore Sigma #UFC803024), transferred to dialysis tubing (Spectrum Labs #132676), and renatured overnight by dialysis in buffer C (50 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol). Recombinant Ulp1 catalytic domain (purified separately from *E. coli*) was added to renatured protein to cleave the 6xHis-SUMO tag, and cleaved protein was injected onto a Superdex 200 10/300 size exclusion column equilibrated in Buffer C. Peak fractions were pooled, concentrated, frozen in LN_2, and stored at -80ºC.

RNA preparation

Sequences of all RNAs used in this study are provided in Table S2-2. The template for in vitro transcription of 5'-600 RNA was a synthetic DNA (IDT), inserted by Gibson assembly into a pUC18 vector with a 5' T7 promoter sequence. The 5'-600 insert, including the 5' T7 sequence, was excised by EcoR1 digestion and purified by size exclusion chromatography on a Sephacryl 1000 column equilibrated in TE buffer (10 mM Tris pH 8, 1 mM EDTA). Peak fractions of the purified DNA insert were pooled and stored at -4°C.

Templates for all other long RNAs (5'-400, 5'-800, Nsp3, Nsp8/9, and Nsp10) were amplified by PCR of a plasmid containing the SARS-CoV-2 genome (a gift from Hiten

Madhani, UCSF). All forward primers included a 5' T7 promoter sequence. The SL8 and mSL8 templates were generated by PCR of synthetic DNA (IDT). The sequence for mutant SL8 (mSL8) was designed manually and checked for predicted secondary structure by RNAfold (http://rna.tbi.univie.ac.at/). PCR-amplified DNA was purified and concentrated by spin column (Zymo Research #D4004) before being used to generate RNA.

RNA synthesis was performed using the HiScribe T7 High Yield RNA synthesis kit (NEB #E2040S) according to the manufacturer's protocol. Following incubation at 37°C for 3 h, in vitro synthesized RNA was purified and concentrated by spin column (Zymo Research #R1018). To promote formation of proper RNA secondary structure, all purified RNAs were heat denatured at 95°C for 2 min in a pre-heated metal heat block, and then removed from heat and allowed to cool slowly to room temperature over the course of ~1 h. RNA concentration (A_{260}) was quantified by nanodrop.

Preparation of ribonucleoprotein complexes

The day before each experiment, N protein was dialyzed into reaction buffer (25 mM HEPES pH 7.5, 70 mM KCl) overnight. RNA was transcribed in vitro the day of analysis, heat-denatured and cooled slowly to allow for proper secondary structure. To assemble vRNP complexes, RNA was mixed with N protein (256 ng/µl RNA and 15 µM N, unless otherwise indicated) in a total volume of 10 µl and incubated for 10 min at 25°C. Samples containing stem-loop RNAs (SL4a, SL7, SL8, SL8m) were crosslinked by addition of 0.1% glutaraldehyde for 10 min at 25°C and then quenched with 100 mM

Tris pH 7.5. Samples containing longer RNAs (5'-400, 5'-600, 5'-800, Nsp3, Nsp8/9, Nsp10) were not crosslinked. After assembly, vRNP complexes were analyzed as described below.

RNA gel electrophoresis

After assembly (and crosslinking in the case of stem-loop RNAs), 10 µl vRNP mixtures were diluted 1:10 in dilution buffer (25 mM HEPES pH 7.5, 70 mM KCl, 10% glycerol). 2 µl of diluted vRNP mixtures was loaded onto a 5% polyacrylamide native TBE gel (Bio-Rad) and run at 125 V for 80 min at 4°C. 1 µl of the diluted samples was then denatured by addition of 4 M urea and Proteinase K (40 U/ml; New England Biolabs #P8107S), incubated for 5 min at 65°C, loaded onto a 6% polyacrylamide TBE-Urea Gel (Thermo Fisher), and run at 160 V for 50 min at room temperature. Gels were stained with SYBR Gold (Invitrogen) and imaged on a Typhoon FLA9500 Multimode imager set to detect Cy3.

Mass photometry

Mass photometry experiments were performed using a OneMP instrument (Refeyn). A silicone gasket well sheet (Grace Bio-Labs) was placed on top of a microscope coverslip and positioned on the microscope stage. 10 µl reaction buffer (25 mM HEPES pH 7.5, 70 mM KCl) was first loaded into the well to focus the objective, after which 1 µl of vRNP complex sample was added to the reaction buffer, mixed, and measured immediately. Samples containing stem-loop RNA were diluted 1:10 before a second

1:10 dilution directly on the coverslip, while samples containing longer RNAs were only diluted 1:10 on the coverslip.

The mass photometer was calibrated with NativeMark™ Unstained Protein Standard (Thermo #LC0725). Mass photometry data were acquired with AcquireMP and analyzed with DiscoverMP software (Refeyn). Mass photometry data are shown as histograms of individual mass measurements. Peaks were fitted with Gaussian curves to determine the average molecular mass of the selected distributions. Each condition was independently measured at least twice.

Turbidity analysis

Freshly prepared and renatured RNA was mixed with dialyzed N protein and incubated for 2 min at room temperature. Absorbance was measured at 260 nm and 340 nm using the Nanodrop Micro-UV/Vis Spectrophotometer. Turbidity was calculated by normalization of the 340 nm measurements to the absorbance value at 260 nm.

Negative stain

For negative-stain EM, 2.5 µl of vRNP samples were applied to a glow discharged Cu grid covered by continuous carbon film and stained with 0.75% (w/v) uranyl formate. A Tecnai T12 microscope (ThermoFisher FEI Company) operated at 120 kV was employed to analyze these negatively stained grids. Micrographs were recorded at a nominal magnification of 52,000X using a Gatan Rio 16 camera, corresponding to a pixel size of 1.34 Å on the specimen. All images were processed using cryoSPARC.

Micrographs were processed with Patch-Based CTF Estimation, and particles were picked using the blob picker followed by the template picker. Iterations of 2D classification generated final 2D averages.

Glycerol gradient centrifugation

Glycerol gradients were assembled as previously described, with slight modifications[136]. Briefly, 10-40% glycerol gradients (dialysis buffer containing 10% or 40% glycerol) were poured and mixed with the Gradient Master (BioComp). For GraFix purification, fresh 0.1% glutaraldehyde was added to the 40% glycerol buffer prior to gradient assembly. vRNP samples (generally 75 µl of 15 µM N with 256 ng/µl RNA) were gently added on top of the assembled 5 ml gradients and samples were centrifuged in a prechilled Ti55 rotor at 35,000 rpm for 17 h. Gradient fractions were collected by puncturing the bottom of the tube with a butterfly needle and collecting two drops per well. For analysis by negative stain electron microscopy and mass photometry, peak fractions were combined and buffer exchanged using centrifugal concentrators (Millipore Sigma #UFC510024). Concentrated samples were then re-diluted 1:10 with dialysis buffer (0% glycerol) and re-concentrated. Samples were diluted and re-concentrated three times.

Kinase reactions

Kinases were purchased from Promega (SRPK1: #VA7558, GSK-3β: #V1991, CK1ε: V4160). 1.25 µM N protein was incubated with 80 nM kinase for 30 min at 30^{o}C in kinase reaction buffer (25 mM HEPES pH 7.5, 35 mM KCl, 10 mM $MgCl_2$, 1 mM DTT,

0.5 mM ATP and 0.001 mCi/mL ^{32}P-γ-ATP). Reactions were quenched upon addition of SDS loading buffer for analysis by SDS-PAGE and autoradiography.

Phosphorylated protein for vRNP analysis was prepared in 90 µl reactions containing 16.5 µM N (WT or S188A + S206A) and 80 nM SRPK, GSK3, and CK1 in kinase reaction buffer. Reactions were incubated 30 min at 30ºC before addition of 5 mM EDTA. RNA was added to a final concentration of 256 ng/µl (which diluted N protein to a final concentration of 15 µM) and incubated at room temperature for 15 min. vRNP samples were analyzed by gradient centrifugation with crosslinker (GraFix) as described above.

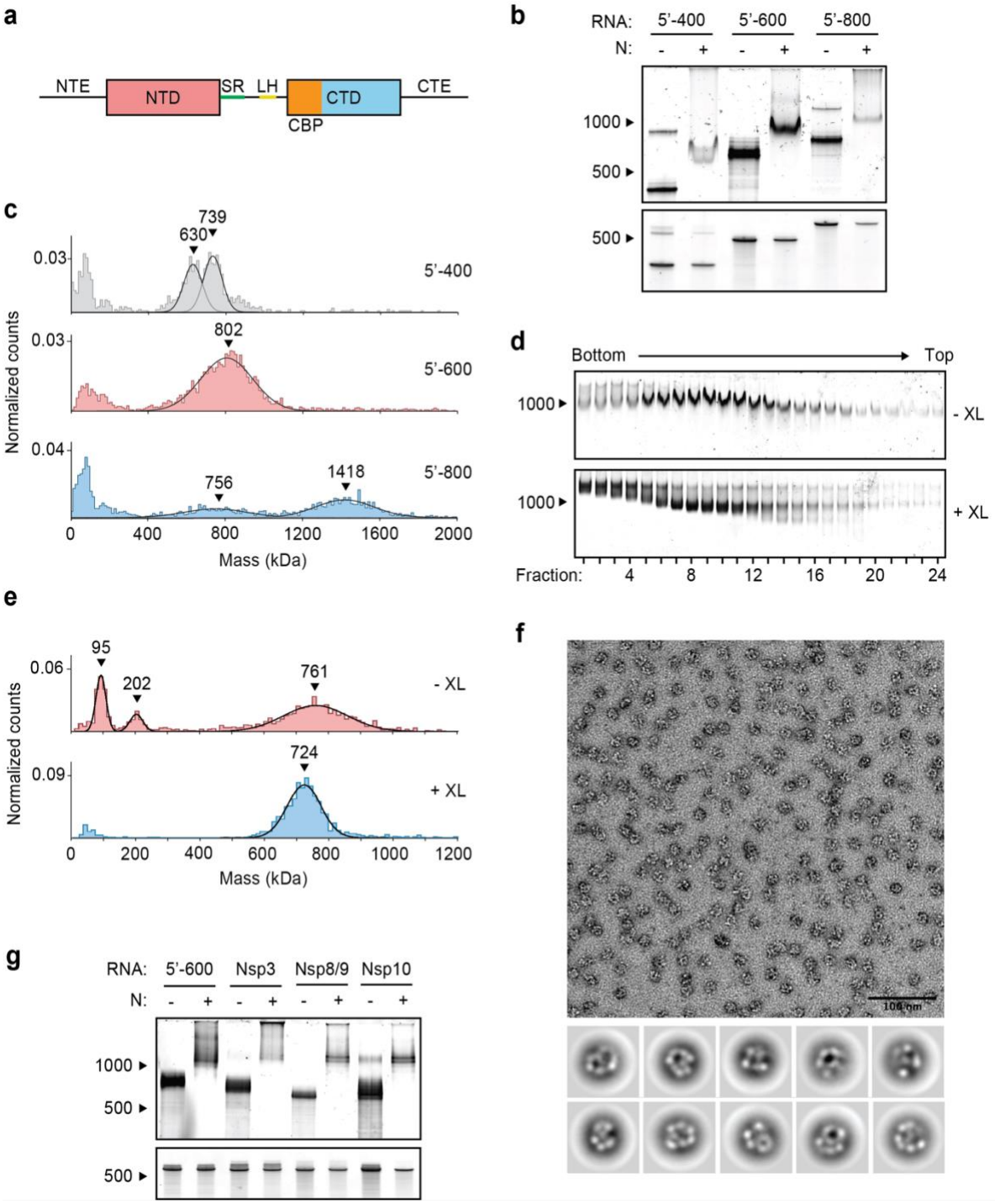

Fig 2-1. Viral RNA promotes formation of the SARS-CoV-2 ribonucleosome.

a Schematic of N protein domain architecture, including the N-terminal extension (NTE), N-terminal domain (NTD), Serine/Arginine region (SR), Leucine Helix (LH), C-terminal basic patch (CBP), C-terminal domain (CTD), and C-terminal extension (CTE). **b** Native (top) and denaturing (bottom) PAGE analysis of 15 µM N protein mixed with 256 ng/µl of the indicated RNA, stained with SYBR Gold to detect RNA species. RNA length

standards shown on left (nt). **c** Mass photometry analysis of vRNP complexes formed in the presence of 15 µM N and 256 ng/µl RNA. Data were fit to Gaussian distributions, with mean molecular mass indicated above each peak. Representative of two independent experiments (Table S2-1). **d** Native gel analysis of glycerol gradient separated vRNP complexes. Top: no crosslinker added (-XL); bottom: 0.1% glutaraldehyde added (+XL) to 40% glycerol buffer (GraFix). **e** Fractions 7 and 8 (from **d**) were combined and analyzed by mass photometry, as in **c**. Top: no crosslinker (-XL); bottom: GraFix-purified vRNP (+XL). Representative of two independent experiments (Table S2-1). **f** Negative stain electron microscopy and two-dimensional classification of GraFix-purified vRNPs (combined fractions 7 and 8 from **d**). **g** Native (top) and denaturing gel analysis (bottom) of 15 µM N protein mixed with 256 ng/µl of the indicated 600 nt RNA molecules. See Table S2-2 for sequences.

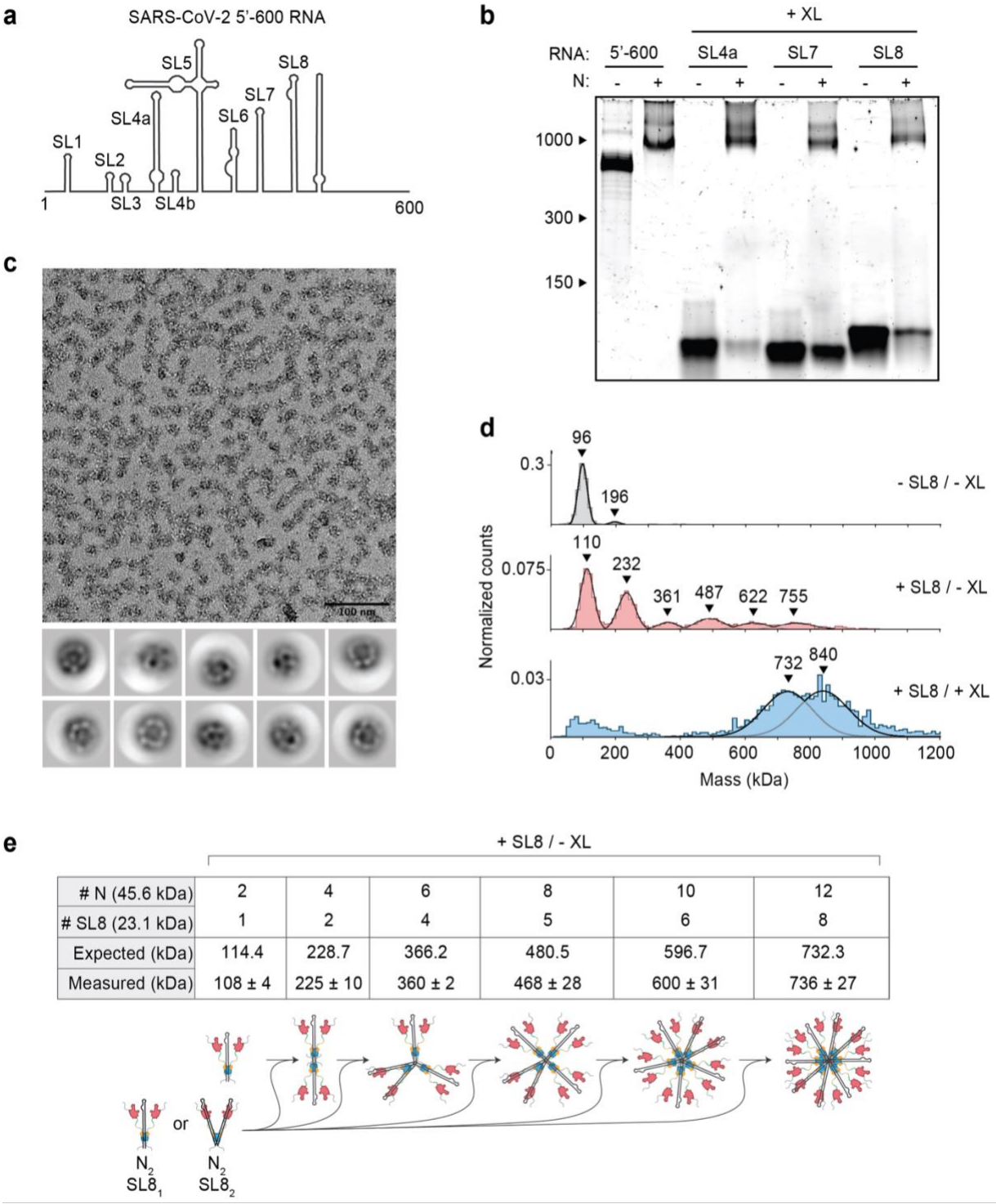

Fig 2-2. **Stem-loop RNA, in complex with N protein, drives ribonucleosome formation.**

a Schematic of RNA secondary structure in the 5'-600 RNA[119]. **b** Native gel analysis of 15 µM N protein mixed with 256 ng/µl of the indicated RNAs. Samples containing stem-loop RNA were crosslinked to stabilize the resulting complex, while the 5'-600 RNA

sample was left un-crosslinked. Corresponding denaturing gel analysis shown in Fig. S2-1a. **c** Fractions 7 and 8 of GraFix-purified SL8 assembled vRNPs were combined and analyzed by negative stain electron microscopy and two-dimensional classification. **d** Mass photometry analysis of indicated N protein-RNA mixtures. Top: N protein alone; middle: N protein mixed with SL8, un-crosslinked; bottom: crosslinked complexes of N protein bound to SL8 (data reproduced from Fig. S2-1b for ease of comparison). Representative of two independent experiments (Table S2-1). **e** Predictions of N protein and RNA stoichiometry, based on measured masses of N protein in complex with SL8 RNA without crosslinker (**d**, middle panel). Measured masses are means ± standard deviation in two independent experiments (Table S2-1). Below the table is a schematic of a proposed assembly mechanism in which N protein dimers, bound to one or two stem-loop RNAs, iteratively assemble to the full vRNP.

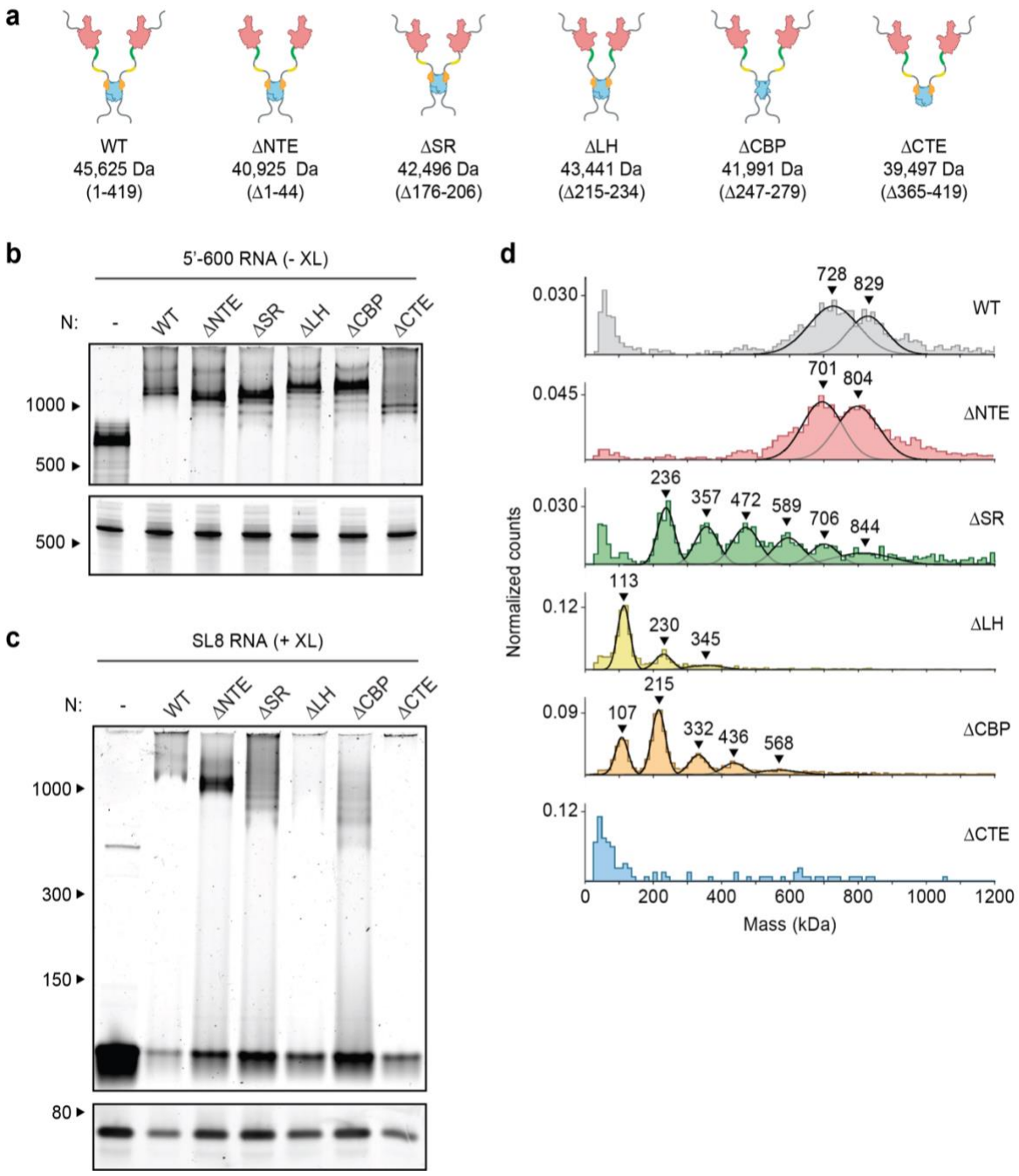

Fig 2-3. Disordered regions contribute to vRNP formation.

a Schematic of wild-type (WT) N protein and deletion mutants, as described in the text. Mass is that of monomeric N protein. **b** 15 µM N protein mutants were mixed with 256 ng/µl 5'-600 RNA and analyzed by native (top) and denaturing (bottom) gel electrophoresis. **c** 20 µM N protein mutants were mixed with 256 ng/µl SL8 RNA and analyzed by native (top) and denaturing (bottom) gel electrophoresis. SL8 ribonucleoprotein complexes were crosslinked prior to analysis. **d** Mass photometry analysis of crosslinked N protein mutants (20 µM) bound to SL8 RNA (256 ng/µl). Representative of at least two independent experiments (Table S2-1).

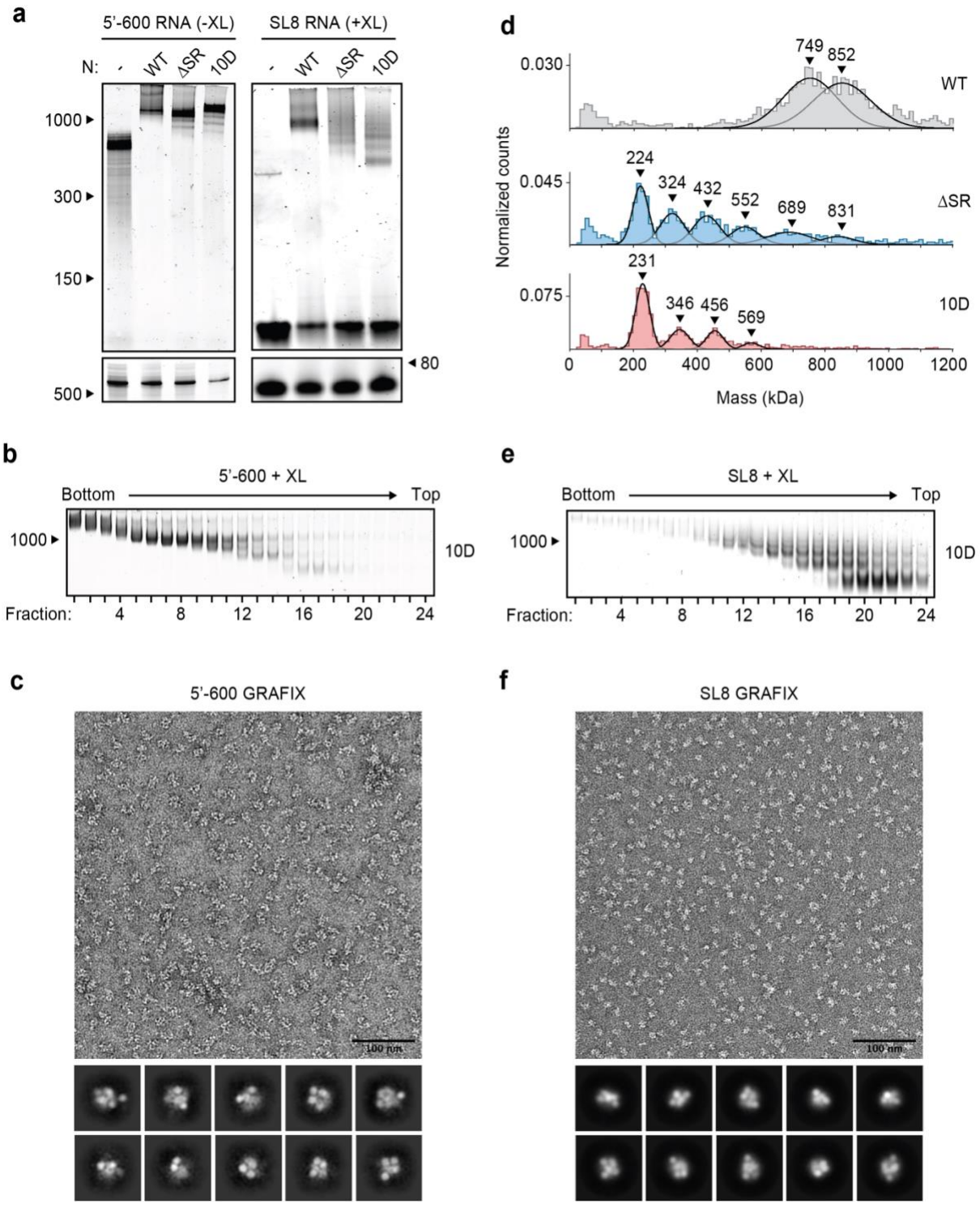

Fig 2-4. Phosphomimetic mutations in the SR region of N prevent vRNP assembly.

a 15 µM N protein constructs were combined with 256 ng/µl 5'-600 RNA (left) or 256 ng/µl SL8 RNA (right) and analyzed by native (top) and denaturing (bottom) gel electrophoresis. SL8 ribonucleoprotein complexes were crosslinked prior to native gel electrophoresis. WT, wild type. See Fig. 2-5a for the ten sites of phosphorylation mutated to aspartic acid in the 10D mutant. **b** 15 µM phosphomimetic N protein (10D) was mixed with 256 ng/µl 5'-600 RNA and separated by glycerol gradient centrifugation in the presence of crosslinker (GraFix). Fractions were collected and analyzed by native gel electrophoresis. **c** Fractions 7 and 8 of GraFix-separated vRNPs (from **b**) were combined and analyzed by negative stain electron microscopy and two-dimensional classification. **d** 15 µM N protein mutants were mixed with 256 ng/µl SL8 RNA, crosslinked, and analyzed by mass photometry. Representative of at least two independent experiments (Table S2-1). A separate analysis of ΔSR mutant is also shown in Fig. 2-3d. **e** 15 µM 10D N protein was mixed with 256 ng/µl SL8 RNA and separated by GraFix. Fractions were analyzed by native gel electrophoresis. **f** Fractions 19 and 20 of GraFix-purified 10D N in complex with SL8 RNA (from **e**) were combined and visualized by negative stain electron microscopy and two-dimensional classification.

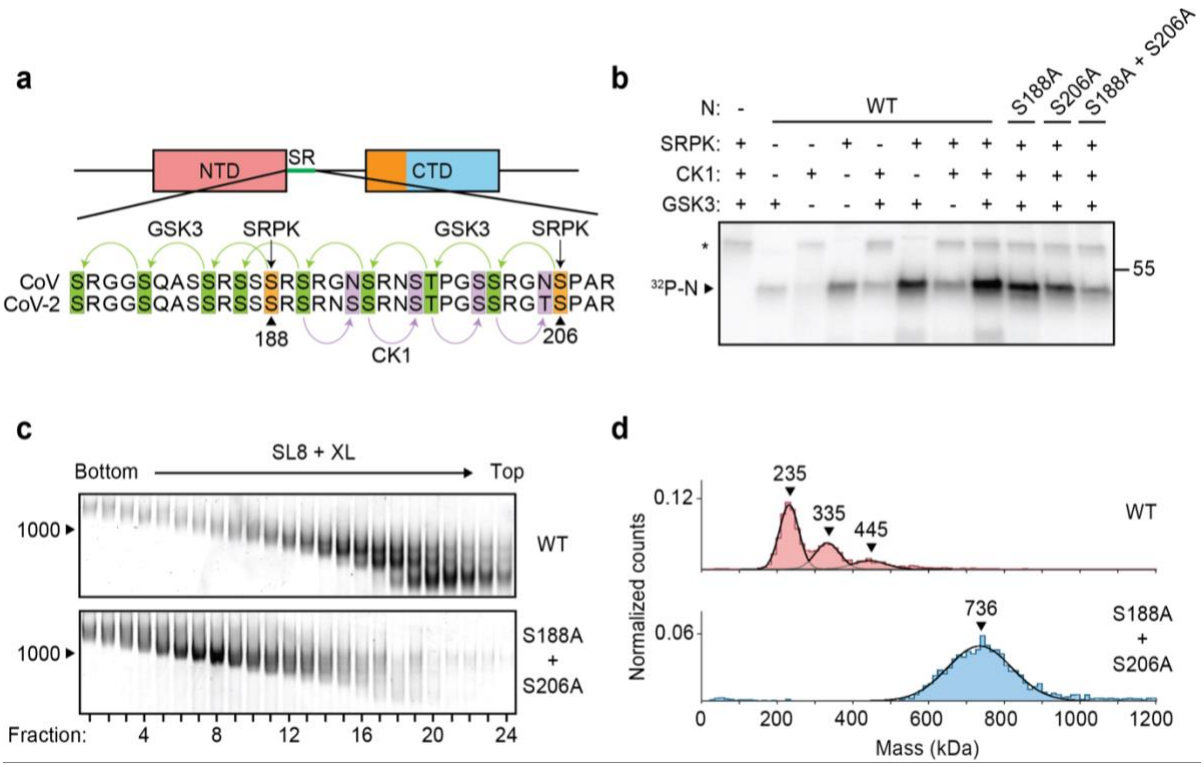

Fig 2-5. Phosphorylation of N protein inhibits ribonucleosome formation.

a Sequence of N protein SR regions from SARS-CoV (aa 177-210) and SARS-CoV-2 (aa 176-209). The proposed mechanism of sequential phosphorylation[107] is initiated by SRPK at S188 and S206 (orange), which leads to downstream phosphorylation of eight sites by GSK3 (green), allowing for final phosphorylation of four additional sites by CK1 (purple). In the phosphomimetic 10D mutant used in Fig. 2-4, the SRPK and GSK3 sites are changed to aspartic acid. **b** Wild-type (WT) and mutant N protein constructs were incubated with the indicated kinases in the presence of radiolabeled ATP and analyzed by SDS-PAGE and autoradiography. Phosphorylated N is indicated. Asterisk denotes autophosphorylation of CK1. Molecular mass marker shown on right (kDa). **c** N protein (WT or S188A + S206A) was phosphorylated by SRPK, GSK3, and CK1, and then mixed with SL8 RNA. The resulting ribonucleoprotein complexes were separated by glycerol gradient centrifugation in the presence of crosslinker (GraFix) and analyzed by native gel electrophoresis. **d** Peak fractions from the GraFix analyses in **c** were analyzed by mass photometry. Top: fractions 19 + 20 of wild-type N; bottom: fractions 7 + 8 of S188A + S206A mutant N. Representative of two independent experiments (Table S2-1).

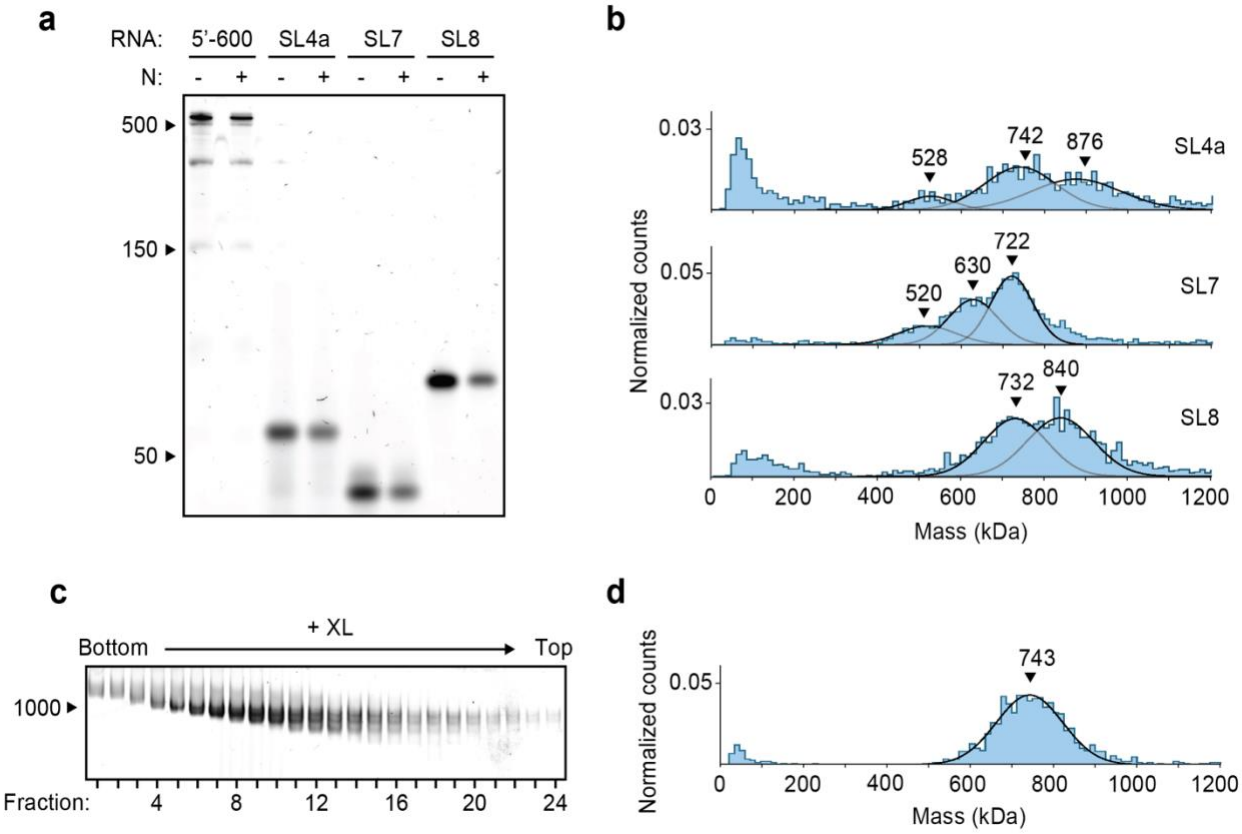

Fig S2-1. vRNP formation with stem-loop RNAs.

a, Denaturing gel electrophoresis of N protein mixed with indicated RNAs, related to Fig. 2-2b. **b,** Mass photometry analyses of crosslinked N protein complexes with indicated RNAs. Results with SL8 are reproduced in Fig. 2-2d. Representative of two independent experiments (Table S2-1). **c,** N protein in complex with SL8 RNA was separated by glycerol gradient centrifugation in the presence of crosslinker (GraFix) and analyzed by native gel electrophoresis. **d,** Fractions 7 and 8 of GraFix-purified N-SL8 vRNPs (from **c**) were combined and analyzed by mass photometry. Representative of two independent experiments (Table S2-1).

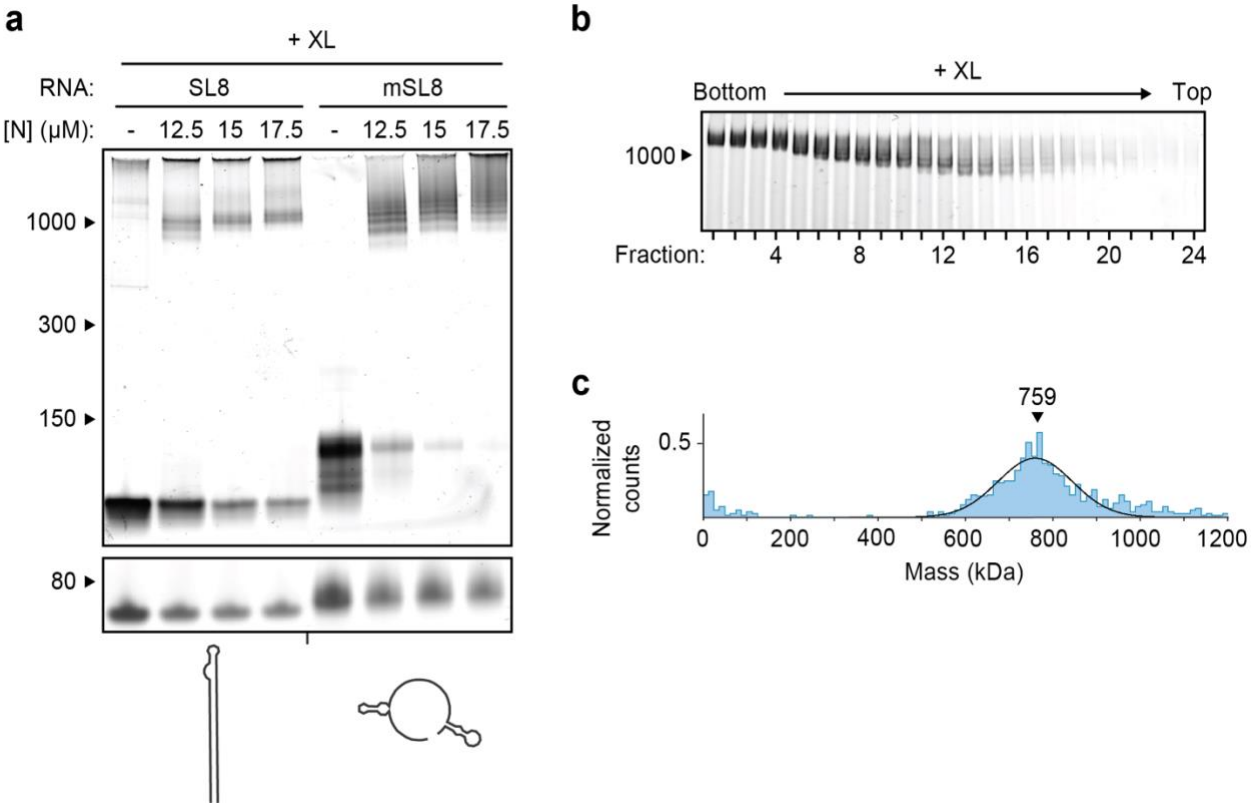

Fig S2-2. Analysis of vRNP formation with mSL8 RNA.

a, The indicated concentrations of N protein were combined with SL8 RNA or mutant SL8 RNA (mSL8), crosslinked, and analyzed by native (top) and denaturing (bottom) gel electrophoresis. Predicted secondary structures are shown below. See Table S2-2 for sequences. **b**, N protein in complex with mSL8 RNA was separated by glycerol gradient centrifugation in the presence of crosslinker (GraFix) and analyzed by native gel electrophoresis. **c**, Fractions 7 and 8 of GraFix-purified N-mSL8 vRNPs (from **b**) were combined and analyzed by mass photometry. Representative of two independent experiments (Table S2-1).

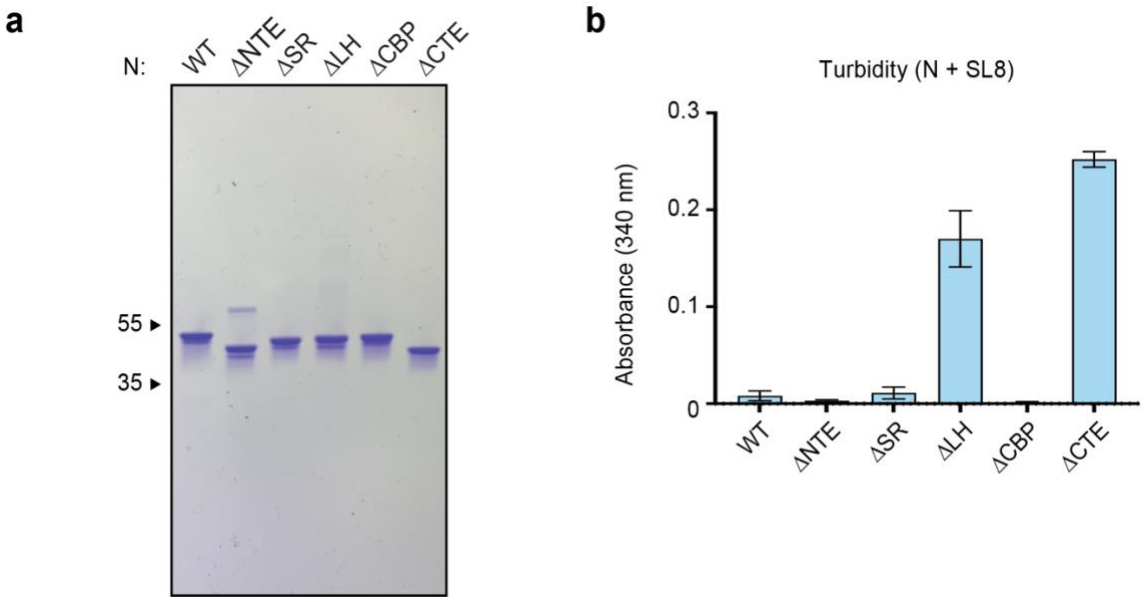

Fig S2-3. Analysis of N protein deletion mutants.

a, SDS-PAGE of N protein constructs used in this study, stained with Coomassie Blue. Molecular weight markers at left (kDa). **b,** Absorbance at 340 nm was used to quantify the turbidity of wild-type and mutant N proteins mixed with SL8 RNA. All values are normalized to absorbance at 260 nm.

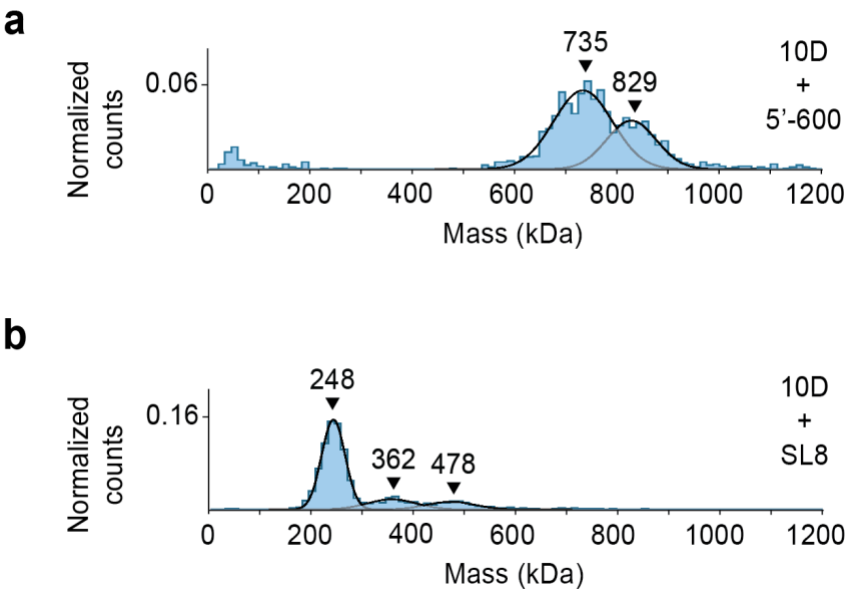

Fig S2-4. Analysis of complex formation by 10D mutant.

Mass photometry analysis of GraFix-purified (**a**) fractions 7 + 8 of N protein in complex with 5'-600 RNA and (**b**) fractions 19 + 20 of N protein in complex with SL8 RNA. Representative of two independent experiments (Table S2-1).

Table S2-1. Summary of mass photometry results (kDa).

	Figure 1								
	Figure 1C								
	5'-400			5'-600			5'-800		
peak	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD
1	630	631	630.5 ± 0.7	802	853	827.5 ± 36	756	721	738.5 ± 24.7
2	739	728	733.5 ± 7.8				1418	1613	1515.5 ± 137.9

	Figure 1E					
	-XL			+XL		
peak	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD
1	95	98	96.5 ± 2.2	724	714	719 ± 7.1
2	202	211	206.5 ± 6.4			
3	761	770	765.5 ± 6.4			

	Figure 2									
	Figure 2D									
	-SL8 / -XL			+SL8 / -XL			+SL8 / +XL (from Fig S1B)			
peak	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD	rep 1	rep 2	rep 3	mean ± SD
1	96	98	97 ± 1.4	110	105	107.5 ± 3.5	732	728	749	736.5 ± 11.2
2	196	202	199 ± 4.2	232	218	225 ± 9.9	840	829	852	840.3 ± 11.5
3				361	358	359.5 ± 2.1				
4				487	448	467.5 ± 27.6				
5				622	578	600 ± 31.1				
6				755	717	736 ± 26.9				

	Figure 3									
	Figure 3D									
	WT (from Fig S1B)				ΔNTE			ΔSR		
peak	rep 1	rep 2	rep 3	mean ± SD	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD
1	732	728	749	736.5 ± 11.2	701	702	701.5 ± 1	236	224	230 ± 8.5
2	840	829	852	840.3 ± 11.5	775	804	789.5 ± 20.5	357	324	340.5 ± 23.3
3								472	432	452 ± 28.3
4								589	552	570.5 ± 26.2
5								706	689	697.5 ± 12
6								844	831	837.5 ± 9.2

	Figure 3D (continued)								
	ΔLH			ΔCBP			ΔCTE		
peak	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD
1	100	113	106.5 ± 9.2	92	107	99.5 ± 10.6	n/a	n/a	n/a
2	214	230	222 ± 11.3	200	215	207.5 ± 10.6			
3	335	345	340 ± 7.1	303	332	317.5 ± 20.5			
4				428	436	432 ± 5.7			
5				562	568	565 ± 4.2			
6									

	Figure 4									
	Figure 4D									
	WT (from Fig S1B)				ΔSR (from fig 3)			10D		
peak	rep 1	rep 2	rep 3	mean ± SD	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD
1	732	728	749	736.5 ± 11.2	236	224	230 ± 8.5	231	230	230.5 ± 0.7
2	840	829	852	840.3 ± 11.5	357	324	340.5 ± 23.3	346	348	347 ± 1.4
3					472	432	452 ± 28.3	456	452	454 ± 2.8
4					589	552	570.5 ± 26.2	569	540	554.5 ± 20.5
5					706	689	697.5 ± 12			
6					844	831	837.5 ± 9.2			

	Figure 5					
	Figure 5D					
	WT (+kinases)			S188A + S206A (+kinases)		
peak	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD
1	235	207	221 ± 19.8	736	730	733 ± 4.2
2	335	308	321.5 ± 19.1			
3	445	447	446 ± 1.4			

	Figure S1									
	Figure S1B									
	SL4a			SL7			SL8			
peak	rep 1	rep 2	mean ± SD	rep 1	rep 2	mean ± SD	rep 1	rep 2	rep 3	mean ± SD
1	528	501	514.5 ± 19.1	520	484	502 ± 25.5	732	728	749	736.5 ± 11.2
2	742	735	738.5 ± 4.9	630	599	614.5 ± 21.9	840	829	852	840.3 ± 11.5
3	876	875	875.5 ± 0.7	722	703	712.5 ± 13.4				

	Figure S1D		
	SL8 (GRAFIX)		
peak	rep 1	rep 2	mean ± SD
1	743	736	739.5 ± 4.9

	Figure S2		
	Figure S2C		
	WT +mSL8 (GRAFIX)		
peak	rep 1	rep 2	mean ± SD
1	759	773	766 ± 9.9

	Figure S3		
	Figure S3A		
	10D + 5'-600 (GRAFIX)		
peak	rep 1	rep 2	mean ± SD
1	735	725	730 ± 7.1
2	829	817	823 ± 8.5

	Figure S3B		
	10D + SL8 (GRAFIX)		
peak	rep 1	rep 2	mean ± SD
1	248	249	248.5 ± 0.7
2	362	367	364.5 ± 3.5
3	478	487	482.5 ± 6.4

Table S2-2. RNA sequences used in this study.

5’-400	auuaaagguuuauaccuucccagguaacaaaccaaccaacuuucgaucucuuguagaucuguucucu aaacgaacuuuaaaaucuguguggcugucacucggcugcaugcuuagugcacucacgcaguauaauu aauaacuaauuacugucguugacaggacacgaguaacucgucuaucuucugcaggcugcuuacgguu ucguccguguugcagccgaucaucagcacaucuagguuucguccgggugugaccgaaagguaagaug gagagccuugucccuggguucaacgagaaaacacacguccaacucaguuugccuguuuuacagguuc gcgacgugcucguacguggcuuuggagacuccguggaggaggucuuaucagaggcacgucaacau
5’-600	auuaaagguuuauaccuucccagguaacaaaccaaccaacuuucgaucucuuguagaucuguucucu aaacgaacuuuaaaaucuguguggcugucacucggcugcaugcuuagugcacucacgcaguauaauu aauaacuaauuacugucguugacaggacacgaguaacucgucuaucuucugcaggcugcuuacgguu ucguccguguugcagccgaucaucagcacaucuagguuucguccgggugugaccgaaagguaagaug gagagccuugucccuggguucaacgagaaaacacacguccaacucaguuugccuguuuuacagguuc gcgacgugcucguacguggcuuuggagacuccguggaggaggucuuaucagaggcacgucaacaucu uaaagauggcacuuguggcuuaguagaaguugaaaaaggcguuuugccucaacuugaacagcccuau guguucaucaaacguucggaugcucgaacugcaccucauggucauguuaugguugagcugguagcag aacucgaaggcauucaguacggucguaguggugagacacuuggugucccuugucccucauguggg
5’-800	auuaaagguuuauaccuucccagguaacaaaccaaccaacuuucgaucucuuguagaucuguucucu aaacgaacuuuaaaaucuguguggcugucacucggcugcaugcuuagugcacucacgcaguauaauu aauaacuaauuacugucguugacaggacacgaguaacucgucuaucuucugcaggcugcuuacgguu ucguccguguugcagccgaucaucagcacaucuagguuucguccgggugugaccgaaagguaagaug gagagccuugucccuggguucaacgagaaaacacacguccaacucaguuugccuguuuuacagguuc gcgacgugcucguacguggcuuuggagacuccguggaggaggucuuaucagaggcacgucaacaucu uaaagauggcacuuguggcuuaguagaaguugaaaaaggcguuuugccucaacuugaacagcccuau guguucaucaaacguucggaugcucgaacugcaccucauggucauguuaugguugagcugguagcag aacucgaaggcauucaguacggucguaguggugagacacuuggugucccuugucccucauguggggcga aauaccaguggcuuaccgcaagguucuucuucguaagaacgguaauaaaggagcugguggccauagu uacggcgccgaucuaaagucauuugacuuaggcgacgagcuuggcacugauccuuaugaagauuuuc aagaaaacuggaacacuaaacauagcaguggguuacccgugaacucaugcgugagcuuaacg
Nsp3	uuaugaaagacauucucucucucauuuuguuaacuuagacaaccugagagcuaauaacacuaaaggu ucauugccuauuaauguuauaguuuuugaugguaaaucaaaaugugaagaaucaucugcaaaauca gcgucuguuuacuacagucagcuuaugugucaaccuauacuguuacuagaucaggcauuagugucug auguuggugauagugcggaaguugcaguuaaaauguuugaugcuuacguuaauacguuuucaucaa cuuuuaacguaccaauggaaaaacucaaaacacuaguugcaacugcagaagcugaacuugcaaagaa ugugucccuuagacaaugucuuaucuacuuuuauuucagcagcucggcaaggguuuguugauucagau guagaaacuaaagauguuguugaaugucuuaaauugucacaucaaucugacauagaaguuacuggc gauaguuguaauaacuauaugcucaccuauaacaaaguugaaaacaugacaccccgugaccuuggug cuuguauugacuguagugcgcgucauauuaaugcgcagguagcaaaaagucacaacauugcuuugau
Nsp8/9	gccaugcaacguaaguuggaaaagauggcugaucaagcuaugacccaaauguauaaacaggcuagau cugaggacaagagggcaaaaguuacuagugcuaugcagacaaugcuuuucacuaugcuuagaaaguu ggauaaugaugcacucaacaacauuaucaacaaugcaagagaugguuguguucccuugaacauaaua ccucuuacaacagcagccaaacuaaugguugucauaccagacuauaacacauauaaaaauacgugug augguacaacauuuacuuaugcaucagcauugugggaaauccaacagguuguagaugcagauagua aaauuguucaacuuagugaaauuaguauggacaauucaccuaauuuagcauggccucuuauuguaac agcuuuaagggccaauucugcugucaaauuacagaauaaugagcuuaguccuguugcacuacgacag augucuugugcugccgguacuacacaaacugcuugcacugaugacaaugcguuagcuuacuacaaca caacaaagggaggguagguuuguacuugcacuguuauccgauuuacaggauuugaaaugggcuaga

Nsp10	ggaagccaauauggaucaagaauccuuuggugggugcaucguguugucuguacugccguugccacaua gaucauccaaauccuaaaggauuuugugacuuaaaagguaaguauguacaaauaccuacaacuugug cuaaugacccuguggguuuuacacuuaaaaacacagucuguaccgucugcgguauguggaaaggguua uggcuguaguugugaucaacuccgcgaacccaugcuucagucagcugaugcacaaucguuuuuaaac ggguuugcgguguaagugcagcccgucuuacaccgugcggcacaggcacuaguacugaugucguaua cagggcuuuugacaucuacaaugauaaaguagcugguuuugcuaaauuccuaaaaacuaauuguug ucgcuuccaagaaaaggacgaagaugacaauuuaauugauucuuacuuuguaguuaagagacacacu uucucuaacuaccaacaugaagaaacaauuuauaauuuacuuaaggauuguccagcuguugcuaaac augacuucuuuaaguuuagaauagacggugacaugguaccacauauaucacgucaacgucuuacu
SL4a	uuaaaaucuguguggcugucacucggcugcaugcuuagugcacucacgcaguauaa
SL7	acguggcuuuggagacuccguggaggaggucuuaucagaggcacgu
SL8	gauggcacuuguggcuuaguagaaguugaaaaaggcguuuugccucaacuugaacagcccuaugugu ucauc
mSL8	gauAgcaAuuguAAcuuaguagUCACGgaaaaagUcguuuugccucaacuugaacagcccuauA ugAucauc

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. in *The Coronaviridae* (ed S. G. Siddell) 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 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).

5 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).

6 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).

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 Snijder, E. J. *et al.* A unifying structural and functional model of the coronavirus replication organelle: Tracking down RNA synthesis. *PLoS Biol* **18**, e3000715, doi:10.1371/journal.pbio.3000715 (2020).

12 Stertz, S. *et al.* The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* **361**, 304-315, doi:10.1016/j.virol.2006.11.027 (2007).

13 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).

14 Wolff, G. *et al.* A molecular pore spans the double membrane of the coronavirus replication organelle. *Science*, doi:10.1126/science.abd3629 (2020).

15 Bouhaddou, M. *et al.* The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **182**, 685-712 e619, doi:10.1016/j.cell.2020.06.034 (2020).

16 Bost, A. G., Carnahan, R. H., Lu, X. T. & Denison, M. R. Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. *J Virol* **74**, 3379-3387, doi:10.1128/jvi.74.7.3379-3387.2000 (2000).

17 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).

18 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).

19 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).

20 Verheije, M. H. *et al.* Mouse hepatitis coronavirus RNA replication depends on GBF1-mediated ARF1 activation. *PLoS Pathog* **4**, e1000088, doi:10.1371/journal.ppat.1000088 (2008).

21 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).

22 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).

23 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).

24 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).

25 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).

26 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).

27 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).

28 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).

29 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).

30 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).

31 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).

32 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).

33 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).

34 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).

35 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).

36 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).

37 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).

38 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).

39 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).

40 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).

41 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).

42 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).

43 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).

44 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).

45 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).

46 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).

47 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).

48 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).

49 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).

50 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).

51 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).

52 Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, doi:10.1126/science.aaf4382 (2017).

53 Tauber, D., Tauber, G. & Parker, R. Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. *Trends Biochem Sci* **45**, 764-778, doi:10.1016/j.tibs.2020.05.002 (2020).

54 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).

55 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).

56 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).

57 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).

58 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).

59 Masters, P. S. Coronavirus genomic RNA packaging. *Virology* **537**, 198-207, doi:10.1016/j.virol.2019.08.031 (2019).

60 Grossoehme, 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).

61 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).

62 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).

63 Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* **13**, 261-291, doi:10.1146/annurev.cellbio.13.1.261 (1997).

64 Hurst, K. R., Koetzner, C. A. & Masters, P. S. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural

protein 3 of the viral replicase-transcriptase complex. *J Virol* **87**, 9159-9172, doi:10.1128/JVI.01275-13 (2013).

65 Hurst, K. R., Ye, R., Goebel, S. J., Jayaraman, P. & Masters, P. S. An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. *J Virol* **84**, 10276-10288, doi:10.1128/JVI.01287-10 (2010).

66 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).

67 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).

68 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).

69 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).

70 Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *BioRxiv*, doi:https://doi.org/10.1101/2020.06.23.167064 (2020).

71 Yao, H. *et al.* Molecular architecture of the SARS-CoV-2 virus. *BioRxiv*, doi:https://doi.org/10.1101/2020.07.08.192104 (2020).

72 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).

73 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).

74 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).

75 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).

76 Chen, H. *et al.* Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res*, doi:10.1038/s41422-020-00408-2 (2020).

77 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).

78 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).

79 Jack, A. *et al.* SARS CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *BioRxiv*, doi:https://doi.org/10.1101/2020.09.14.295824 (2020).

80 Lu, S. *et al.* The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *BioRxiv*, doi:https://doi.org/10.1101/2020.07.30.228023 (2020).

81 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).

82 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).

83 Cascarina, S. M. & Ross, E. D. A proposed role for the SARS-CoV-2 nucleocapsid protein in the formation and regulation of biomolecular condensates. *FASEB J*, doi:10.1096/fj.202001351 (2020).

84 McCormick, C. & Khaperskyy, D. A. Translation inhibition and stress granules in the antiviral immune response. *Nat Rev Immunol* **17**, 647-660, doi:10.1038/nri.2017.63 (2017).

85 Protter, D. S. W. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol* **26**, 668-679, doi:10.1016/j.tcb.2016.05.004 (2016).

86 Tsai, W. C. & Lloyd, R. E. Cytoplasmic RNA Granules and Viral Infection. *Annu Rev Virol* **1**, 147-170, doi:10.1146/annurev-virology-031413-085505 (2014).

87 Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459-468, doi:10.1038/s41586-020-2286-9 (2020).

88 Li, J. *et al.* Virus-host interactome and proteomic survey of PBMCs from COVID-19 patients reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *BioRxiv*, doi:https://doi.org/10.1101/2020.03.31.019216 (2020).

89 Samavarchi-Tehrani, P. *et al.* A SARS-CoV-2 - host proximity interactome. *BioRxiv*, doi:https://doi.org/10.1101/2020.09.03.282103 (2020).

90 Stukalov, A. *et al.* Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. *BioRxiv*, doi:https://doi.org/10.1101/2020.06.17.156455 (2020).

91 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).

92 Wang, Y. *et al.* Low stability of nucleocapsid protein in SARS virus. *Biochemistry* **43**, 11103-11108, doi:10.1021/bi049194b (2004).

93 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).

94 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).

95 Jain, A., Liu, R., Xiang, Y. K. & Ha, T. Single-molecule pull-down for studying protein interactions. *Nat Protoc* **7**, 445-452, doi:10.1038/nprot.2011.452 (2012).

96 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).

97 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).

98 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).

99 Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273, doi:10.1038/s41586-020-2012-7 (2020).

100 Scherer, K. M. *et al.* SARS-CoV-2 nucleocapsid protein adheres to replication organelles before viral assembly at the Golgi/ERGIC and lysosome-mediated egress. *Sci Adv* **8**, eabl4895, doi:10.1126/sciadv.abl4895 (2022).

101 V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19**, 155-170, doi:10.1038/s41579-020-00468-6 (2021).

102 Snijder, E. J. *et al.* Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. *J Virol* **80**, 5927-5940, doi:10.1128/JVI.02501-05 (2006).

103 Sawicki, S. G., Sawicki, D. L. & Siddell, S. G. A contemporary view of coronavirus transcription. *J Virol* **81**, 20-29, doi:10.1128/JVI.01358-06 (2007).

104 McBride, R., van Zyl, M. & Fielding, B. C. The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991-3018, doi:10.3390/v6082991 (2014).

105 Ye, Q., West, A. M. V., Silletti, S. & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci*, doi:10.1002/pro.3909 (2020).

106 Kang, S. *et al.* Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B* **10**, 1228-1238, doi:10.1016/j.apsb.2020.04.009 (2020).

107 Yaron, T. M. *et al.* The FDA-approved drug Alectinib compromises SARS-CoV-2 nucleocapsid phosphorylation and inhibits viral infection in vitro. *bioRxiv*, doi:10.1101/2020.08.14.251207 (2020).

108 Koetzner, C. A., Hurst-Hess, K. R., Kuo, L. & Masters, P. S. Analysis of a crucial interaction between the coronavirus nucleocapsid protein and the major membrane-bound subunit of the viral replicase-transcriptase complex. *Virology* **567**, 1-14, doi:10.1016/j.virol.2021.12.004 (2022).

109 Bessa, L. M. *et al.* The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci Adv* **8**, eabm4034, doi:10.1126/sciadv.abm4034 (2022).

110 Lu, S. *et al.* The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat Commun* **12**, 502, doi:10.1038/s41467-020-20768-y (2021).

111 Perdikari, T. M. *et al.* SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J* **39**, e106478, doi:10.15252/embj.2020106478 (2020).

112 Iserman, C. *et al.* Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol Cell* **80**, 1078-1091 e1076, doi:10.1016/j.molcel.2020.11.041 (2020).

113 Carlson, C. R. *et al.* Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Mol Cell* **80**, 1092-1103 e1094, doi:10.1016/j.molcel.2020.11.025 (2020).

114 Barcena, M. *et al.* Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirion. *Proc Natl Acad Sci U S A* **106**, 582-587, doi:10.1073/pnas.0805270106 (2009).

115 Davies, H. A., Dourmashkin, R. R. & Macnaughton, M. R. Ribonucleoprotein of avian infectious bronchitis virus. *J Gen Virol* **53**, 67-74, doi:10.1099/0022-1317-53-1-67 (1981).

116 Macneughton, M. R. & Davies, H. A. Ribonucleoprotein-like structures from coronavirus particles. *J Gen Virol* **39**, 545-549, doi:10.1099/0022-1317-39-3-545 (1978).

117 Yao, H. *et al.* Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **183**, 730-738 e713, doi:10.1016/j.cell.2020.09.018 (2020).

118 Klein, S. *et al.* SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun* **11**, 5885, doi:10.1038/s41467-020-19619-7 (2020).

119 Huston, N. C. *et al.* Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol Cell* **81**, 584-598 e585, doi:10.1016/j.molcel.2020.12.041 (2021).

120 Cao, C. *et al.* The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat Commun* **12**, 3917, doi:10.1038/s41467-021-22785-x (2021).

121 Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* **5**, 53-55, doi:10.1038/nmeth1139 (2008).

122 Nikolakaki, E. & Giannakouros, T. SR/RS Motifs as Critical Determinants of Coronavirus Life Cycle. *Front Mol Biosci* **7**, 219, doi:10.3389/fmolb.2020.00219 (2020).

123 Luo, H. *et al.* The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *FEBS Lett* **579**, 2623-2628, doi:10.1016/j.febslet.2005.03.080 (2005).

124 Bar-On, Y. M., Flamholz, A., Phillips, R. & Milo, R. SARS-CoV-2 (COVID-19) by the numbers. *Elife* **9**, doi:10.7554/eLife.57309 (2020).

125 Neuman, B. W. *et al.* A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol* **174**, 11-22, doi:10.1016/j.jsb.2010.11.021 (2011).

126 Kuo, L. & Masters, P. S. Functional analysis of the murine coronavirus genomic RNA packaging signal. *J Virol* **87**, 5182-5192, doi:10.1128/JVI.00100-13 (2013).

127 Verma, S., Bednar, V., Blount, A. & Hogue, B. G. Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. *J Virol* **80**, 4344-4355, doi:10.1128/JVI.80.9.4344-4355.2006 (2006).

128 Hurst, K. R. *et al.* A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. *J Virol* **79**, 13285-13297, doi:10.1128/JVI.79.21.13285-13297.2005 (2005).

129 Kuo, L. & Masters, P. S. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. *J Virol* **76**, 4987-4999, doi:10.1128/jvi.76.10.4987-4999.2002 (2002).

130 Narayanan, K., Maeda, A., Maeda, J. & Makino, S. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. *J Virol* **74**, 8127-8134, doi:10.1128/jvi.74.17.8127-8134.2000 (2000).

131 Wang, X., Yang, Y., Sun, Z. & Zhou, X. Crystal structure of the membrane (M) protein from a SARS-CoV-2-related *Betacoronavirus*. *BioRxiv*, doi:10.1101/2022.06.28.497981 (2022).

132 Dolan, K. A. *et al.* Structure of SARS-CoV-2 M protein in lipid nanodiscs. *BioRxiv*, doi:https://doi.org/10.1101/2022.06.12.495841 (2022).

133 Kuo, L., Koetzner, C. A. & Masters, P. S. A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. *Virology* **494**, 100-107, doi:10.1016/j.virol.2016.04.009 (2016).

134 Kuo, L., Koetzner, C. A., Hurst, K. R. & Masters, P. S. Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid protein. *J Virol* **88**, 4451-4465, doi:10.1128/JVI.03866-13 (2014).

135 Liu, X. *et al.* Targeting the coronavirus nucleocapsid protein through GSK-3 inhibition. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2113401118 (2021).

136 Zhou, C. Y. & Narlikar, G. J. Analysis of Nucleosome Sliding by ATP-Dependent Chromatin Remodeling Enzymes. *Methods Enzymol* **573**, 119-135, doi:10.1016/bs.mie.2016.01.015 (2016).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Christopher Carlson
37D58C87BACD49B...
Author Signature

9/13/2022
Date