Title

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

Journal

Journal of Biological Chemistry, 298(11)

ISSN

0021-9258

Authors

Carlson, Christopher R
Adly, Armin N
Bi, Maxine
et al.

Publication Date

2022-11-01

DOI

10.1016/j.jbc.2022.102560

Peer reviewed

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

Received for publication, August 31, 2022, and in revised form, September 17, 2022 Published, Papers in Press, October 4, 2022,
https://doi.org/10.1016/j.jbc.2022.102560

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,*]

From the [1]Department of Physiology, and [2]Department of Biochemistry & Biophysics, University of California, San Francisco, California, USA

Edited by Craig Cameron

The nucleocapsid (N) protein of severe acute respiratory syndrome coronavirus 2 is responsible for compaction of the ∼30-kb RNA genome in the ∼90-nm virion. Previous studies suggest that each virion contains 35 to 40 viral ribonucleoprotein (vRNP) complexes, or ribonucleosomes, arrayed along the genome. There is, however, little mechanistic understanding of the vRNP complex. Here, we show that N protein, when combined *in vitro* with short fragments of the viral genome, forms 15-nm particles similar to the vRNP structures observed within virions. These vRNPs depend on regions of N protein that promote protein–RNA and protein–protein interactions. Phosphorylation of N protein in its disordered serine/arginine region weakens these interactions to generate less compact vRNPs. We propose that unmodified N protein binds structurally diverse regions in genomic RNA to form compact vRNPs within the nucleocapsid, while phosphorylation alters vRNP structure to support other N protein functions in viral transcription.

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 (1). The ∼30-kb single-stranded RNA genome is packed inside the virus in a structure called the nucleocapsid (2, 3). Following infection and genomic RNA unpackaging, the first two-thirds of the genome is translated to produce numerous nonstructural proteins 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 (4–7). The final third of the genome then serves as a template for generation of the four structural proteins that form the mature virus (8–10).

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 (9). 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 (4, 6, 11–14). 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 (16–19).

The 46 kDa N protein contains two globular domains flanked by three regions of intrinsic disorder (Fig. 1*A*) (20). The N-terminal domain and the C-terminal domain (CTD) bind RNA and are highly conserved among coronaviruses (21–26). In solution, N protein exists predominantly as a dimer due to a high-affinity dimerization interface on the CTD and also forms tetramers and higher-order oligomers that are modulated by the disordered N-terminal extension (NTE) and C-terminal extension (CTE, Fig. 1*A*) (21, 23, 26–29). 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, 16, 30–32). 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 (11, 33–35).

Purified N protein and viral RNA form biomolecular condensates *in vitro* (36–39). Unphosphorylated N protein forms gel-like condensates containing discrete substructures, perhaps reflecting N protein function in the nucleocapsid (36, 38, 39). Phosphorylated N forms more liquid-like condensates that are reminiscent of dynamic N protein foci seen at the RTC (36, 39, 40). Phosphorylated N protein condensates might therefore provide a local compartment that facilitates

* For correspondence: David O. Morgan, David.Morgan@ucsf.edu.

viral RNA processing at the RTC, but this possibility remains untested.

During viral assembly, hypophosphorylated 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 (4, 5, 16, 31). Early electron microscopy (EM) studies of coronavirus nucleocapsids demonstrated the existence of viral ribonucleoprotein (vRNP) complexes aligned helically along an RNA strand (41–43). Recent cryo-electron tomography studies of intact SARS-CoV-2 virions revealed that each virus contains 35 to 40 discrete nucleosome-like vRNP complexes (44, 45). These vRNPs or ribonucleosomes are ∼15 nm in diameter and, through low resolution modeling efforts, are speculated to contain 12 N proteins in complex with up to 800 nt of RNA (30,000 nucleotides [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 (46, 47). 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 (39). 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 (39, 44, 45). 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. 1*B*), indicating that N protein bound the RNAs and retarded their electrophoretic mobility.

We used mass photometry to 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. 1*C*). 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. 1*C*). The 5′-600 RNA was therefore chosen as a representative viral RNA for further study.

To purify the vRNP complex for more detailed analysis, we used velocity sedimentation on a glycerol gradient, which is commonly used for purification of nucleosomes (48). N protein was mixed with 5′-600 RNA and separated by centrifugation on a 10 to 40% glycerol gradient. Individual fractions were analyzed by native gel electrophoresis (Fig. 1*D*, top). We observed a broad RNA–protein peak that migrated in the gel at the same size (at the 1000 nt marker) as that seen in native gel electrophoresis of unpurified complexes (Fig. 1*B*). Peak 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. 1*E*, top; Table S1). These peaks likely correspond to free N protein dimer (predicted mass 91.2 kDa; see Fig. 2*D*, 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 to crosslink the protein complex during centrifugation (a technique known as gradient fixation, or GraFix) (49). Analysis of the GraFix-purified fractions by native gel electrophoresis revealed sharper, more discrete bands compared to the noncrosslinked sample (Fig. 1*D*, bottom). The distribution of vRNP complexes across the gradient was similar in the two conditions, although crosslinked preparations contained additional large species at the bottom of the gradient, which are likely to represent crosslinked complexes of multiple vRNPs. The GraFix-purified main peak (fractions 7 + 8) was analyzed by mass photometry, revealing one sharp peak with an approximate mass of 719 ± 7 kDa (Fig. 1*E*, bottom; Table S1). This is consistent with the idea that the noncrosslinked sample dissociates upon dilution for mass photometry and is consistent with a 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 EM of the GraFix-purified sample revealed discrete 15-nm particles with an electron-dense center

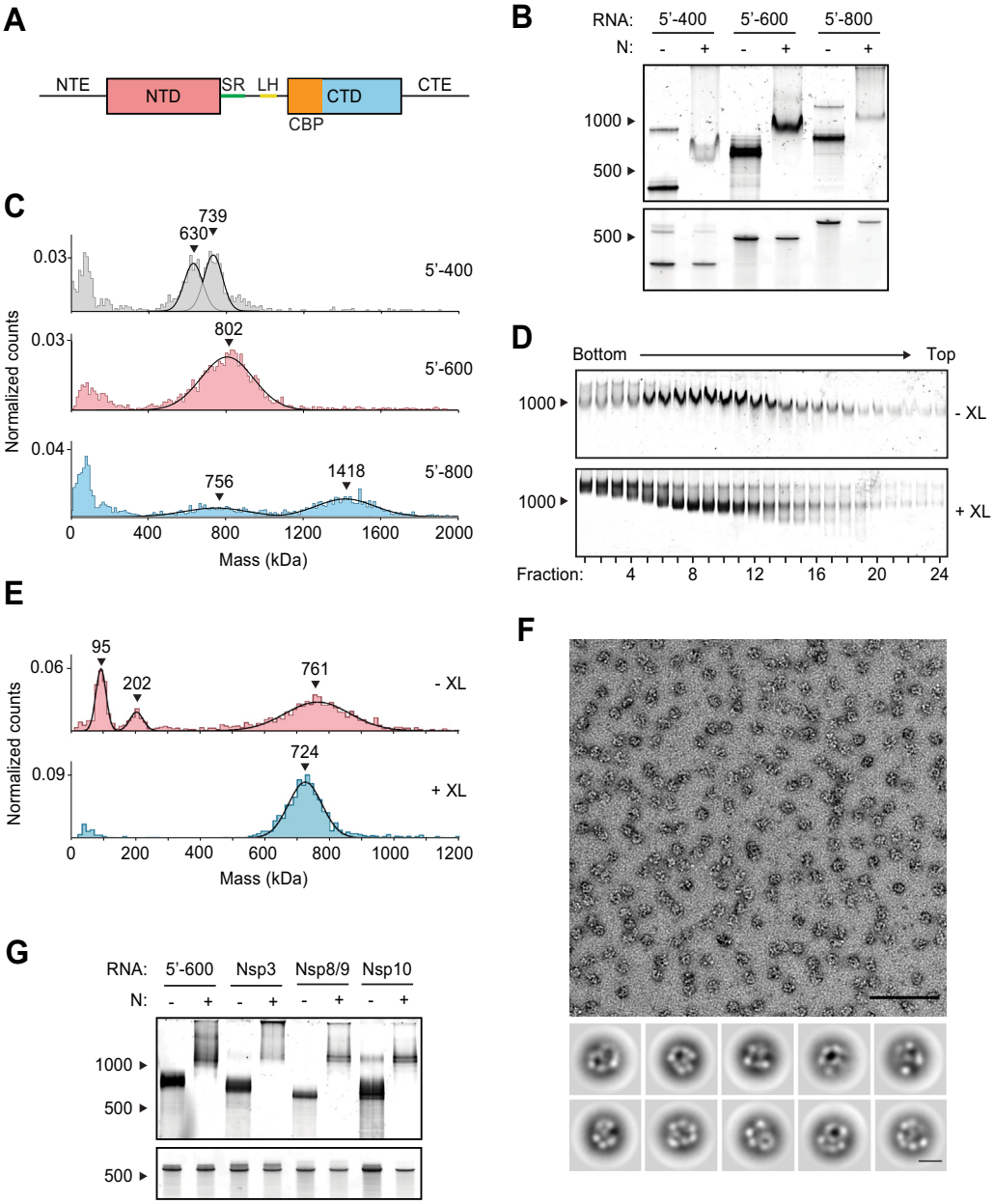
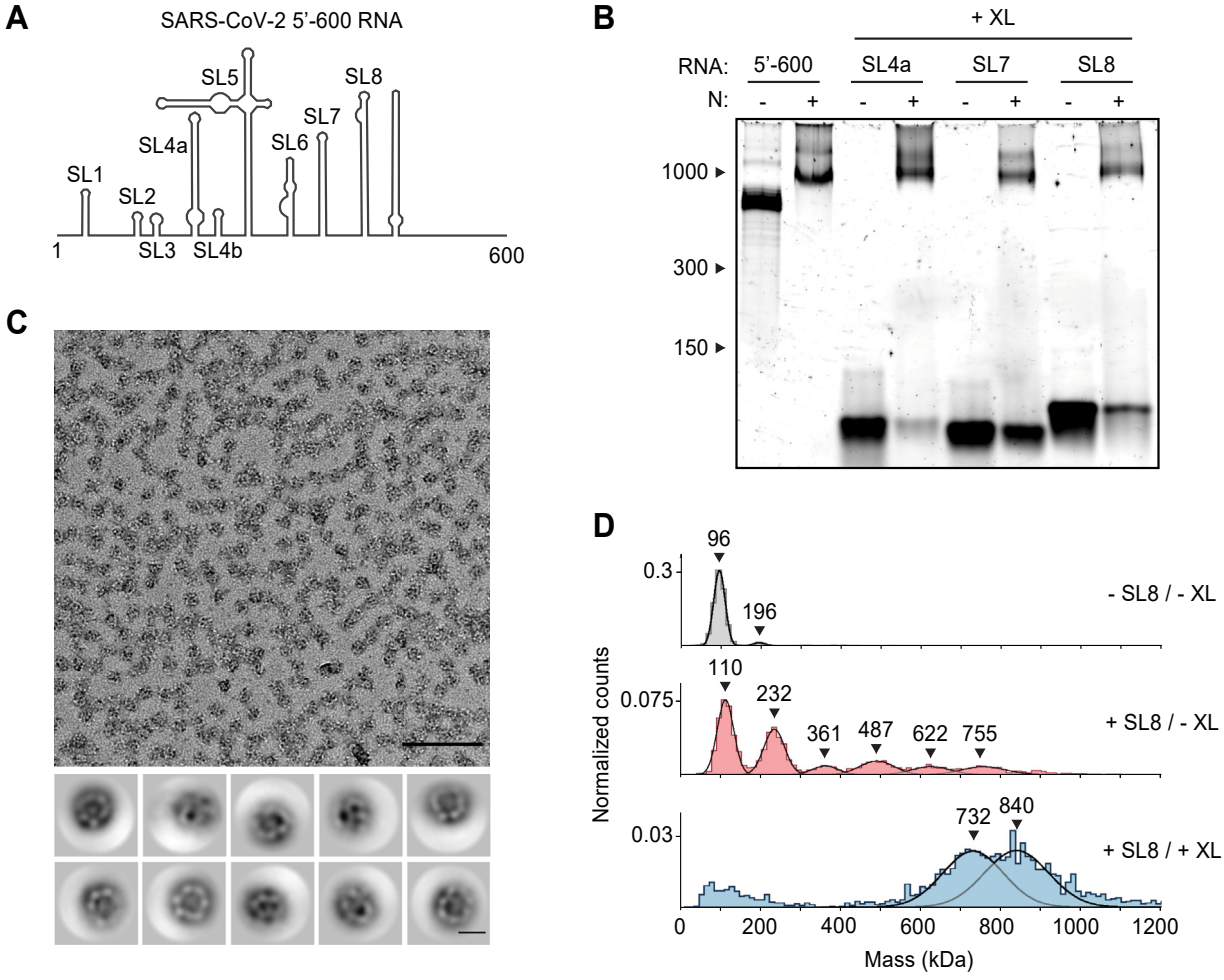


Figure 1. Viral RNA promotes formation of the SARS-CoV-2 ribonucleosome. *A*, schematic of N protein domain architecture, including the N-terminal extension (NTE), N-terminal domain (NTD), serine/arginine region (SR), leucine helix (LH), C-terminal basic patch (CBP), C-terminal domain (CTD), and C-terminal 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). RNA concentration in these and other experiments was 256 ng/μl, regardless of RNA length, to ensure that all mixtures contain the same nucleotide concentration. *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 S1). *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). RNA length standard shown on *left* (nt). *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 S1). *F*, negative stain electron microscopy and two-dimensional classification of GraFix-purified vRNPs (combined fractions 7 and 8 from *D*). Scale bars are 100 nm (*top*) and 10 nm (*bottom*). *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. RNA length standards shown on *left* (nt). See Table S2 for sequences. N, nucleocapsid; vRNP, viral ribonucleoprotein.

surrounded by an outer ring (Fig. 1*F*). 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*A*). 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 (44, 45).

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 7800–8400), (3) Nsp8/9 (nucleotides 12250–12850), and (4), Nsp10/12 (nucleotides 13200–13800). All RNAs appeared to form vRNPs (Fig. 1*G*). Each of these regions is known to contain a diverse



E

+ SL8 / - XL

# N (45.6 kDa)	2	4	6	8	10	12
# SL8 (23.1 kDa)	1	2	4	5	6	8
Expected (kDa)	114.4	228.7	366.2	480.5	596.7	732.3
Measured (kDa)	108 ± 4	225 ± 10	360 ± 2	468 ± 28	600 ± 31	736 ± 27

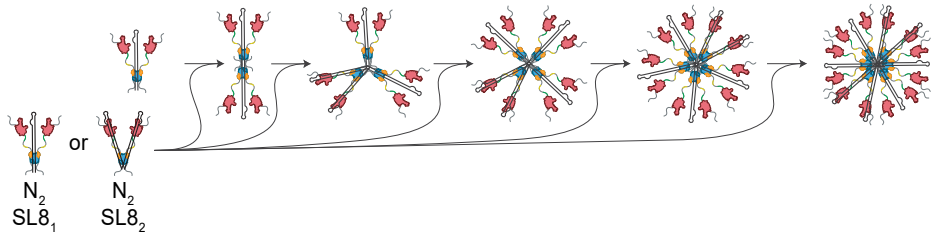


Figure 2. Stem-loop RNA, in complex with N protein, drives ribonucleosome formation. *A*, schematic of RNA secondary structure in the 5′-600 RNA (46). *B*, native gel analysis of 15 μM N protein mixed with 256 ng/μl of the indicated RNAs. Samples containing stem-loop RNA were crosslinked to stabilize the resulting complex, while the 5′-600 RNA sample was left un-crosslinked. RNA length standards shown on left (nt). Corresponding denaturing gel analysis shown in Fig. S1*A*. *C*, fractions 7 and 8 of GraFix-purified SL8 assembled vRNPs were combined and analyzed by negative stain electron microscopy and two-dimensional classification. Scale bars are 100 nm (*top*) and 10 nm (*bottom*). *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. S1*B* for ease of comparison). Representative of two independent experiments (Table S1). *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 S1). 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. N, nucleocapsid; vRNP, viral ribonucleoprotein.

array of secondary structure in infected cells and in the mature virion (46, 47). 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*A*) (46, 50). 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 (Figs. 2*B* and S1*A*). 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. S1*B*, top; Table S1). SL7 generated three broad mass peaks at 502 ± 26 kDa, 615 ± 22 kDa, and 713 ± 13 kDa (Fig. S1*B*, middle; Table S1). SL8 generated two broad peaks at 737 ± 11 kDa and 840 ± 12 kDa (Fig. S1*B*, bottom; Table S1) and was chosen for further analysis due to less heterogeneity in the composition of the complex.

SL8-containing vRNPs were purified by GraFix (Fig. S1*C*). Analysis of peak fractions (7 + 8) by mass photometry (Fig. S1*D*) indicated that SL8 vRNPs were similar in mass to vRNPs assembled with 5′-600 RNA (Fig. 1*E*). Negative stain EM and two-dimensional class averages revealed ring structures that resemble vRNPs assembled with the 5′-600 RNA (Fig. 2*C* compared to Fig. 1*F*).

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*A*, bottom). The mSL8 RNA promoted vRNP formation across a range of N protein concentrations (Fig. S2*A*). Analysis by GraFix and mass photometry confirmed that mSL8 forms a heterogeneous vRNP that is similar in size to vRNPs containing SL8 (Fig. S2, *B* and *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 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 noncrosslinked 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 to 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*D*, middle; Table S1). N protein alone exists primarily as a ∼96 kDa dimer (97 ± 1 kDa) at the low concentration used for mass photometry (Fig. 2*D*, top; Table S1; predicted mass 91.2 kDa), so the ∼110 kDa peak likely represents 1 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 to 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 12 N proteins and 6 to 12 stem-loop RNAs (Fig. 2*E*). These data support the possibility that the vRNP assembled with 5′-600 RNA (Fig. 1*E*) 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. S1*B*, bottom; also shown in Fig. 2*D*, bottom). Based on the similar molecular mass of the smaller peak in the crosslinked sample (737 ± 11 kDa) to the noncrosslinked 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 NTE, a poorly conserved disordered sequence (39, 51); (2) The conserved 31-aa SR region that has been implicated in RNA binding, oligomerization, and phosphorylation (15, 16, 30–32, 52–54); (3) the 20-aa leucine helix (LH), an alpha helix downstream of the SR region that interacts with Nsp3 (34); (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 (34) and helical stacking of N protein (23, 55); and (5) the 55-aa CTE, which has been implicated in tetramerization and oligomerization of N (21, 52, 54, 56) (Figs. 3*A* and S3*A*).

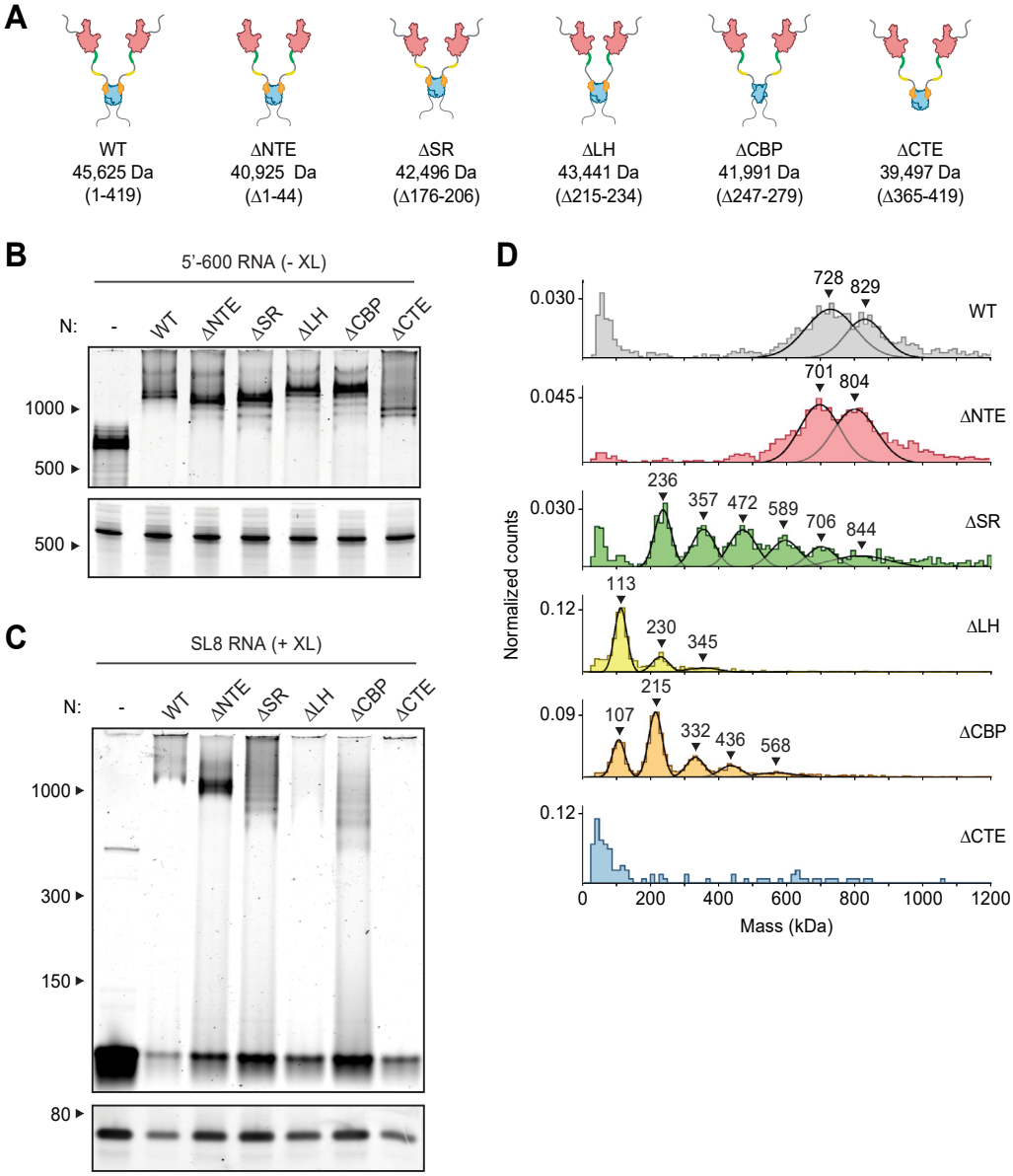


Figure 3. Multiple N protein regions contribute to vRNP formation. *A*, schematic of wildtype (WT) N protein and deletion mutants, as described in the text. Mass is that of monomeric N protein. *B*, 15 μM N protein mutants were mixed with 256 ng/μl 5′-600 RNA and analyzed by native (*top*) and denaturing (*bottom*) gel electrophoresis. RNA length standards shown on *left* (nt). *C*, 20 μM N protein mutants were mixed with 256 ng/μl SL8 RNA, crosslinked, and analyzed by native (*top*) and denaturing (*bottom*) gel electrophoresis. RNA length standards shown on *left* (nt). *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 S1). CBP, C-terminal basic patch; CTD, C-terminal domain; CTE, C-terminal extension; LH, leucine helix; N, nucleocapsid; NTD, N-terminal domain; NTE, N-terminal extension; SR, serine/arginine region; vRNP, viral ribonucleoprotein.

Mutant N proteins were mixed with 5′-600 RNA and analyzed by native gel electrophoresis (Fig. 3*B*). All mutant N proteins, with the exception of the CTE deletion, appeared to form fully assembled vRNPs. Most mutants displayed a small population of lower bands beneath the fully shifted vRNP. These lower bands might represent subcomplexes 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. 3, *C* and *D* and Table S1). Deletion of the NTE had little effect, other than to decrease the size of the vRNP complexes by ∼30 to 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. 3*C*). These mutants appeared cloudy, and turbidity analysis revealed a higher absorbance at 340 nm compared to wildtype, suggesting formation of biomolecular condensates (Fig. S3*B*) (39). 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. 3*D* and Table S1). The smallest peak represents an N dimer bound to one SL8 RNA, with the next two representing stepwise additions of one or 2 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. 3*D*), 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 or CBP region 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 to 130 kDa increases in molecular mass revealed by mass photometry (Fig. 3, *C* and *D* and Table S1). 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 1 N protein dimer bound to one SL8 RNA molecule. Interestingly, the SR deletion resulted in a minimal ribonucleoprotein complex of ∼230 kDa, consistent with 1 N protein tetramer bound to two SL8 RNAs.

To test if SR deletion causes defects by reducing the length of the central disordered region, we also constructed an 'SR linker' mutant in which the 31-aa SR region is replaced with a random sequence of glycines, alanines, and serines. This mutant displayed defects similar to those seen with the SR deletion, confirming that specific sequence features of the SR region are required for vRNP formation (Fig. S3*C*).

Native gel analysis revealed an increase in free SL8 RNA in some mutant N protein samples compared to wildtype (Figs. 3*C* and S3*C*), 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, 16, 30–32, 57). In contrast, N protein in the virion is thought to be poorly phosphorylated (16, 31). We previously observed defects in vRNP formation when 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) (39), 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. 4*A*, left). The mutant formed an appropriately sized vRNP, as well as minor subcomplexes 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 wildtype N protein (Figs. 4*B* and 1*D*, bottom). Mass photometry of fractions 7 + 8 confirmed a similar mass of the 10D and wildtype vRNPs, apart from a minor ∼830 kDa peak observed with the 10D mutant (Figs. S4*A* and S1*E*, 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 wildtype vRNPs (Fig. 4*C* compared with Fig. 1*F*). The 10D complex appears extended and heterogeneous, unlike the compact structure of the wildtype 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 wildtype 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. 4*A*, right, and Fig. 4*D*). 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 wildtype N (Fig. 4*E* compared to Fig. S1*C*). This result was confirmed by mass photometry analysis of fractions 19 + 20 (Fig. S4*B*). 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 wildtype N (Fig. 4*F* compared to Fig. 2*C*).

We next tested vRNP assembly with N protein that had been phosphorylated *in vitro*. Multiple protein kinases are thought to act in sequence to catalyze N protein phosphorylation (15, 16, 30, 31, 39, 57, 58). In recent work, Yaron *et al.* (30) elegantly demonstrated a multikinase cascade that results in maximally phosphorylated N protein: serine-arginine protein kinase (SRPK) phosphorylates S188 and S206, which primes the protein for subsequent phosphorylation of eight more sites within the SR by glycogen-synthase kinase 3 (GSK3), which then primes a final four sites for phosphorylation by casein kinase 1 (CK1) (Fig. 5*A*). Consistent with this model, we observed maximal phosphorylation of N in the presence of all three kinases (Fig. 5*B*). Phosphorylation was greatly reduced when both SRPK priming sites were mutated to alanine (S188A + S206A mutant) (Fig. 5*B*). We mixed kinase-treated wildtype or S188A + S206A N proteins with SL8 RNA and purified the resulting vRNP complexes by GraFix (Fig. 5*C*). Wildtype phosphorylated N protein migrated as a low molecular weight ribonucleoprotein complex similar

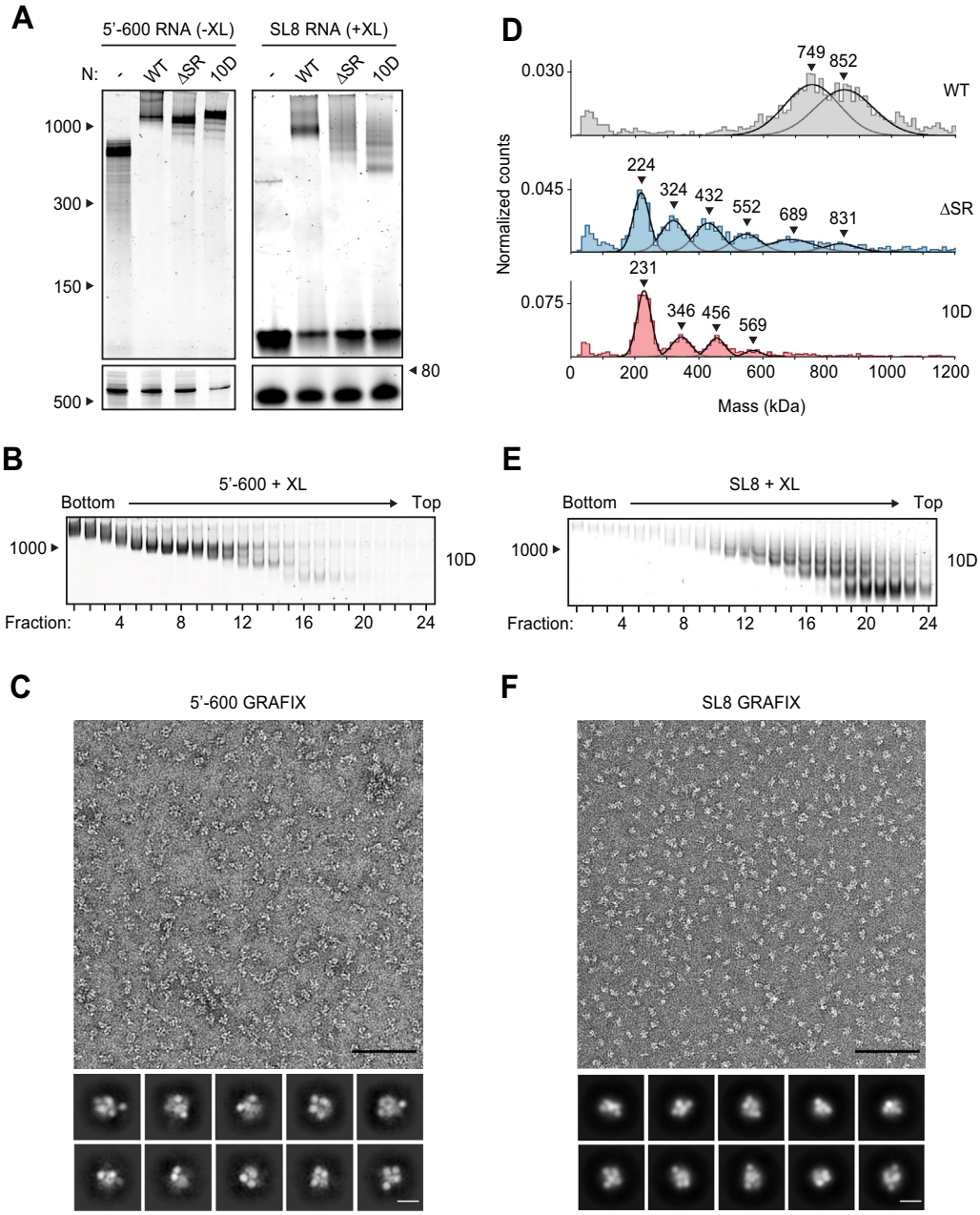


Figure 4. Phosphomimetic mutations in the SR region of N prevent vRNP assembly. *A*, 15 μM N protein constructs were combined with 256 ng/μl 5′-600 RNA (*left*) or SL8 RNA (*right*) and analyzed by native (*top*) and denaturing (*bottom*) gel electrophoresis. SL8 ribonucleoprotein complexes were crosslinked prior to native gel electrophoresis. RNA length standards shown on *left* (nt). See Figure 5*A* 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. RNA length standard shown on *left* (nt). *C*, fractions 7 and 8 of GraFix-separated vRNPs (from *B*) were combined and analyzed by negative stain electron microscopy and two-dimensional classification. Scale bars are 100 nm (*top*) and 10 nm (*bottom*). *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 S1). A separate analysis of ΔSR mutant is also shown in Figure 3*D*. *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. Scale bars are 100 nm (*top*) and 10 nm (*bottom*). N, nucleocapsid; SR, serine/arginine region; vRNP, viral ribonucleoprotein; WT, wild type.

to the 10D mutant. The poorly phosphorylated S188A + S206A mutant, however, formed a vRNP similar to wildtype unphosphorylated N protein (Fig. 5*C*). Mass photometry of the GraFix-purified samples further substantiated the defect in wildtype phospho-N vRNP assembly (Fig. 5*D*, top), which is rescued by mutation of the two priming phosphorylation sites (the S188A + S206A mutant) (Fig. 5*D*, 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 (44, 45), seem to

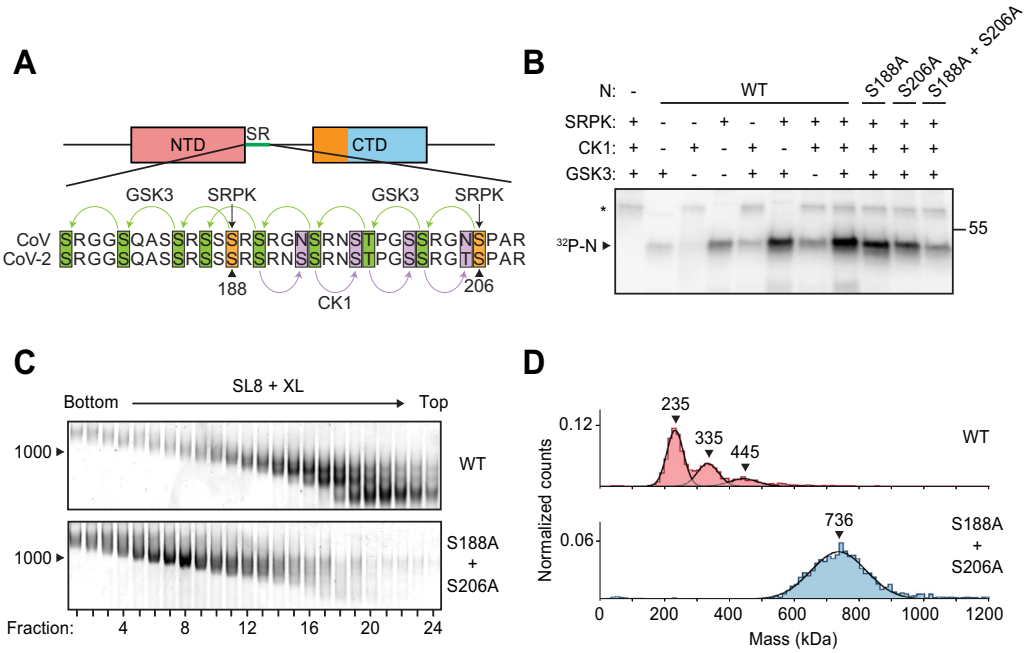


Figure 5. Phosphorylation of N protein inhibits ribonucleosome formation. *A*, sequence of N protein SR regions from SARS-CoV (aa 177–210) and SARS-CoV-2 (aa 176–209). The proposed mechanism of sequential phosphorylation (30) is initiated by SRPK at S188 and S206 (*orange*), which leads to upstream phosphorylation of eight sites by GSK3 (*green*), allowing for final phosphorylation of four additional sites by CK1 (*purple*). In the phosphomimetic 10D mutant used in Figure 4, the SRPK and GSK3 sites are changed to aspartic acid. *B*, wildtype (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. RNA length standard shown on left (nt). *D*, peak fractions from the GraFix analyses in *C* were analyzed by mass photometry. *Top*, fractions 19 + 20 of wildtype N; *bottom*, fractions 7 + 8 of S188A + S206A mutant N. Representative of two independent experiments (Table S1). CK1, casein kinase 1; GSK3, glycogen-synthase kinase 3; N, nucleocapsid; SR, serine/arginine region; SRPK, serine-arginine protein kinase.

contain 12 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 (54), LH binding to other regions of the N protein, helical stacking of the CBP (23), and tetramerization driven by the CTE (21, 52, 56, 59). 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 12 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 (46, 47), 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 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 (44, 45), 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. The total number of N proteins in a virus is not well defined but has been estimated at 730 to 2200 copies (60). The viral genome is 30,000 nt in length. Thus, it seems likely that some N proteins and RNA in the virion are not incorporated in 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 (41, 44, 60). 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 (36, 38, 39).

During viral assembly, one copy of the ∼30 kb viral genome is packaged in each virus, while cellular and subgenomic viral RNAs are excluded (61). 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 (62). 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 M protein likely functions in this capacity.

M protein is a 25 kDa structural protein containing three transmembrane helices followed by a ∼100 aa CTD that faces the interior of the virion and is thought to interact with the C terminus of N protein (61, 63–69). This interaction is required for maintaining packaging specificity in MHV (70, 71). The soluble CTD of M protein triggers RNA-independent phase separation when mixed with N protein (36), 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 (44, 45, 60). With these lines of evidence in mind, it is 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 (66). 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, 16, 30, 31, 39, 57, 58). Yaron *et al.* (30) recently provided evidence for sequential phosphorylation of N by SRPK, GSK3, and CK1 (Fig. 5*A*). 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 (30). 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 (72). 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.

Experimental procedures

N protein preparation

Wildtype and mutant N proteins were produced as described previously (39). 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 absorbance 0.6, and induced with 0.4 mM IPTG. Cells were harvested, washed with PBS, snap frozen in LN_2, and stored at −80 °C until use. Thawed cells were resuspended in buffer A (50 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol, and 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 4 °C. Ni-NTA beads were washed three times with ten bed volumes of buffer A and eluted with buffer B (50 mM Hepes pH 7.5, 500 mM NaCl, 10% glycerol, 250 mM imidazole, and 6 M 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 −80 °C.

RNA preparation

Sequences of all RNAs used in this study are provided in Table S2. 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 preheated 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. RNA concentration was 256 ng/μl, regardless of RNA length, to ensure that all mixtures contained the same nucleotide concentration. Samples containing stem-loop RNAs (SL4a, SL7, SL8, and mSL8) 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, and 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, and 10% glycerol). Diluted vRNP mixture (2 μl) was loaded onto a 5% polyacrylamide native TBE gel (Bio-Rad) and run at 125 V for 80 min at 4 °C. Another aliquot of diluted sample (1 μl) was 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. Reaction buffer (10 μl) (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.

Glycerol gradient centrifugation

Glycerol gradients were assembled as previously described, with slight modifications (48). Briefly, 10 to 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.

Negative stain electron microscopy

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

52,000× 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.

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.

Protein kinase reactions

Protein kinases were purchased from Promega (SRPK1: #VA7558, GSK-3β: #V1991, CK1ε: V4160). N protein (1.25 μM) was incubated with 80 nM kinase for 30 min at 30 °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 30 °C 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.

Data availability

All data are included in the article or available from the corresponding author D. O. M.

Supporting information—This article contains supporting information.

Acknowledgments—We thank the members of the Morgan laboratory for discussions and comments on the manuscript and Elise Muñoz, Hayden Saunders, and Rhiju Das for discussions and technical assistance.

Author contributions—C. R. C. and D. O. M. conceptualization; C. R. C., A. N. A., M. B., and C. J. H. methodology; C. R. C., A. N. A., M. B., and C. J. H. validation; C. R. C., A. N. A., M. B., and C. J. H. investigation; C. R. C. and D. O. M. writing-original draft; A. F., Y. C., and D. O. M. supervision.

Funding and additional information—This work was supported by a graduate research fellowship from ARCS (C. R. C.), and grants (to D. O. M.) from the National Institute of General Medical Sciences (R35-GM118053) and the UCSF Program for Breakthrough Biomedical Research, which is funded in part by the Sandler Foundation. A. F. is supported by the Chan Zuckerberg Biohub and the QBI Coronavirus Research Group at UCSF. Y. C. is an investigator of the Howard Hughes Medical Institute.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: CBP, CTD basic patch; CK1, casein kinase 1; CTD, C-terminal domain; CTE, C-terminal extension; E, envelope; EM, electron microscopy; GraFix, gradient fixation; GSK3, glycogen-synthase kinase 3; LH, leucine helix; M, membrane; MHV, murine hepatitis virus; N, nucleocapsid; Nsp, nonstructural protein; nt, nucleotides; NTD, N-terminal domain; NTE, N-terminal extension; RTC, replication–transcription complex; S, spike; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SR, serine/arginine; SRPK, serine-arginine protein kinase; vRNP, viral ribonucleoprotein.

References

1. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273
2. Fung, T. S., and Liu, D. X. (2019) Human coronavirus: host-pathogen interaction. *Annu. Rev. Microbiol.* **73**, 529–557
3. Masters, P. S. (2006) The molecular biology of coronaviruses. *Adv. Virus Res.* **66**, 193–292
4. Scherer, K. M., Mascheroni, L., Carnell, G. W., Wunderlich, L. C. S., Makarchuk, S., Brockhoff, M., *et al.* (2022) SARS-CoV-2 nucleocapsid protein adheres to replication organelles before viral assembly at the Golgi/ERGIC and lysosome-mediated egress. *Sci. Adv.* **8**, eabl4895
5. V'Kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170
6. Snijder, E. J., Limpens, R., de Wilde, A. H., de Jong, A. W. M., Zevenhoven-Dobbe, J. C., Maier, H. J., *et al.* (2020) A unifying structural and functional model of the coronavirus replication organelle: tracking down RNA synthesis. *PLoS Biol.* **18**, e3000715
7. Snijder, E. J., van der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J. J., van der Meulen, J., Koerten, H. K., *et al.* (2006) Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. *J. Virol.* **80**, 5927–5940
8. Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., and Chang, H. (2020) The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921.e10
9. Sola, I., Almazan, F., Zuniga, S., and Enjuanes, L. (2015) Continuous and discontinuous RNA synthesis in coronaviruses. *Annu. Rev. Virol.* **2**, 265–288
10. Sawicki, S. G., Sawicki, D. L., and Siddell, S. G. (2007) A contemporary view of coronavirus transcription. *J. Virol.* **81**, 20–29
11. Cong, Y., Ulasli, M., Schepers, H., Mauthe, M., V'Kovski, P., Kriegenburg, F., *et al.* (2020) Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. *J. Virol.* **94**, e01925-19
12. V'Kovski, P., Gerber, M., Kelly, J., Pfaender, S., Ebert, N., Braga Lagache, S., *et al.* (2019) Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* **8**, e42037
13. Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martinez-Sobrido, L., Garcia-Sastre, A., *et al.* (2007) The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* **361**, 304–315
14. Bost, A. G., Carnahan, R. H., Lu, X. T., and Denison, M. R. (2000) 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

15. Bouhaddou, M., Memon, D., Meyer, B., White, K. M., Rezelj, V. V., Correa Marrero, M., *et al.* (2020) The global phosphorylation landscape of SARS-CoV-2 infection. *Cell* **182**, 685–712.e19
16. Wu, C. H., Chen, P. J., and Yeh, S. H. (2014) Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* **16**, 462–472
17. Zuniga, S., Cruz, J. L., Sola, I., Mateos-Gomez, P. A., Palacio, L., and Enjuanes, L. (2010) Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J. Virol.* **84**, 2169–2175
18. Almazan, F., Galan, C., and Enjuanes, L. (2004) The nucleoprotein is required for efficient coronavirus genome replication. *J. Virol.* **78**, 12683–12688
19. McBride, R., van Zyl, M., and Fielding, B. C. (2014) The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991–3018
20. Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., and Huang, T. H. (2014) The SARS coronavirus nucleocapsid protein–forms and functions. *Antiviral Res.* **103**, 39–50
21. Ye, Q., West, A. M. V., Silletti, S., and Corbett, K. D. (2020) Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* **29**, 1890–1901
22. Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., *et al.* (2020) Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* **10**, 1228–1238
23. Chen, C. Y., Chang, C. K., Chang, Y. W., Sue, S. C., Bai, H. I., Riang, L., *et al.* (2007) 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
24. Jayaram, H., Fan, H., Bowman, B. R., Ooi, A., Jayaram, J., Collisson, E. W., *et al.* (2006) X-Ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. *J. Virol.* **80**, 6612–6620
25. Fan, H., Ooi, A., Tan, Y. W., Wang, S., Fang, S., Liu, D. X., *et al.* (2005) The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. *Structure* **13**, 1859–1868
26. Luo, H., Ye, F., Sun, T., Yue, L., Peng, S., Chen, J., *et al.* (2004) In vitro biochemical and thermodynamic characterization of nucleocapsid protein of SARS. *Biophys. Chem.* **112**, 15–25
27. Cong, Y., Kriegenburg, F., de Haan, C. A. M., and Reggiori, F. (2017) Coronavirus nucleocapsid proteins assemble constitutively in high molecular oligomers. *Sci. Rep.* **7**, 5740
28. Chang, C. K., Chen, C. M., Chiang, M. H., Hsu, Y. L., and Huang, T. H. (2013) Transient oligomerization of the SARS-CoV N protein–implication for virus ribonucleoprotein packaging. *PLoS One* **8**, e65045
29. He, R., Dobie, F., Ballantine, M., Leeson, A., Li, Y., Bastien, N., *et al.* (2004) Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem. Biophys. Res. Commun.* **316**, 476–483
30. [preprint] Yaron, T. M., Heaton, B. E., Levy, T. M., Johnson, J. L., Jordan, T. X., Cohen, B. M., *et al.* (2020) The FDA-approved drug alectinib compromises SARS-CoV-2 nucleocapsid phosphorylation and inhibits viral infection in vitro. *bioRxiv*. https://doi.org/10.1101/2020.08.14.251207
31. Wu, C. H., Yeh, S. H., Tsay, Y. G., Shieh, Y. H., Kao, C. L., Chen, Y. S., *et al.* (2009) Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. *J. Biol. Chem.* **284**, 5229–5239
32. Peng, T. Y., Lee, K. R., and Tarn, W. Y. (2008) 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
33. Koetzner, C. A., Hurst-Hess, K. R., Kuo, L., and Masters, P. S. (2022) 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
34. Bessa, L. M., Guseva, S., Camacho-Zarco, A. R., Salvi, N., Maurin, D., Perez, L. M., *et al.* (2022) The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a. *Sci. Adv.* **8**, eabm4034
35. Wolff, G., Limpens, R., Zevenhoven-Dobbe, J. C., Laugks, U., Zheng, S., de Jong, A. W. M., *et al.* (2020) A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **369**, 1395–1398
36. Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J. K., Yates, J. R., 3rd, *et al.* (2021) The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* **12**, 502
37. Perdikari, T. M., Murthy, A. C., Ryan, V. H., Watters, S., Naik, M. T., and Fawzi, N. L. (2020) SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J.* **39**, e106478
38. Iserman, C., Roden, C. A., Boerneke, M. A., Sealfon, R. S. G., McLaughlin, G. A., Jungreis, I., *et al.* (2020) Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol. Cell* **80**, 1078–1091.e6
39. Carlson, C. R., Asfaha, J. B., Ghent, C. M., Howard, C. J., Hartooni, N., Safari, M., *et al.* (2020) Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol. Cell* **80**, 1092–1103.e4
40. Verheije, M. H., Hagemeijer, M. C., Ulasli, M., Reggiori, F., Rottier, P. J., Masters, P. S., *et al.* (2010) The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes. *J. Virol.* **84**, 11575–11579
41. Barcena, M., Oostergetel, G. T., Bartelink, W., Faas, F. G., Verkleij, A., Rottier, P. J., *et al.* (2009) Cryo-electron tomography of mouse hepatitis virus: insights into the structure of the coronavirion. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 582–587
42. Davies, H. A., Dourmashkin, R. R., and Macnaughton, M. R. (1981) Ribonucleoprotein of avian infectious bronchitis virus. *J. Gen. Virol.* **53**, 67–74
43. Macneughton, M. R., and Davies, H. A. (1978) Ribonucleoprotein-like structures from coronavirus particles. *J. Gen. Virol.* **39**, 545–549
44. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., *et al.* (2020) Molecular architecture of the SARS-CoV-2 virus. *Cell* **183**, 730–738.e13
45. Klein, S., Cortese, M., Winter, S. L., Wachsmuth-Melm, M., Neufeldt, C. J., Cerikan, B., *et al.* (2020) SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat. Commun.* **11**, 5885
46. Huston, N. C., Wan, H., Strine, M. S., de Cesaris Araujo Tavares, R., Wilen, C. B., and Pyle, A. M. (2021) Comprehensive *in vivo* secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell* **81**, 584–598.e5
47. Cao, C., Cai, Z., Xiao, X., Rao, J., Chen, J., Hu, N., *et al.* (2021) The architecture of the SARS-CoV-2 RNA genome inside virion. *Nat. Commun.* **12**, 3917
48. Zhou, C. Y., and Narlikar, G. J. (2016) Analysis of nucleosome sliding by ATP-dependent chromatin remodeling enzymes. *Methods Enzymol.* **573**, 119–135
49. Kastner, B., Fischer, N., Golas, M. M., Sander, B., Dube, P., Boehringer, D., *et al.* (2008) GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55
50. Yang, D., and Leibowitz, J. L. (2015) The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **206**, 120–133
51. Chang, C. K., Hsu, Y. L., Chang, Y. H., Chao, F. A., Wu, M. C., Huang, Y. S., *et al.* (2009) 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
52. Hurst, K. R., Koetzner, C. A., and Masters, P. S. (2009) Identification of in vivo-interacting domains of the murine coronavirus nucleocapsid protein. *J. Virol.* **83**, 7221–7234
53. Grossoehme, N. E., Li, L., Keane, S. C., Liu, P., Dann, C. E., 3rd, Leibowitz, J. L., *et al.* (2009) 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
54. Luo, H., Ye, F., Chen, K., Shen, X., and Jiang, H. (2005) SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization. *Biochemistry* **44**, 15351–15358
55. Takeda, M., Chang, C. K., Ikeya, T., Guntert, P., Chang, Y. H., Hsu, Y. L., *et al.* (2008) 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
56. Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006) Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* **45**, 11827–11835
57. Nikolakaki, E., and Giannakouros, T. (2020) SR/RS motifs as critical determinants of coronavirus life cycle. *Front. Mol. Biosci.* **7**, 219
58. [preprint] Zhu, P., Franklin, R., Vogel, A., Stanisheuski, S., Reardon, P., Sluchanko, N. N., *et al.* (2021) PermaPhosSer: autonomous synthesis of functional, permanently phosphorylated proteins. *bioRxiv*. https://doi.org/10.1101/2021.10.22.465468
59. Luo, H., Chen, Q., Chen, J., Chen, K., Shen, X., and Jiang, H. (2005) The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. *FEBS Lett.* **579**, 2623–2628
60. Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., *et al.* (2011) A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **174**, 11–22
61. Masters, P. S. (2019) Coronavirus genomic RNA packaging. *Virology* **537**, 198–207
62. Kuo, L., and Masters, P. S. (2013) Functional analysis of the murine coronavirus genomic RNA packaging signal. *J. Virol.* **87**, 5182–5192
63. Verma, S., Bednar, V., Blount, A., and Hogue, B. G. (2006) Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. *J. Virol.* **80**, 4344–4355
64. Hurst, K. R., Kuo, L., Koetzner, C. A., Ye, R., Hsue, B., and Masters, P. S. (2005) A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. *J. Virol.* **79**, 13285–13297
65. Kuo, L., and Masters, P. S. (2002) 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
66. Narayanan, K., Maeda, A., Maeda, J., and Makino, S. (2000) Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. *J. Virol.* **74**, 8127–8134
67. [preprint] Wang, X., Yang, Y., Sun, Z., and Zhou, X. (2022) Crystal structure of the membrane (M) protein from a SARS-CoV-2-related *Betacoronavirus*. *bioRxiv*. https://doi.org/10.1101/2022.06.28.497981
68. [preprint] Dolan, K. A., Dutta, M., Kern, D. M., Kotecha, A., Voth, G. A., and Brohawn, S. G. (2022) Structure of SARS-CoV-2 M protein in lipid nanodiscs. *bioRxiv*. https://doi.org/10.1101/2022.06.12.495841
69. Zhang, Z., Nomura, N., Muramoto, Y., Ekimoto, T., Uemura, T., Liu, K., *et al.* (2022) Structure of SARS-CoV-2 membrane protein essential for virus assembly. *Nat. Commun.* **13**, 4399
70. Kuo, L., Koetzner, C. A., and Masters, P. S. (2016) A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. *Virology* **494**, 100–107
71. Kuo, L., Koetzner, C. A., Hurst, K. R., and Masters, P. S. (2014) 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
72. Liu, X., Verma, A., Garcia, G., Jr., Ramage, H., Lucas, A., Myers, R. L., *et al.* (2021) Targeting the coronavirus nucleocapsid protein through GSK-3 inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2113401118

Christopher R. Carlson is a PhD student in the lab of David Morgan at the University of California in San Francisco. At the onset of the COVID-19 pandemic, Chris became fascinated by the question of how coronaviruses package their large RNA genomes. Together with his collaborators, he began a 2-years exploration of this problem that led to the current paper. He is graduating soon and will move on to a position at Orbital Therapeutics, with the goal of developing the next generation of RNA delivery therapeutics. Twitter: @chris_carlson_.