SARS CoV-2 nucleocapsid protein forms condensates with viral genomic RNA

Amanda Jack[1], Luke S. Ferro[2], Michael J. Trnka[3], Eddie Wehri[4], Amrut Nadgir[5], Katelyn Costa, Julia Schaletzky[4], Ahmet Yildiz[1,2,5,†]

[1]Department of Biophysics, University of California, Berkeley CA.
[2]Department of Molecular and Cellular Biology, University of California, Berkeley CA.
[3]Department of Pharmaceutical Chemistry, University of California, San Francisco CA.
[4]Center for Emerging and Neglected Diseases, University of California, Berkeley CA.
[5]Department of Physics, University of California, Berkeley CA.
[†]Correspondence: yildiz@berkeley.edu

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection causes COVID-19, a pandemic that seriously threatens global health. SARS CoV-2 propagates by packaging its RNA genome into membrane enclosures in host cells. The packaging of the viral genome into the nascent virion is mediated by the nucleocapsid (N) protein, but the underlying mechanism remains unclear. Here, we show that the N protein forms biomolecular condensates with viral RNA both in vitro and in mammalian cells. While the N protein forms spherical assemblies with unstructured RNA, it forms mesh like-structures with viral RNA strands that contain secondary structure elements. Cross-linking mass spectrometry identified an intrinsically-disordered region that forms interactions between N proteins in condensates, and truncation of this region disrupts phase separation. By screening 1,200 FDA approved drugs in vitro, we identified a kinase inhibitor nilotinib, which affects the morphology of N condensates in vitro and disrupts phase separation of the N protein in vivo. These results indicate that the N protein compartmentalizes viral RNA in infected cells through liquid-liquid phase separation, and this process can be disrupted by a possible drug candidate.

Introduction

The majority of current efforts to treat COVID-19 have focused on targeting the interactions between the SARS-CoV-2 spike protein and the ACE2 receptor on human cells, which drive the initial infection of the host cells (*1-3*). However, not much work has been done to stop proliferation of the virus in host cells following infection. To propagate, the viral genome is first replicated inside the viral replication organelle, a double-membrane vesicle derived from the endoplasmic reticulum (*4*). Genomic RNA is then exported to the cytosol through a molecular pore complex and bound by viral N protein (*4*). The N protein restructures viral genomic RNA into shell-shaped structures (~15 nm in diameter), which contain approximately 12 N proteins and 800 nucleotides of viral RNA (*4, 5*). These viral ribonucleoprotein complexes (vRNPs) form asymmetric “beads on a string” structures which then bind to the viral membrane (M) protein on the surface of the ER-Golgi intermediate compartment (ERGIC) (*4, 6*). This interaction triggers the budding of the vRNP complex into the ERGIC, and the enveloped virus is released into the extracellular space via exocytosis.

The N protein could be a potential drug target because of its fundamental role in the packaging of the viral genome. However, the mechanism by which N remodels the viral RNA and packages it into a viral particle is not well understood. Previous studies showed that SARS-CoV N protein overexpressed in cells forms puncta that resemble biomolecular condensates (*7*). Similarly, in the SARS-CoV-2 virus, the N protein's domain organization and its involvement with phase-separated structures, such as stress granules (*8, 9*), suggest that N protein phase separates with the SARS-CoV-2 genome in infected cells.

Phase separation of N can form membrane-less compartments that mediate viral packaging in host cells (*10*).

Phase separation has been shown to create membraneless organelles, including the nucleolus, centrosome, heterochromatin, stress granules, and P granules when networks of weak interactions drive the de-mixing of proteins from the surrounding environment (*11, 12*). Nucleic acids (RNA, DNA) are highly involved in the formation of biomolecular condensates because they can scaffold multivalent interactions (*13-15*). Nucleotide sequences without the ability to base pair (e.g. poly-G) are stimulated to phase separate by weak electrostatic interactions between the polycation and the phosphate backbone (*13*).

The N protein contains many of the characteristic domain features common in phase separating systems, including an uneven electrostatic charge distribution, an RNA binding domain (RBD), a dimerization domain, and intrinsically disordered regions (Figure 1A). The RBD interacts nonspecifically with RNA and recognizes a nucleotide sequence in the 3' end of the viral genome (*16*). Both the N- and C-termini are disordered and can potentially form higher order protein/protein interactions (Figure 1B) (*17*). These disordered regions are arginine- and lysine-rich (Figure 1B), which may facilitate additional interactions with the negatively-charged RNA backbone (*18*) and drive biomolecular condensation of RNA (*19*). The N protein also contains a serine/arginine-rich (SR) motif, which has been associated with phase separation in other ribonucleoproteins (*16, 20*), and a predicted prion-like domain (PLD) that can potentially trigger protein demixing (*21*). However, the precise roles of these domains in packaging viral RNA remain to be elucidated.

The N protein could be a potential drug target because of its fundamental role in the packaging of the viral genome and possibility of targeting of liquid condensates small-molecule therapeutics (*22*). However, the mechanism by which N remodels the viral RNA and packages it into a viral particle is not well understood. In this study, we expressed the SARS-CoV-2 N protein in human embryonic kidney (HEK 293) cells and tested if it could create a biomolecular condensate with viral RNA *in vitro*. We observed that N and viral RNA form higher-order structures that resemble the "beads on a string" pattern observed in infected cells (*6*). The N protein nonspecifically interacts with RNA homopolymers to form liquid droplets, indicating that the length and structure of the RNA determine the macroscopic properties of the condensate. Cross-linking mass spectrometry (CLMS) identifies two regions flanking the dimerization domain where interactions are enriched within phase separated droplets. The deletion of the C-terminal oligomerization sequence in particular fully disrupts phase separation. Finally, we tested a library of FDA approved small molecules and identified drugs that affect the number, size, or shape of the N protein droplets. Of these, the kinase inhibitor nilotinib affects N protein condensation with viral RNA in vitro and disrupts condensates formed by the N protein in mammalian cells.

Results

The N protein phase separates with RNA

We first asked whether the N protein forms bimolecular condensates in the presence or absence of viral RNA in vitro. In the absence of RNA, N protein formed small puncta only at the highest protein concentration tested (72 μM) under physiological salt concentration (150 mM NaCl) (Figure 1D). However, N formed spherical droplets at lower salt with half maximal inhibition constant (IC_{50}) of 94 ± 1 μM NaCl (±s.e.) (Figure 1C). The addition of 50 ng/μL polyC homopolymer RNA substrate triggered phase separation of N even at physiological salt with a half saturation constant (k_{sat}) of 11.0 ± 0.7 μM, ± s.e., Figure 1D). These condensates were dissolved under higher salt concentrations (IC_{50} = 163 ± 8 mM NaCl), suggesting that they are liquid-like rather than solid structures (Figure 1C). Similar to polyC RNA, polyA and polyU homopolymer RNA substrates produced liquid droplets at micromolar concentrations of the N protein (Figure 1E). However, mixing the N protein with polyG, which forms G-quadruplexes, or polyAU, which forms stable base-pairing interactions, led to the formation of more gel-like condensates with asymmetric shapes (Figure 1E, Figure 1_figure supplement

1). We concluded that unstructured and flexible polyanions form complex coacervates with the positively charged N protein (*20*) and trigger formation of liquid condensates. The structure of the RNA affects N phase separation, with secondary structures of RNA triggering formation of gel-like droplets (*13, 23, 24*). Phase separation of N with RNA was also dependent on protein-RNA stoichiometry. Increasing the concentration of the N protein promoted phase separation, whereas high concentrations of RNA dissolved these N condensates (Figure 1E, Figure 1_figure supplement 2A,B), presumably by reducing the local concentration of the N protein on an RNA scaffold (*25, 26*).

We next sought to characterize how the N protein could structure viral RNA into condensates. Recently, the SARS-CoV-2 RNA genome was reverse transcribed and assembled into a DNA plasmid (*27*). We transcribed six 5 kb fragments and two 1 kb fragments from the genome in vitro (Figure 1F, Figure 1_figure supplement 1B) (*28*). In silico methods predict that these RNA fragments contain secondary structure elements (Figure 1F, Figure 1_figure supplement 1B) (*23, 29, 30*). Similar to synthetic RNA substrates that form a secondary structure (Figure 1D), the N protein formed asymmetric gel-like condensates with viral RNA fragments *in vitro* (Figure 1F, Figure 1_figure supplement 1B,C). The shapes of these asymmetric structures were consistent with remodeling of vRNPs into “beads on a string”, as observed by cryoEM (*6*). These asymmetric structures formed across a wide range of protein and RNA concentrations, did not change shape over time, and were not strongly affected by raising the temperature from 20 ºC to 37 ºC (Figure 1_figure supplement 3).

Cross-linking mass spectrometry identifies N protein interacting sites

We next investigated how interactions between N proteins mediate phase separation using cross-linking mass spectrometry (CLMS) (*31*), which detects protein-protein contacts by covalently capturing nearby residues with bifunctional reagents. Because CLMS cannot distinguish whether crosslinks are formed between two N protein homodimers or within a homodimer (*32*), we specifically investigated the crosslinks that are depleted or enhanced upon phase separation. We first crosslinked the soluble (not phase separated) N protein in 300 mM KAc without RNA using a bifunctional crosslinker bis(sulfosuccinimidyl) suberate (BS3) (Figure 2_figure supplement 1A). We detected that the N-terminal half of the protein, including the RNA-binding domain, makes diverse contacts throughout the entire protein (Figure 2_figure supplement 1A). There was also an abundance of contacts between the disordered regions immediately flanking the dimerization domain on either side (referred to as R1 and R2).

Next, we performed quantitative CLMS measurements (*33*) comparing the soluble N protein in 300 mM KAc with phase separated N protein in 100 mM KAc (Figure 2A). The soluble N protein was crosslinked with heavy (D12) BS3 whereas the phase separated protein was crosslinked with light (H12) BS3. As a result, the crosslinked precursor ions from high and low salt conditions were spaced by 12 Da, allowing us to compare interactions between the two conditions (Figure 2A). Interactions between specific regions that promote droplet formation are implicated by the ratio of the crosslinked precursor ion signal and its corresponding isotopic doublet (Figure 2A) (*33*). This experiment was repeated by reversing the labels, such that (H12) BS3 was used to crosslink the soluble N protein and (D12) BS3 was used for the phase separated N protein. Across two independent experiments, 29 unique crosslinks were enriched and 30 crosslinks were depleted upon phase separation (Figure 2B, Figure 2_figure supplement 1B). Remarkably, the analysis of the crosslink fold change (Table S1) revealed that nearly all of the enriched interactions are concentrated in R1 and R2, whereas depleted crosslinks more uniformly spanned the entire primary sequence (Figure 2C). These results indicate that R1 and R2 drive phase separation of the N protein (Figure 2D, Figure 2_figure supplement 1C).

Our mass spectrometry (MS) analysis also found phosphorylation sites on the N protein (Figure 2E). While some of these sites have been identified in previous studies (*9, 34, 35*), we also identified several novel sites (Table S2). Although one of the phosphorylation sites (S176) is involved in a crosslink, the phosphorylated and unphosphorylated peptides were both strongly depleted in droplets, suggesting that S176 phosphorylation

does not play a major role in phase separation (Tables S1 and S2, Figure 2_figure supplement 1B). Additionally, MS identified native proteins that co-purified with the N protein in 1 M salt (Table S3). Consistent with the recruitment of N to stress granules in cells (*8, 16, 36*), two of the most frequently identified proteins were stress granule proteins G3BP1 and G3BP2, with R2 of the N protein interacting with G3BP1 (Figure 2_figure supplement 1D).

The C-terminal disordered region and phosphorylation modulate phase separation

To test the predictions of the CLMS results, we tested how deletion of the PLD, SR, R1, or R2 regions affects phase separation of the N protein with polyC RNA (Figure 3A). Phase separation was only moderately reduced in the ΔSR, ΔPLD, and ΔR1 mutants (Figure 3B, Figure 3_figure supplement 1), suggesting that these regions are not essential for phase separation. However, deletion of the R2 sequence almost fully abolished the formation of condensates with polyC RNA (Figure 3B, Figure 3_figure supplement 1). None of the deletion mutants disrupted the asymmetric condensates formed by the N protein with viral RNA fragments (Figure 3_figure supplement 1E). These results show that R2 primarily drives phase separation of the N protein with unstructured RNA, consistent with the role of this sequence in oligomerization of N (*37*).

Recent studies proposed that phosphorylation state on the N protein regulates genome packaging in SARS-CoV and SARS-CoV-2 (*36, 38-41*), and the mass spectrometry analysis of our N protein purified from mammalian cells identified several phosphorylation sites (Figure 2E, Table S2). Treating the full-length N protein with casein kinase 2 did not affect its phase separation with polyC (Figure 3C,D) or its migration on a denaturing gel (Figure 3_figure supplement 2A), suggesting that the N protein expressed in human cells is already phosphorylated. However, treating N with λ protein phosphatase produced a noticeable shift in a denaturing gel (Figure 3_figure supplement 2A). This dephosphorylated N phase separated similarly to the mammalian expressed protein (Figure 3C,D, Table S4), suggesting that phosphorylation does not affect the concentration of N required to form higher-order structures. To further characterize whether phosphorylation affects phase separation, we dephosphorylated the ΔSR, ΔR1 and ΔR2 mutants. Dephosphorylated ΔSR or ΔR1 did not produce noticeable changes in molecular weight (Figure 3_figure supplement 2B,C) and had only a moderate increase in phase separation (Figure 3_figure supplement 2). However, dephosphorylated ΔR2 had a reduced molecular weight (Figure 3_figure supplement 2B). Although regular ΔR2 was unable to phase separate with polyC RNA, dephosphorylated ΔR2 exhibited robust phase separation with polyC RNA, similar to the wild-type N protein (Figure 3E,F). Together, these results suggest the R2 region promotes phase separation and in the absence of R2, phosphorylation of the N protein inhibits phase separation.

Targeting phase separation of the N protein with small molecules

Phase separation may also provide a macroscopic readout to study the N protein and RNA interactions (*42*) and suggest novel strategies to disrupt packaging of the viral genome. We attempted to identify small molecules that affect N protein phase separation *in vitro*. We triggered the formation of liquid condensates by mixing 7.8 μM Cy5-labeled N protein and 50 ng/μL polyC RNA in 150 mM NaCl. Lipoic acid, a potential drug candidate that increases the liquidity of stress granules (*42*), did not affect phase separation of the N protein (Figure 4_figure supplement 1). We also utilized a high throughput fluorescence microscopy combined with computational image analysis for direct observation of in vitro phase separation of the N protein and determination of mean the droplet size. By screening a 1200-molecule FDA approved library, we identified two molecules that reduced the number of droplets (Class I), three molecules that increased the number and size of the droplets (Class II), and two molecules that changed droplet morphology (Class III).

Class I molecules, nelfinavir mesylate and LDK378, produced fewer but larger droplets (Figure 4A,B, Figure 4_figure supplements 1 and 2), and both have been identified by other SARS-CoV-2 screens. Nelfinavir mesylate is an HIV protease inhibitor that was identified in multiple *in silico* screens as binding to the SARS-CoV-2 main protease (*43-46*) and was shown to block viral production without affecting viral entry in SARS-

CoV (*47*). Nelfinavir mesylate reduced droplet number in our screen with a half maximal effective concentration (EC_{50}) of 4.3 ± 1.3 μM (Table S4), suggesting that it could act by disrupting virion assembly. LDK378, a tyrosine kinase inhibitor, was shown to reduce the N protein signal in infected cells (*48*). In our screen, LDK378 produced a similar phenotype as nelfinavir but with an EC_{50} greater than 40 μM. Of the three Class II molecules that increase the size and number of N-RNA condensates (Figure 4A,B, Figure 4_figure supplements 1 and 2), tolcapone, a methyltransferase inhibitor, had been implicated in reducing viral production of an alpha coronavirus (*49*). The Class III molecules, nilotinib (a kinase inhibitor) and L-thyroxine (a thyroid hormone receptor agonist), did not quantitatively change droplet count or area (EC_{50} > 40 μM) but had morphological effects on the droplets at 40 μM (Figure 4A,B, Figure 4_figure supplements 1 and 2). Nilotinib could be a promising drug target, as it was shown to bind the S protein (*50*), reduce viral titer in SARS-CoV infected cells (*51*), and inhibit SARS-CoV-2 VeroE6 cell infection with an EC_{50} of 60 nM (*52*). Similarly, L-thyroxine has been suggested to reduce SARS-CoV-2 uptake through its ability to bind to integrin (*53*).

We next tested the drugs we identified in our initial screen against the condensates formed by N and viral RNA in vitro. Nilotinib addition resulted in formation of thread-like filaments of N and viral RNA (Figure 4C), whereas other drugs did not affect the condensates containing longer RNA (Figure 4_figure supplement 1). To test whether nilotinib affects phase separation of the N protein in vivo, we transiently expressed N-GFP in VeroE6 cells. Consistent with our in vitro results, cells expressing N-GFP produced bright puncta, suggesting that the N protein forms higher order structures in mammalian cells (Figure 4D, Figure 4_figure supplement 3A). Incubating these cells with at least 3 μM nilotinib led to a significant decrease in the number (Figure 4E) and the total area of the puncta per cell (Figure 4_figure supplement 3B) but did not significantly affect the GFP signal per cell (Figure 4_figure supplement 3C). Therefore, nilotinib disrupts the association of the N protein into higher order structures in vivo and could serve as a potential drug candidate against packaging of SARS-CoV-2 virus in host cells.

Discussion

Recent studies showed that SARS-CoV-2 N protein expressed in bacteria forms biomolecular condensates with RNA in vitro (*17, 28, 36, 38, 40, 54, 55*). In this study, we expressed the N protein using a mammalian expression system, which more closely resembles the N protein produced in infected human cells (*9*). Consistent with Carlson et al. and Chen et al., we observed that the N protein requires unstructured RNA fragments to form liquid condensates in physiological salt (*38, 54*), suggesting that the positively charged protein and polyanion RNA form complex coacervates. The N protein is also capable of forming a biomolecular condensate in the presence of 1-5 kb viral RNA; however, these structures appear closer to gel-like filaments (*38*) and not as temperature-sensitive, liquid-like droplets observed by Iserman et al. using bacterially expressed N (*28*). However, in agreement with Iserman et al., we found that the structure of the RNA itself affects N condensation; base-paired or G-quadruplex-containing RNA produce irregularly shaped, gel-like condensates, similar to viral RNA (*28*). RNA may drive the formation of abnormal droplet shapes by forming networks of intermolecular base pairs, as previously reported (*13, 23, 24, 56*).

We also investigated which domain of the N protein drives phase separation by CLMS experiments and testing a series of deletion mutants. PLD and SR motifs are common features of phase separating proteins (*20, 21*), and the linker sequence outside the SR motif (amino acids 210-247) is essential for phase separation of N expressed in bacteria (*55*). The SR motif was required for N/N interactions in SARS-CoV and for forming puncta in SARS-CoV infected cells (*41*). In addition, the SR motif is highly phosphorylated in human cells (*57*), and this promotes interactions with host proteins, nuclear targeting, and transcription of the viral genome, and either enhances or inhibits oligomerization in SARS-CoV (*16, 39, 57, 58*). In SARS-CoV-2, SR phosphorylation is reported to make N condensates more liquid (*38*). In contrast to these reports, we did not see evidence of the SR motif driving the phase separation of SARS-CoV-2 N expressed in human cells. Instead, we found that the R2 motif, which forms a transient alpha helix (*17*), is essential for phase separation with nonviral RNA.

Phosphorylation of the ΔR2 mutant fully disrupts phase separation, suggesting that phosphorylation interferes with phase separation, and this is countered by the R2 motif. Because R2 promotes oligomerization and phase separation of the N protein (*37*), targeting of this motif may be a viable approach to disrupt packaging of SARS-CoV-2 in infected cells. Although we found that phosphorylated and dephosphorylated full-length N protein phase separate at a similar protein concentration, our results are not inconsistent with phosphorylation of the N protein to reduce the amount of RNA required to form a condensate (*36*), as well as to affect the transcription of the genomic RNA (*58*).

Although it remains to be demonstrated whether phase separation is essential for propagation of SARS-CoV-2, the filament structures we observe recapitulate vRNPs observed in infected cells (*4-6*). These “beads on a string” structures are seen both in vitro and in vivo, suggesting that the assembly of a virus inside host cells may involve the formation of gel-like condensates of the N protein (Figure 5). In the measles virus, nucleoprotein (N) contains an RNA binding domain while phosphoprotein (P) contains an oligomerization domain. N and P form condensates when mixed together (*59*). In coronaviruses, the N protein possesses both an RNA-binding and oligomerization domain in a single peptide chain. Vesicular stomatitis virus and rabies virus create membraneless compartments as sites of RNA synthesis in the cytoplasm of the host cell, and these compartments have characteristics of liquid condensates (*60, 61*). These results indicate that phase separation may serve as a general mechanism used by RNA viruses to propagate in host cells. Future studies are required to test whether mutations or deletion of the R2 motif of the viral genome alter localization of the N protein and reduce the efficiency of SARS-CoV-2 propagation in infected cells.

We tested phase separation of the SARS-CoV-2 N protein to gain mechanistic insight into how existing drugs can modulate a specific viral process, rather than broad cell pathology. The most specific drug we identified is nilotinib, which affected phase separation with nonspecific polyC RNA and specific viral RNA in vitro, as well as decreased the number of N puncta in cells. Previous studies showed that nilotinib inhibits the proliferation of both SARS-CoV and SARS-CoV-2 (*50-52*), but the mechanism of this inhibition was unknown. Our results indicate that nilotinib affects the formation of the N/RNA condensates, which could potentially impair viral replication. Nilotinib has been shown to be readily absorbed in patients, with a minimal steady state plasma concentration of about 1000 ng/mL (~2 μM) (*62*). Several COVID-19 clinical trials of imatinib, the parent drug of nilotinib, are underway (*63*). Although nilotinib has not been a part of a COVID-19 clinical trial, by using phase separation as a proxy for vRNP formation, our results indicate that nilotinib could represent a viable approach in future studies of COVID-19.

Figures

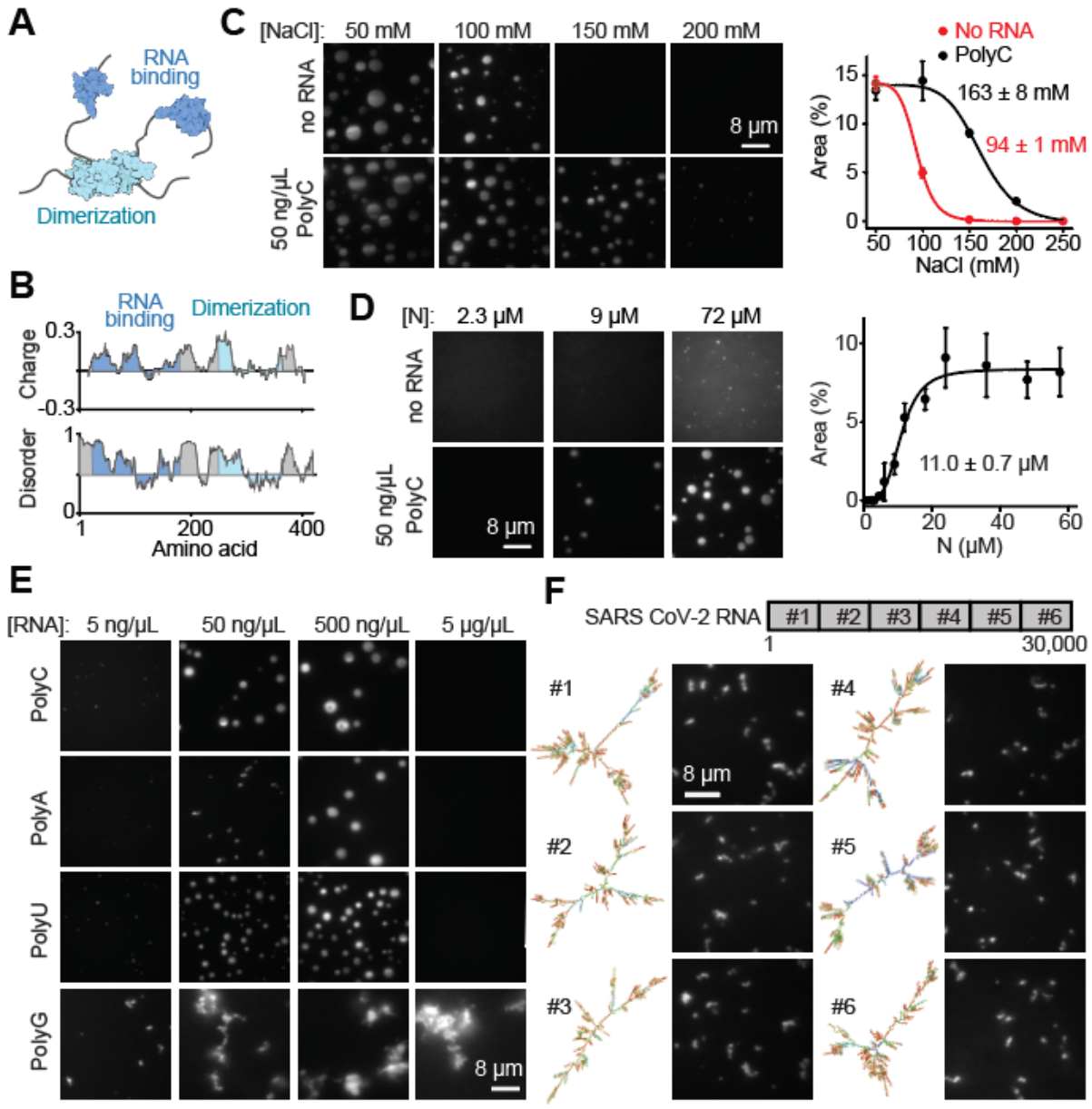

Figure 1. The SARS-CoV-2 N protein phase separates with RNA *in vitro*. (A) Schematic highlights the structured and disordered regions of the N protein dimer. **(B)** Sliding window plot of charge distribution and disorder prediction for the N protein. **(C)** (Left) Spherical droplets formed by 24 µM N protein in the presence or absence of 50 ng/µL polyC RNA dissolve by increasing NaCl concentration. (Right) The changes in the total area covered by N droplets with increasing salt concentration (mean ± s.d., n = 20 with two technical replicates). Solid curves represent a fit to a dose response equation to determine IC_{50} (±s.e.). **(D)** (Left) Images of the N protein in the presence and absence of polyC RNA in 150 mM NaCl. (Right) The percentage of surface area covered by N droplets in 150 ng/µL polyC RNA (mean ± s.d., n = 20 with two technical replicates). Solid curve represents a fit to a Hill equation to determine k_{sat} (± s.e.). **(E)** The N protein forms spherical droplets with unstructured RNA (polyA, polyC, and polyU) but forms asymmetric condensates with structured RNA (polyG). The N protein concentration was set to 18.5 µM. **(F)** (Top) SARS-Cov-2 genomic RNA was divided into 6 sections. (Bottom) Structure prediction of each section (left) and the formation of asymmetric N condensates in the presence of 18 nM RNA (right). The N protein concentration was set to 18.5 µM.

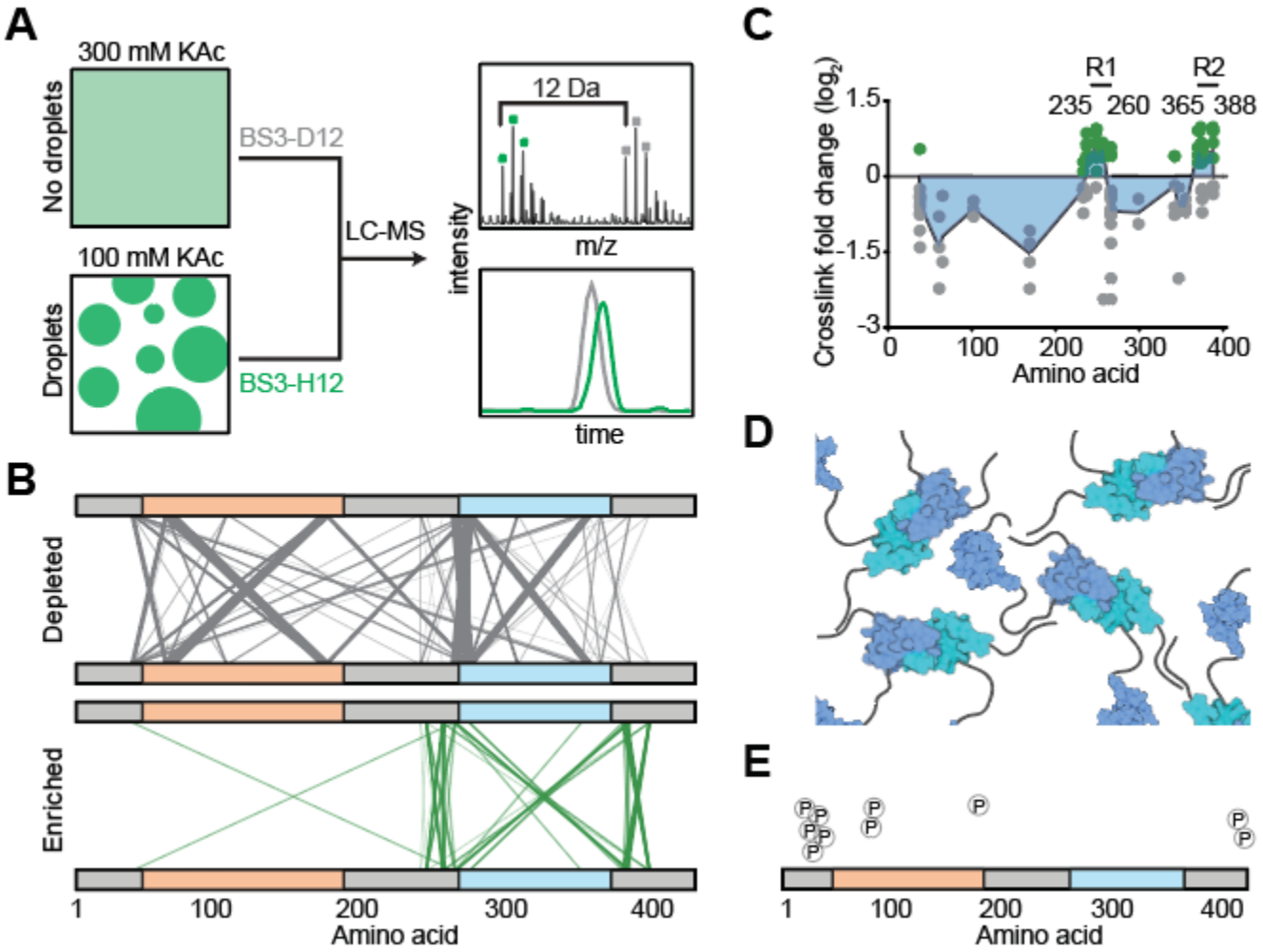

Figure 2. CLMS reveals interdomain interactions of the N protein. (A) Schematic of the CLMS experiment. (Left) A high salt (300 mM KAc) buffer disrupts N droplets, whereas a low salt (100 mM KAc) buffer promotes droplet formation. (Right, top) Example of an individual crosslinked peptide in quantitative CLMS analysis. Precursor ions from the high salt (gray) and low salt (green) BS3 crosslinking conditions show the 12 Da shift between light (H12) and heavy (D12) crosslinkers. (Right, bottom) Ion chromatograms from the first three isotopes of each doublet were extracted and expressed as the ratio of peak areas. **(B)** The plot of crosslinks depleted and enriched in the droplet condition. Width and transparency of the lines scale with the number of times the crosslink was detected across 3 independent experiments. **(C)** Fold changes of crosslink abundance upon droplet formation of N. As crosslinks contain two positions, fold change information is plotted at both positions. Only crosslinks with p-values less than 0.05 are included. Green and grey dots represent crosslinks enriched and depleted in the droplet condition, respectively. The blue area represents a plot of median crosslink fold change. **(D)** Model for how multiple N dimers could phase separate via their disordered regions. **(E)** Phosphorylation sites detected by the CLMS experiment in 300 mM KAc.

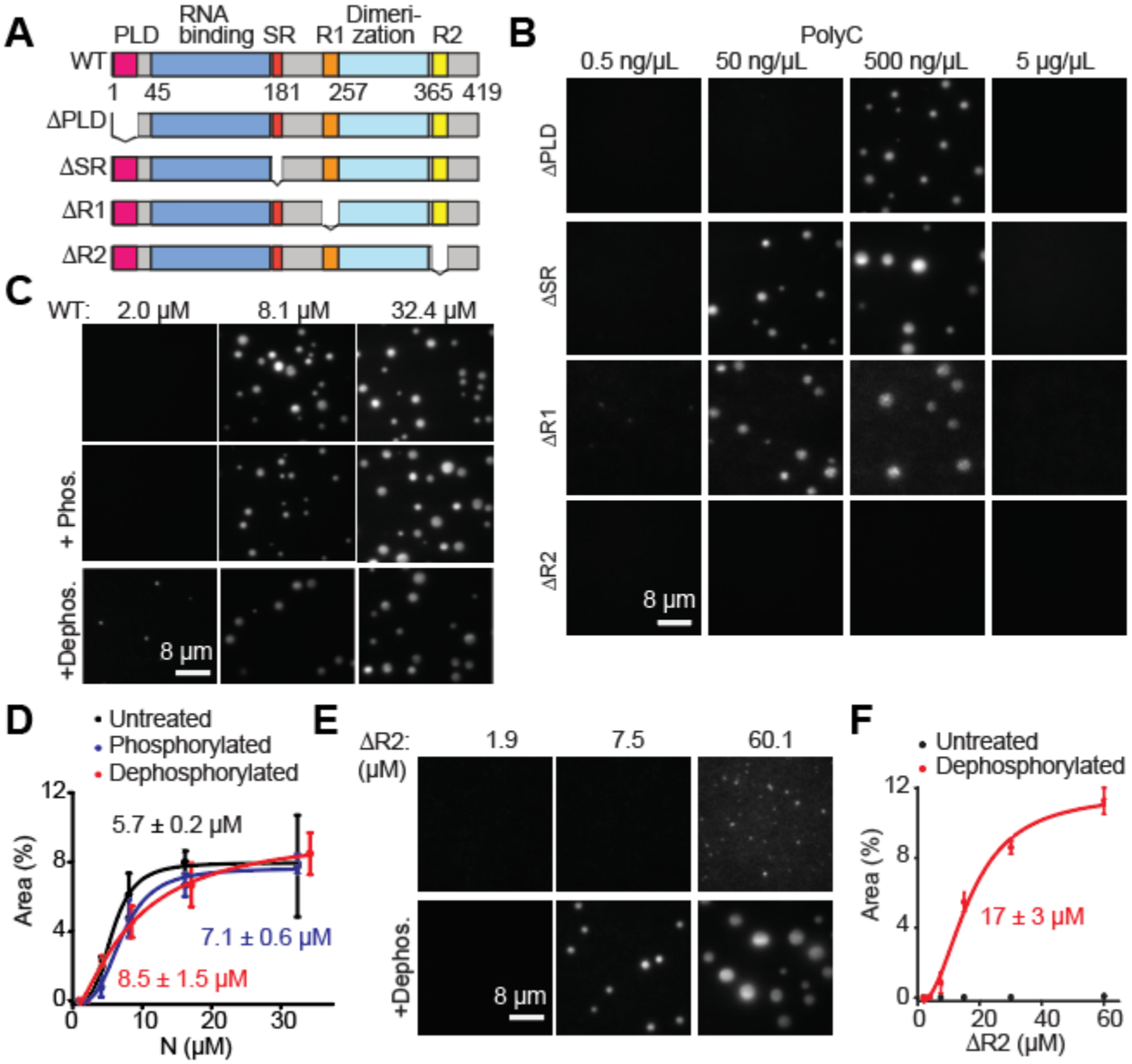

Figure 3. The effect of domain truncations and phosphorylation on phase separation of the N protein. **(A)** Truncation mutants of the N protein. **(B)** While ΔPLD, ΔSR and ΔR1 phase separate, ΔR2 does not phase separate when mixed with polyC RNA. **(C)** Images of condensates formed by untreated, phosphorylated, and dephosphorylated N protein in 50 ng/μL polyC RNA. **(D)** Quantification of N droplets formed under different phosphorylation conditions (mean ± s.d., n = 20 with two technical replicates). Fits to a Hill equation (solid curves) reveal k_{sat} (± s.e.). **(E)** Images of condensates formed by untreated and dephosphorylated ΔR2 in 50 ng/μL polyC RNA. **(F)** Quantification of droplet area as a function of ΔR2 concentration (mean ± s.d., n = 20 with two technical replicates). Fit to a Hill equation (solid curves) reveals k_{sat} (± s.e.).

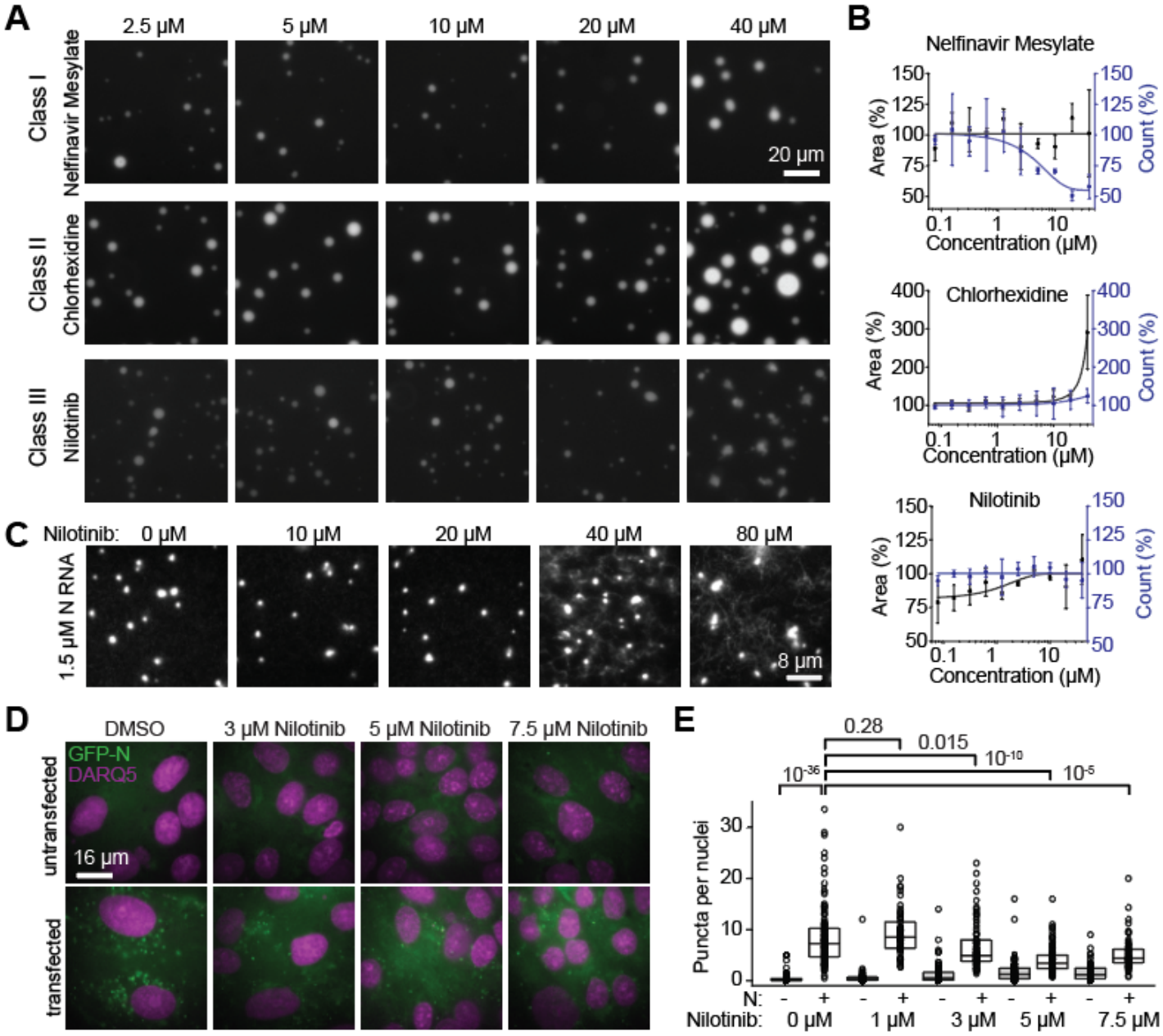

Figure 4. Identification of small molecules that alter phase separation of N in vitro and in vivo. (A) Examples of Class I, II, and III drugs. The N protein and polyC RNA concentrations were set to 7.8 μM and 50 ng/uL, respectively. **(B)** Changes in the total surface area and the number of droplets formed by N and polyC RNA upon drug treatment (mean ± s.d., n = 8 with two technical replicates). Solid curves represent a fit to a dose response equation (see Methods). **(C)** 57.6 μM N protein and N gene RNA mixtures as a function of nilotinib. **(D)** VeroE6 cells transfected with N-GFP form puncta. Nilotinib treatment reduces the number of puncta per cell. **(E)** The number of puncta per nucleus under different nilotinib treatments. The center and edges of boxes represent median with first and third quartile (four technical replicates for the 0 μM and 5 μM conditions; two technical replicates for all other conditions; from left to right, n = 118, 122, 62, 57, 64, 78, 88, 95, 67, 64). p values are calculated from a two-tailed t-test.

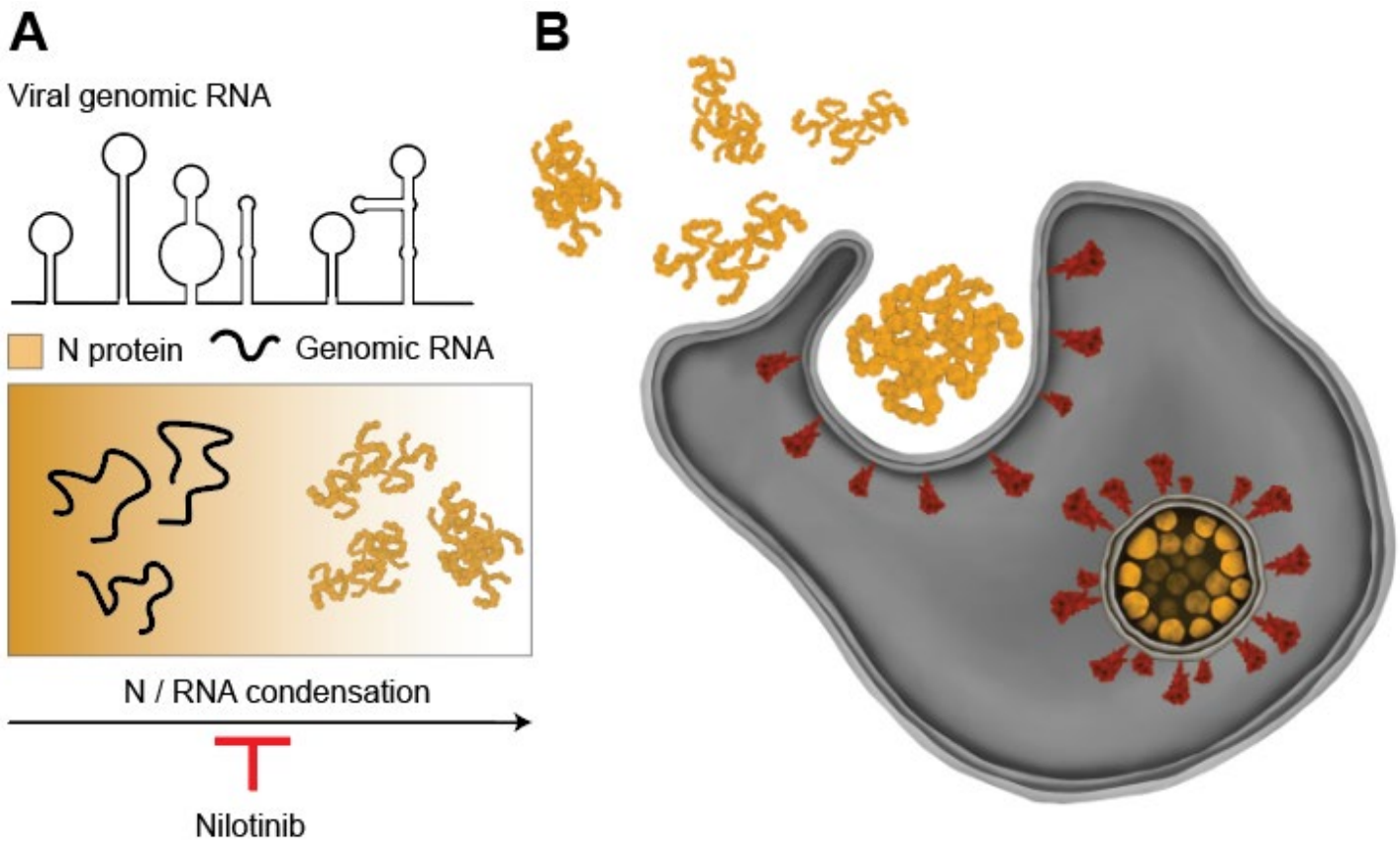

Figure 5. Model for remodeling of viral RNA genome by the SARS CoV-2 N protein. **(A)** The N protein packages viral genomic RNA into asymmetric gel-like structures through phase separation. Nilotinib disfavors formation of higher order structures by the N protein, and may disrupt packaging of the RNA genome into the viral envelope. **(B)** vRNPs interact with the viral membrane proteins on the surface of the ER-Golgi intermediate compartment and form the enveloped virus.

Supplemental tables

Fold change (log_2)	p-value	Xlink position 1	Xlink position 2
-2.441	1.05E-04	257	266
-2.226	5.88E-06	169	61
-2.026	8.84E-08	266	347
-1.697	1.23E-09	169	65
-1.409	1.21E-05	38	61
-1.324	3.20E-05	169	266
-1.079	6.84E-04	169	38
-0.941	3.61E-12	266	299
-0.796	8.00E-07	102	61
-0.765	2.94E-06	266	266
-0.762	3.87E-04	342	38
-0.730	5.18E-07	233	266
-0.729	8.10E-05	266	38
-0.715	0.003417897	355	375
-0.648	0.008659112	102	266
-0.607	0.001310878	342	355
-0.594	2.84E-05	347	38
-0.488	0.030217643	102	38
-0.450	8.76E-04	266	375
-0.448	1.53E-13	299	355
-0.415	3.30E-06	233	38
-0.382	0.048322087	38	65
-0.380	2.21E-04	237	266
-0.351	4.41E-04	266	388
-0.277	0.0366727	266	374
-0.264	2.72E-04	233	233
-0.249	0.004338941	375	38
-0.222	0.015033594	233	248
-0.220	0.034461117	347	388
-0.185	0.041924355	342	388
0.103	0.038548375	233	249
0.271	0.002666995	370	375
0.291	6.17E-06	233	237
0.347	6.53E-05	373	375
0.361	3.17E-06	237	388
0.378	5.90E-09	248	256
0.396	5.40E-05	256	342
0.398	1.41E-05	266	373
0.402	1.40E-06	257	342
0.402	9.55E-12	372	375
0.403	0.00139739	237	249
0.408	1.05E-07	248	257
0.418	0.002486181	370	374
0.428	3.83E-07	372	374

0.455	1.34E-06	266	372
0.522	1.41E-07	256	375
0.540	5.63E-06	256	38
0.578	3.49E-05	266	370
0.578	2.94E-05	257	374
0.580	4.27E-10	257	375
0.610	1.18E-07	237	257
0.641	1.58E-07	237	256
0.668	4.45E-04	248	388
0.670	4.93E-06	256	374
0.857	3.80E-12	237	248
0.902	1.31E-06	372	388
0.909	0.018777529	370	388
0.941	2.54E-05	248	249
0.976	2.58E-06	373	388

Table S1. Crosslink fold changes with p-values less than 0.05 upon phase separation. Xlink position indicates the amino acid number of the N protein sequence. 169 in Xlink position 1 represents the K169-K65 crosslink, which is phosphorylated at S176.

A

```
1   MSDNGPQNQR NAPRITFGGP SDSTGSNQNG ERSGARSKQR RPQGLPNNTA SWFTALTQHG KEDLKFPRGQ
71  GVPINTNSSP DDQIGYYRRA TRRIRGGDGK MKDLSPRWYF YYLGTGPEAG LPYGANKDGI IWVATEGALN
141 TPKDHIGTRN PANNAAIVLQ LPQGTTLPKG FYAEGSRGGS QASSRSSSRS RNSSRNSTPG SSRGTSPARM
211 AGNGGDAALA LLLLDRLNQL ESKMSGKGQQ QQGQTVTKKS AAEASKKPRQ KRTATKAYNV TQAFGRRGPE
281 QTQGNFGDQE LIRQGTDYKH WPQIAQFAPS ASAFFGMSRI GMEVTPSGTW LTYTGAIKLD DKDPNFKDQV
351 ILLNKHIDAY KTFPPTEPKK DKKKKADETQ ALPQRQKKQQ TVTLLPAADL DDFSKQLQQS MSSADSTQAL
421 EGGGGWSHPQ FEKGGGSGGG SGGGSWSHPQ FEK
```

B

Phosphosite	Found in	Site Localization (this study)
S21	This study	Unambiguous
S23	This study,	Unambiguous
T24	This study,	Unambiguous
S26	This study, (*34*)	Ambiguous (S26|S33)
S33	This study	Ambiguous (S26|S33)
T76	(*9, 34*)	
S78	This study, (*9*)	Ambiguous (S78|S79)
S79	This study, (*9, 34*)	Unambiguous
S105	(*9, 34*)	
T141	(*9*)	
T166	(*9*)	
S176	This study, (*9, 34*)	Unambiguous
S180	(*9, 34*)	
S183	(*9, 34*)	
S184	(*9, 34*)	
S194	(*9, 34*)	
S197	(*34*)	
T198	(*9, 34*)	
S201	(*9, 34*)	
S202	(*9, 34*)	
T205	(*9, 34*)	
S206	(*9, 34*)	
T391	(*9*)	
T393	(*35*)	
S412	This study	Ambiguous (S412|S413)
S413	This study	Ambiguous (S412|S413)
S441	This study	Unambiguous

Table S2. A compilation of the SARS-CoV2 N protein phosphosites identified by mass spectrometry. A. The coverage map and phosphorylation sites for the N protein detected in proteomics experiments. Text in red indicates peptide regions that were not detected; green highlight indicates an unambiguous phosphorylation site; yellow highlight indicates an ambiguous phosphorylation site. **B.** Each site identified by our study is characterized as being unambiguously or ambiguously localized based on manual inspection of the product ion series. See Methods for the full protein sequence and a link to supporting evidence. S176 is a phosphorylation site identified on a crosslinked peptide.

A

Gene	Unique Peptides	Percent Coverage	SAF	Species	UniProt Accession	Protein Name
N	57	70.0	104.6	SARS2	P0DTC9	Nucleoprotein
HSPA1A	54	59.3	27.5	HUMAN	P0DMV8	Heat shock 70 kDa protein 1A
G3BP2	38	42.5	19.5	HUMAN	Q9UN86	Ras GTPase-activating protein-binding protein 2
G3BP1	37	60.5	17.8	HUMAN	Q13283	Ras GTPase-activating protein-binding protein 1
HSPA8	27	37.8	9.8	HUMAN	P11142	Heat shock cognate 71 kDa protein
RPS12	2	17.4	8.3	HUMAN	P25398	40S ribosomal protein S12
NA	4	17.3	7.4	PIG*	P00761	Trypsin
TUBB	20	56.3	7.0	HUMAN	P07437	Tubulin beta chain
TUBA1B	19	54.3	6.7	HUMAN	P68363	Tubulin alpha-1B chain
TUBB4B	18	50.6	6.5	HUMAN	P68371	Tubulin beta-4B chain
KRT9	14	52.6	6.1	HUMAN	P35527	Keratin, type I cytoskeletal 9
A2M	27	20.7	5.8	BOVIN*	Q7SIH1	Alpha-2-macroglobulin
HSPA5	17	29.5	5.2	HUMAN	P11021	Endoplasmic reticulum chaperone BiP
NPM1	6	37.1	5.1	HUMAN	P06748	Nucleophosmin

B

Protein 1	Xlink position 1	Score	Protein 2	Xlink position 2	Score
G3BP2_HUMAN	50	67.2	G3BP2_HUMAN	281	31.4
Nucleoprotein	375	12.2	G3BP1_HUMAN	453	5.6
G3BP2_HUMAN	50	41.5	Nucleoprotein	266	28.5
G3BP2_HUMAN	50	53.8	Nucleoprotein	102	20.9
G3BP2_HUMAN	370	68.6	G3BP2_HUMAN	407	13.2
G3BP2_HUMAN	365	73.7	G3BP2_HUMAN	407	0.8
G3BP2_HUMAN	50	56.5	Nucleoprotein	38	8.6
Nucleoprotein	388	48.1	G3BP1_HUMAN	453	5.1
Nucleoprotein	266	41.4	G3BP1_HUMAN	393	8.6
Nucleoprotein	374	35.3	G3BP1_HUMAN	453	10.2

Table S3. Proteins identified by mass spectrometry in the high salt sample of N protein from HEK293 cells. A. The proteins were ranked by spectral abundance factor (SAF). These protein sequences were included in the CLMS search. Asterisks indicate exogenous proteins. **B.** Crosslinks between N and the G3BP1 and G3BP2 stress granule proteins identified in the qualitative CLMS experiment. Values are frequency of detection (two conditions with two replicates each). Xlink position indicates the amino acid number of the given protein sequence. Score indicates confidence in the identification of the peptide.

A

Condition	$Area_{max}$ (%)	k_{sat} (μM)	*n*	R^2	Figure
WT N + 10% PEG	15 ± 2	4.7 ± 1.2	1.3 ± 0.4	0.98	1C
WT N + PolyC	8.4 ± 0.4	11.0 ± 0.7	3.8 ± 0.9	0.97	1C
WT N, pre-phosphorylation	7.96 ± 0.18	5.7 ± 0.2	3.5 ± 0.4	0.996	3C,D
WT N, post-phosphorylation	7.6 ± 0.4	7.1 ± 0.6	3.2 ± 0.9	0.98	3C,D
WT N, post-dephosphorylation	9.7 ± 1.2	8.5 ± 1.5	1.4 ± 0.4	0.99	3C,D
ΔSR N, pre-dephosphorylation	4.6 ± 0.3	6.2 ± 0.6	2.9 ± 0.7	0.98	S3_2A,B
ΔSR N, post-dephosphorylation	12 ± 4	13 ± 8	1.4 ± 0.5	0.96	S3_2A,B
ΔR1 N, pre-dephosphorylation	7.2 ± 0.6	12 ± 2	2.2 ± 0.6	0.97	S3_2C,D
ΔR1 N, post-dephosphorylation	19 ± 12	44 ± 73	0.8 ± 0.3	0.94	S3_2C,D
ΔR2 N, pre-dephosphorylation	N/A	N/A	N/A	N/A	3E,F
ΔR2 N, post-dephosphorylation	11.7 ± 1.3	17 ± 3	2.2 ± 0.7	0.98	3E,F

B

Condition	IC_{50} / EC_{50}	*n*	R^2	Figure
N, droplet area	93.74 ± 0.08 mM	9.74 ± 0.12	1	1D
N + PolyC, droplet area	162 ± 8 mM	8 ± 3	0.98	1D
N + PolyC with nelfinavir mesylate, droplet number	4.3 ± 1.3 μM	1.9 ± 1.0	0.90	4A,B
N + PolyC with tolcapone, droplet area	28 ± 12 μM	1.6 ± 0.4	0.98	S4_2

Table S4. Quantitative analysis of N droplets. A. The parameters of fitting of N droplets to a Hill curve in the presence and absence of a crowding agent and polyC RNA. **B.** The parameters of fitting to a dose response curve in the presence and absence of a crowding agent and polyC RNA.

Supplemental figures

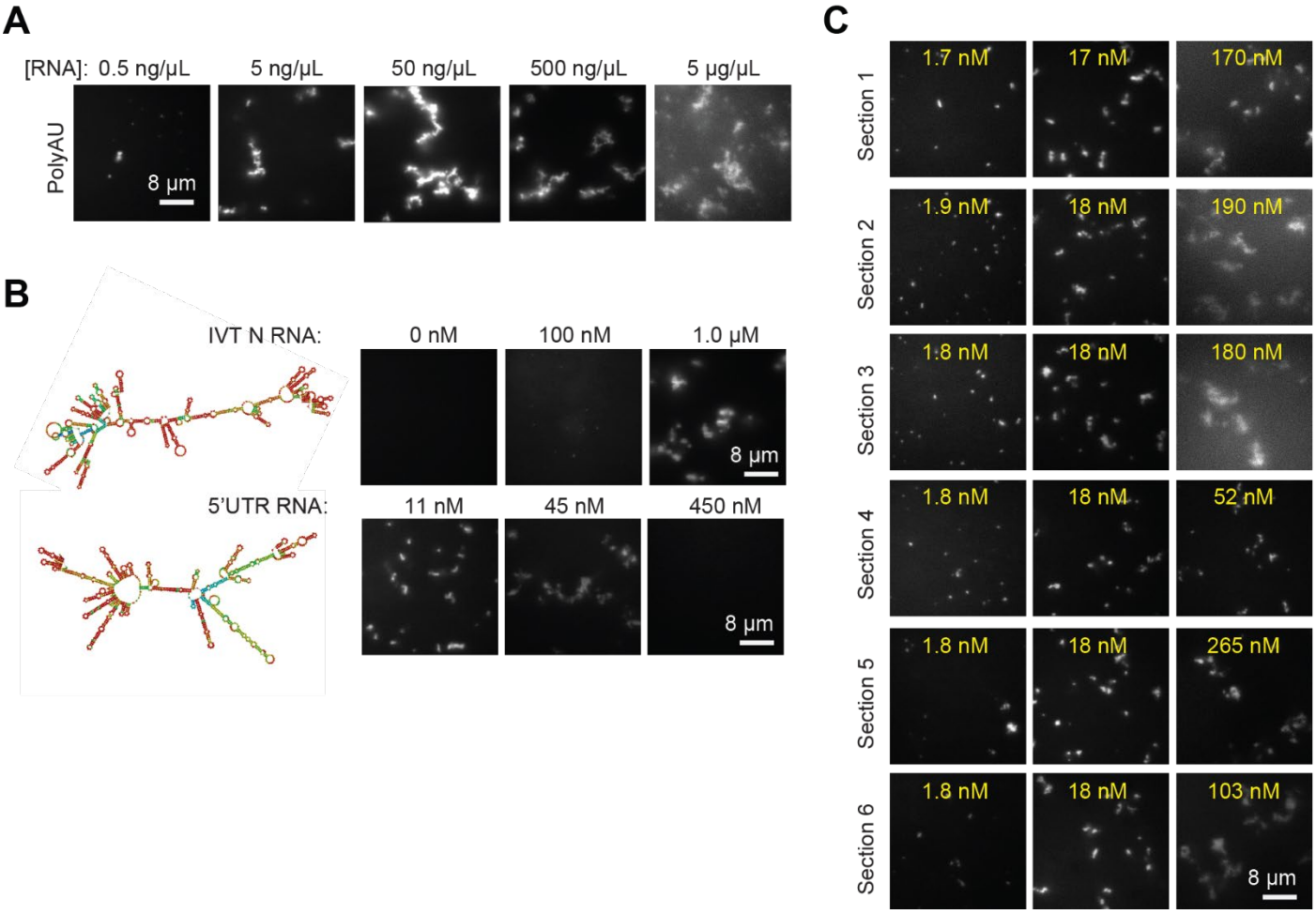

Figure 1_figure supplement 1. N protein mixed with viral RNA produces filaments resembling "beads on a string." (A) Example pictures show that the N protein forms asymmetric condensates with polyAU RNA, which forms basepairing interactions. The N protein concentration was set to 18.5 μM. **(B)** Structure prediction of viral RNA (left) and the formation of asymmetric N condensates under different RNA concentrations (right). The N protein concentration was set to 18.5 μM (IVT: in vitro transcribed; UTR: untranslated region). **(C)** The formation of asymmetric N condensates under different viral RNA concentrations. The N protein concentration was set to 18.5 μM.

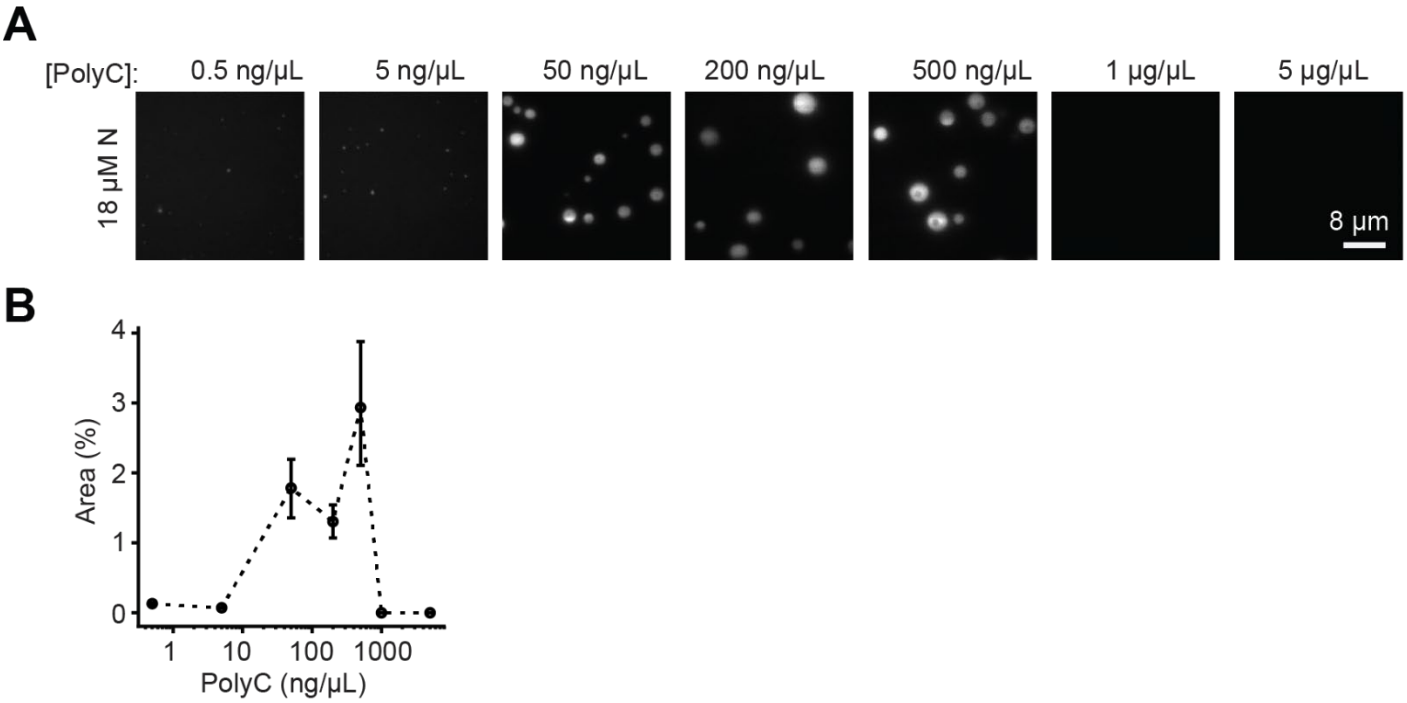

Figure 1_figure supplement 2. The stoichiometry of the N protein and RNA affect phase separation. (A) Example pictures show that the N protein forms spherical condensates with polyC RNA under different RNA concentrations. The N protein concentration was set to 18.5 μM. **(B)** The percentage of the surface area covered by N-RNA droplets under different RNA concentrations (mean ± s.d.; n = 20, two technical replicates).

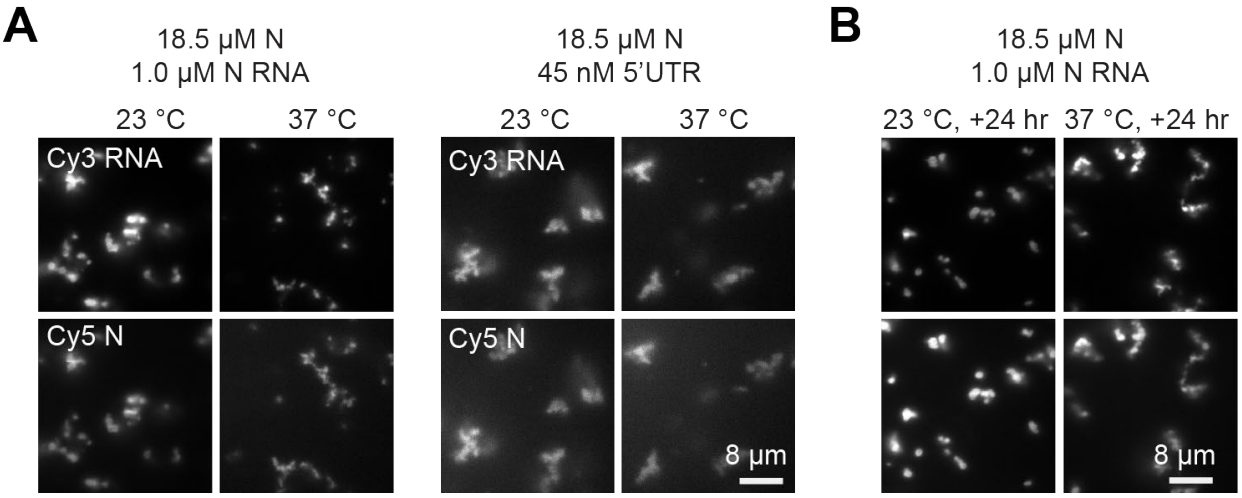

Figure 1_figure supplement 3. Condensates formed by the N protein and viral RNA do not change shape over time or by an increase of temperature to 37 ℃. (A) Condensates formed by the N protein with in vitro transcribed N RNA or the 5'UTR of the viral RNA are not sensitive to an increase of temperature to 37 ℃. **(B)** Condensates formed by the N protein with 5'UTR of viral RNA do not change shape over 24 hours.

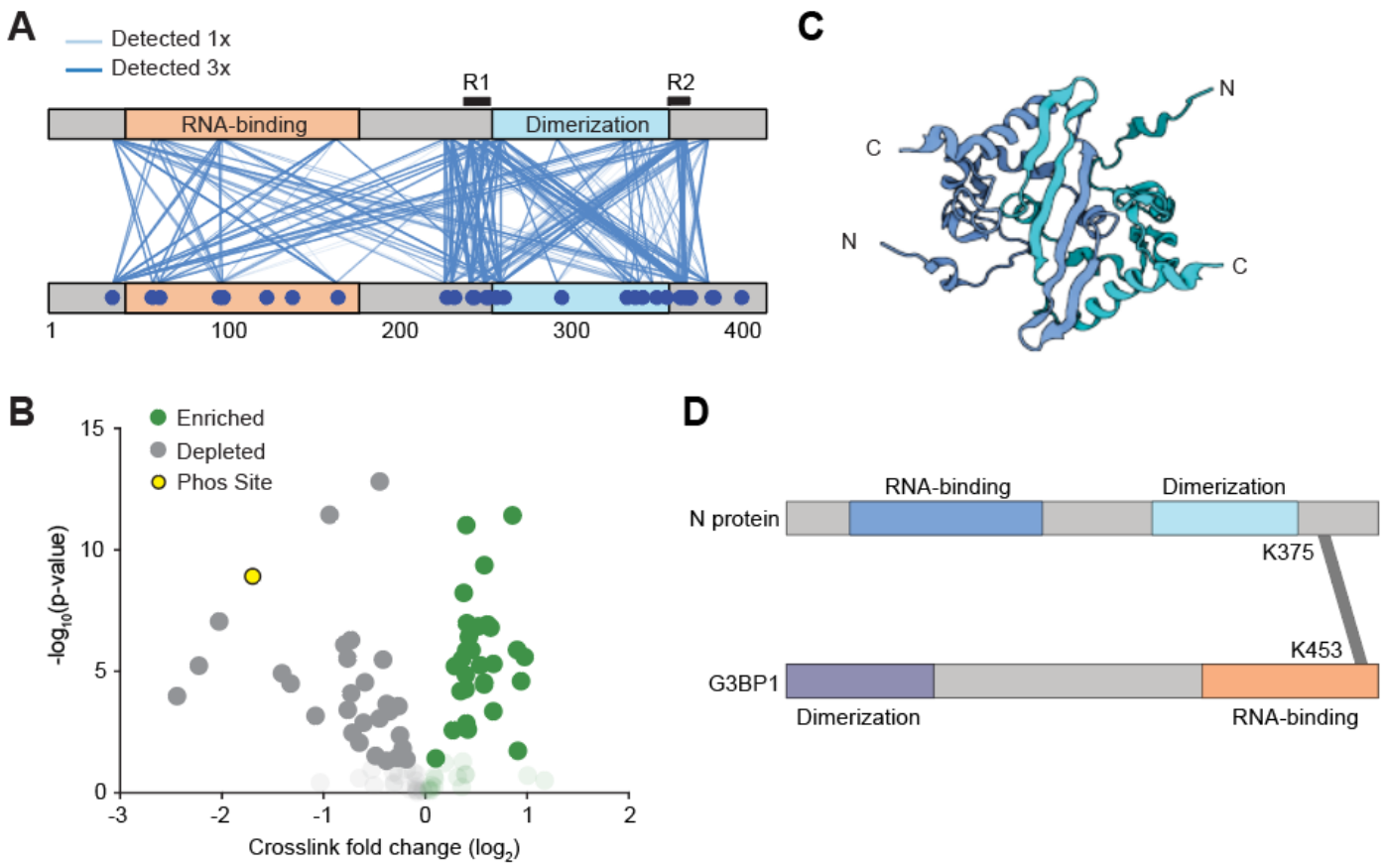

Figure 2_figure supplement 1. Interactions between N proteins and between N and G3BP1. (A) In an initial qualitative experiment, pairwise interactions were detected by CLMS at 300 mM salt. Lines depict a unique crosslink detected. The regions of N protein interactions flank the dimerization domain. **(B)** Volcano plot of the quantitative CLMS data comparing the droplet and no droplet condition. Opaque data points have a p-value below 0.05, transparent data points have p-value greater than 0.05. Green represents unique crosslinks that are enriched in the droplet condition. The yellow marker represents the K169-K65 crosslink, which also contains a phosphorylation site. **(C)** The structure of the dimerization domain (pdb# 2JW8 (*64*)). Because the N and C termini of the protomers are positioned away from each other, R1 and R2 within the same dimer are unlikely to interact with each other. **(D)** Mass spectrometry identified that the RNA binding domain of the stress granule protein G3BP1 interacts with the R2 region of N.

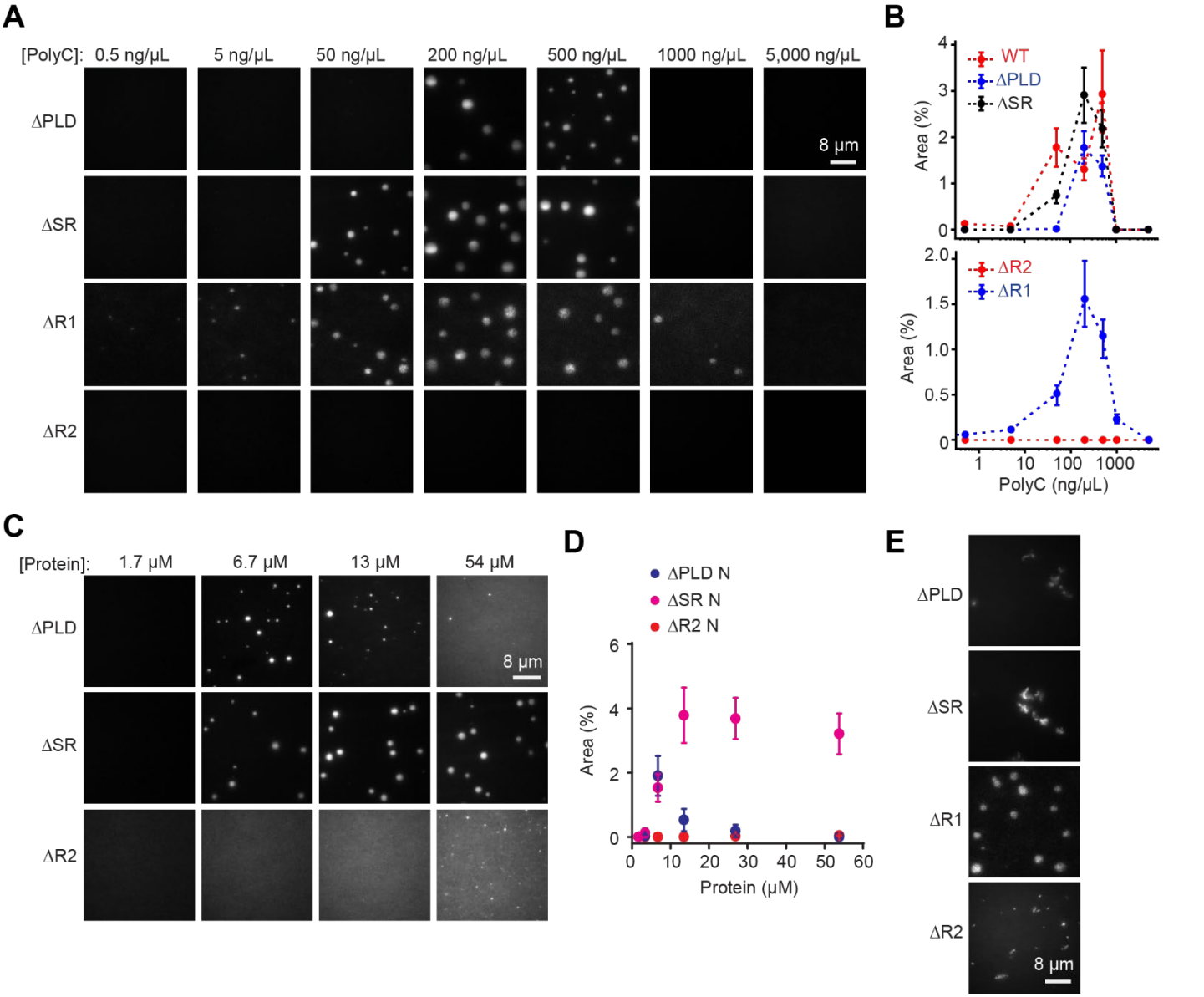

Figure 3_figure supplement 1. The in vitro phase of the truncation mutants of the N protein. (A) Example pictures show that the N truncation mutants, except ΔR2, form spherical condensates with polyC RNA under different RNA concentrations. The N protein concentration was set to 18.5 μM. **(B)** The percentage of the surface area covered by N-RNA droplets (mean ± s.d.; n = 20, two technical replicates) exhibits a reentrant behavior under an increasing RNA concentration. **(C)** Example pictures show that the N truncations mutants, except ΔR2, form spherical condensates with polyC RNA under different protein concentrations. The polyC RNA concentration was set to 50 ng/μl. **(D)** The percentage of the surface area covered by N-RNA droplets (mean ± s.d.; n = 20, two technical replicates) exhibits a reentrant behavior under an increasing protein concentration. **(E)** Phase separation of truncated N protein constructs with 0.74 μM in vitro transcribed N RNA. The protein concentration was set to 18.5 μM.

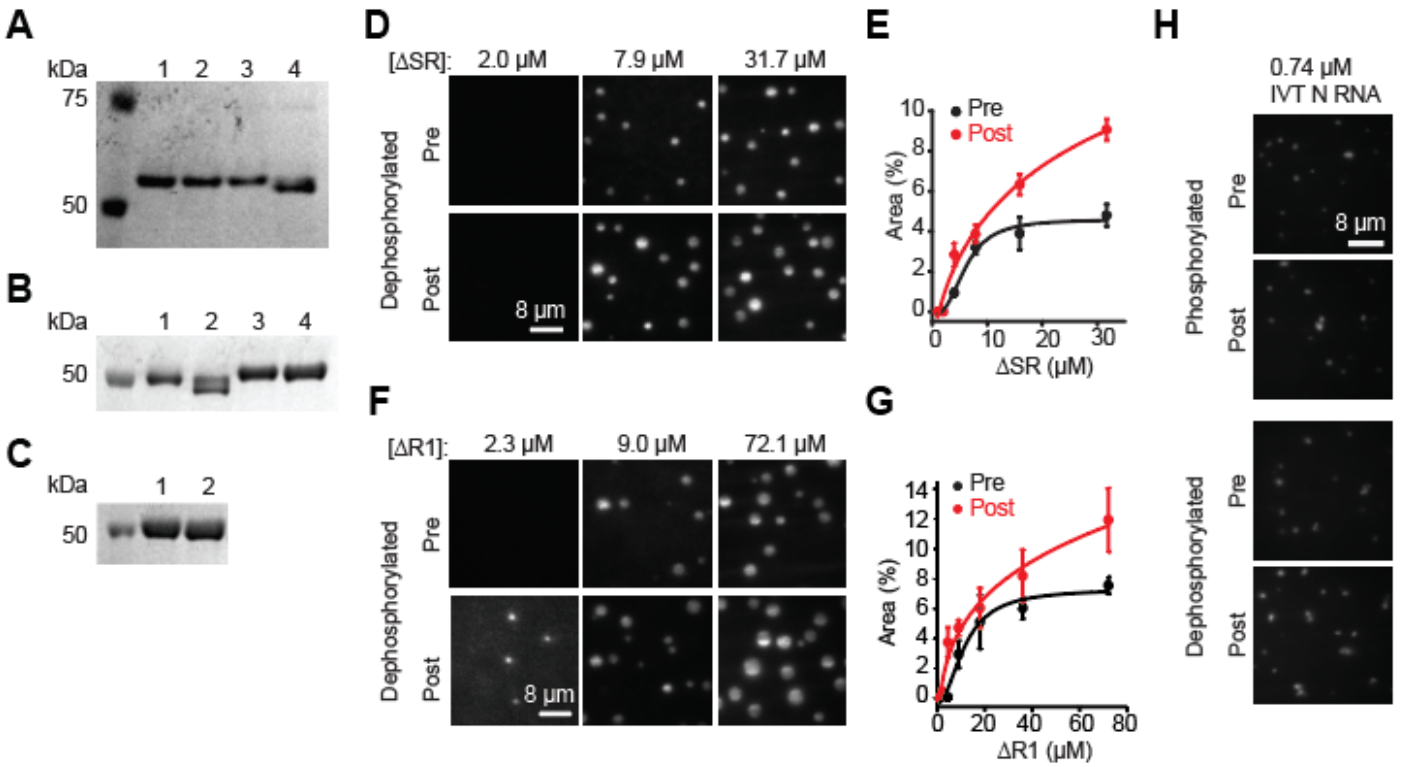

Figure 3_figure supplement 2. Phase separation of full-length and truncated N proteins under different phosphorylation conditions. (A) A denaturing gel shows the molecular weight of full-length N protein under different phosphorylation conditions (Lanes 1: pre-phosphorylation, 2: post-phosphorylation, 3: pre-dephosphorylation, 4: post-dephosphorylation). **(B)** A denaturing gel shows the molecular weight of ΔR2 and ΔSR under different phosphorylation conditions (Lanes 1: ΔR2 pre-dephosphorylation, 2: ΔR2 post-dephosphorylation, 3: ΔSR pre-dephosphorylation, 4: ΔSR post-dephosphorylation). **(C)** A denaturing gel shows the molecular weight of ΔR1 under different phosphorylation conditions (Lanes 1: pre-dephosphorylation, 2: post-dephosphorylation). **(D)** Images of condensates formed by untreated and dephosphorylated ΔSR in 50 ng/μL polyC RNA. **(E)** Quantification of droplet area as a function of ΔSR concentration (mean ± s.d., n = 20 with two technical replicates). Fits to a Hill equation (solid curves) reveal k_{sat} (± s.e.). **(F)** Images of condensates formed by untreated and dephosphorylated ΔR1 in 50 ng/μL polyC RNA. **(G)** Quantification of droplet area as a function of ΔR1 concentration (mean ± s.d., n = 20 with two technical replicates). Fits to a Hill equation (solid curves) reveal k_{sat} (± s.e.). **(H)** Full-length N protein forms asymmetric condensates with 0.74 μM in vitro transcribed N RNA before and after phosphorylation and dephosphorylation. The protein concentration was set to 35 μM.

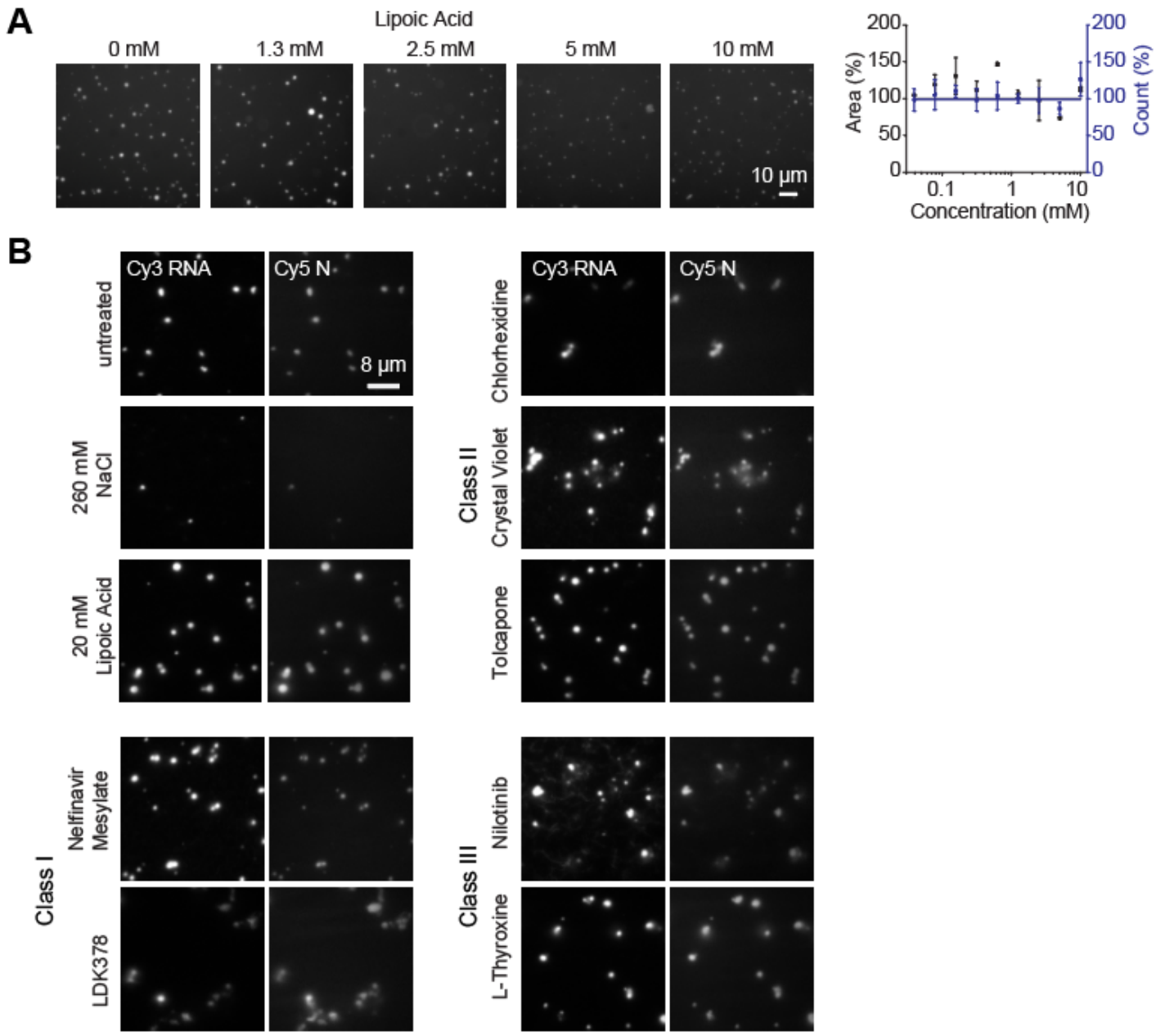

Figure 4_figure supplement 1. Phase separation of the N protein at different drug treatment conditions. (A) (Left) Example pictures show phase separation of 7.8 µM N protein and 50 ng/µL polyC RNA at different lipoic acid concentration. (Right) Quantification of the effect of lipoic acid treatment on the percent change of the total area and the number of N-RNA droplets (mean ± s.d., n = 8 with two technical replicates). **(B)** Example pictures show phase separation of 57.6 µM Cy5-labaled N protein and 1.5 µM Cy3-labeled in vitro transcribed N RNA in the presence of different drug treatment. The concentration of Class I-III drugs was set to 40 µM.

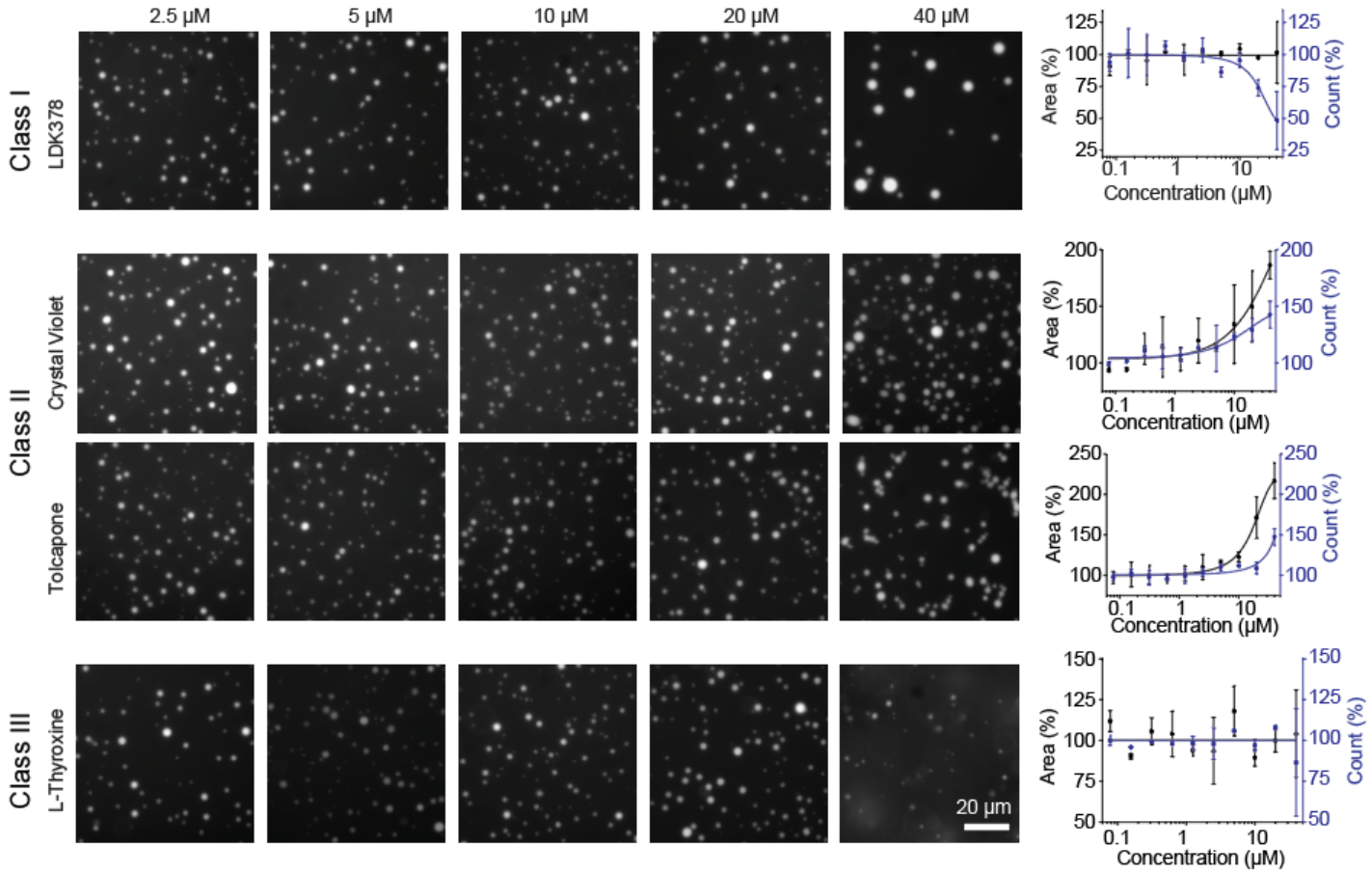

Figure 4_figure supplement 2. Phase separation of the N protein under Class I-III drug treatment. (Left) Example pictures show phase separation of 7.8 μM N protein and 50 ng/μL polyC RNA at different concentrations of LDK378, crystal violet, tolcapone, and L-thyroxine. (Right) The effect of drug treatment in the percent change of the total area and the number of N-RNA droplets (mean ± s.d., n = 8 with two technical replicates).

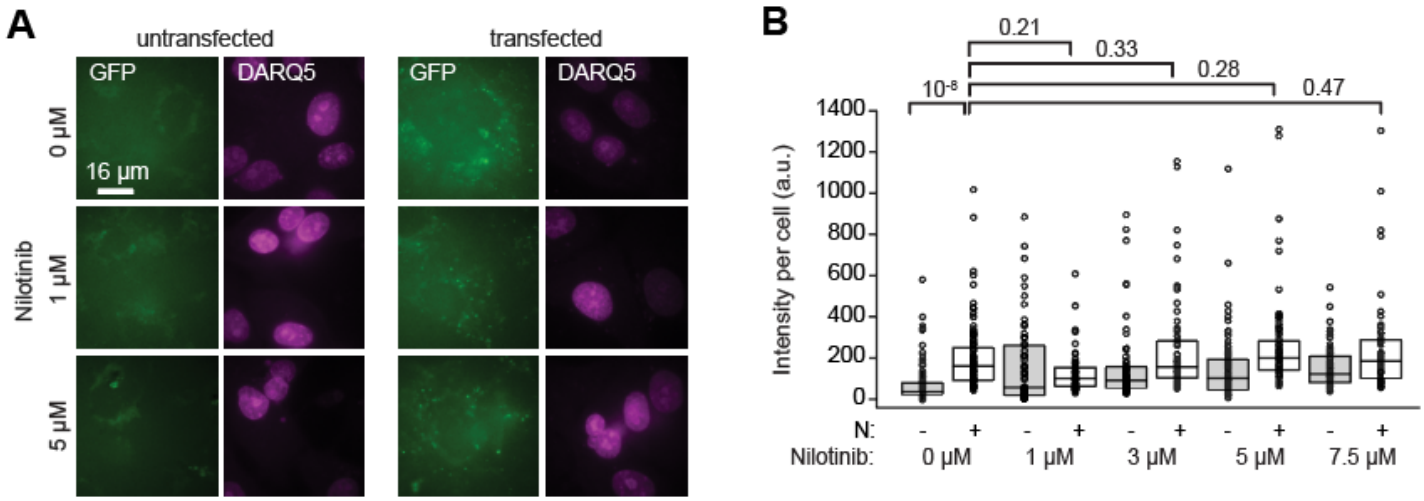

Figure 4_figure supplement 3. Nilotinib decreases the total area covered by N droplets in VeroE6 cells. **(A)** Additional examples of VeroE6 cells before or after being transfected with N-GFP under increasing nilotinib concentrations. DARQ5 marker was used to stain the nuclei. **(B)** Quantification of the background-subtracted GFP intensity per number of cells in the images. Center and edges of boxes represent median with first and third quartile. n = 118, 122, 62, 57, 64, 78, 88, 95, 67, 64 from left to right (four technical replicates for the 0 μM and 5 μM conditions; two technical replicates for all other conditions). p values are calculated from a two-tailed t-test.

Methods

Protein purification

A construct for expressing N protein with a C-terminal Strep tag was obtained from the Krogan lab (UCSF). A C-terminal yBBr labeling site was added, and deletion mutants were generated using Gibson cloning (New England Biosciences). Sufficient amounts of DNA were obtained growing 1 L of transfected XL1Blue *Escherichia coli* cells overnight and performing a gigaprep (Zymo). For protein expression, HEK GNTI cells were grown in suspension in Freestyle media (Gibco) supplemented with 2% FBS (VWR) and 1% penicillin-streptomycin (Gemini Bio-products) to 2 million cells/mL. Cells were spun down for 10 min at 1,200 rcf and resuspended in fresh, antibiotic-free media. For 250 mL cells, the transfection solution was created by mixing 1.8 mL PEI (1mg/mL, pH 7.0 in PBS) dissolved in 20 mL Freestyle media and 0.66 mg DNA dissolved in 20 mL Freestyle media. The mixture was incubated at room temperature for 15 min before being adding it to the cell culture. Transfected cell culture was grown for 72 h at 125 rpm at 37° with 5% CO_2 and 8% humidity.

Cells were then harvested at 4,000 *g* for 10 min and resuspended in 50 mL lysis buffer (50 mM HEPES pH 7.4, 1 M NaCl, 1 mM PMSF, 1 mM DTT, and 1 tablet of protease inhibitor (Sigma)). Lysis was performed using 15 loose and 15 tight plunges of a Wheaton glass dounce. The lysate was clarified using a 45 min, 360,000 *g* spin in a Ti70 rotor. The supernatant was incubated with 1 mL Streptactin sepharose beads (IBA Life Sciences) for 1 h. Beads were washed with 40 mL of lysis buffer followed by 30 mL labelling buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 10 mM $MgCl_2$, 1 mM EGTA, 10% glycerol, 1 mM DTT). Beads were then collected and incubated with purified SFP protein and a fluorescent dye functionalized with CoA (Lumidyne) at room temperature for 30 min. Beads were washed with 30 mL labeling buffer. If proteins were to be kinase or phosphatase treated, they were additionally washed with 30 mL kinase (20 mM HEPES pH 7.5, 300mM NaCl, 10 mM MgCl2, 200 μM ATP, 10% glycerol, 1 mM DTT) or phosphatase (20 mM HEPES pH 7.5, 300mM NaCl, 1mM MnSO4, 10% glycerol, 1 mM DTT) buffer. Protein was eluted in 1 mL fractions in its final buffer supplemented with 10 mM desthiobiotin, and concentrated using Amicon Ultra 30K concentrators. For kinase and phosphatase treatment, 5 μL of CKII kinase (New England Biolabs) and 2.5 μL λ phosphatase (New England Biolabs) was added per 50 μL concentrated protein, respectively, and the samples were incubated at 30 ºC for 1 h. Final protein concentration was measured using Bradford reagent, and aliquots were snap-frozen in liquid nitrogen.

In Vitro Transcription and RNA Labeling

For *in vitro* transcription of long viral RNA, the region of interest was first PCR amplified from a plasmid (N plasmid was a generous gift from the Krogan lab (*8*), 5'UTR plasmid was a generous gift from the Gladfelter lab (*28*), and SARS-CoV-2 cDNA plasmid was a generous gift of the Thiel lab (*27*)) using a forward primer with a T7 polymerase binding site (TAATACGACTCACTATAGGG). The amplified DNA was tested for purity on a 0.8% agarose gel and cleaned up using GlycoBlue and ethanol precipitation. RNA was generated using the HiScribe T7 Quick kit (New England Biolabs) and extracted using trizol and isopropanol precipitation. RNA was Cy3 labeled using a Label IT kit (Mirus Bio), and RNA purity was verified using a 0.8% agarose gel. All RNA structure predictions were done on the RNAfold server (*29, 30*).

Sample Preparation and Microscopy

Purified protein and RNA samples were diluted into the imaging buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM EGTA, 1 mM DTT, 1% pluronic) to their final concentration and introduced into the flow chamber. Samples were settled onto the coverslip for 25 min before imaging. Different buffers were used for phosphatase-treated (50 mM HEPES pH 7.4, 150 mM NaCl, 0.5 mM MnSO4, 1 mM DTT, 1% pluronic) and kinase treated (20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 100 μM ATP, 1 mM DTT, 1% pluronic) samples.

Imaging was performed using a custom-built fiber-coupled Nikon Ti-E Eclipse microscope equipped with an objective-type total internal reflection fluorescence (TIRF) illuminator and 100X 1.49 N.A. Plan Apo oil immersion objective (Nikon). The samples were excited in near-TIRF using 561 and 633 nm laser beams (Coherent). The fluorescent signal was detected by Andor Ixon electron-multiplying CCD (EMCCD) Camera (512x512 pixels). The effective pixel size was 160 nm after magnification. 10 single-frame images were acquired for each condition in each replicate. To calculate the saturation concentration of the proteins, the percent area of the slide covered by the droplets was quantified with Fiji using the Phansalkar function with a 30-pixel radius. Data fitting was performed in Origin.

Mass Spectrometry

Crosslinking analysis was performed as described by McGilvray et al. (*65*) with deviations outlined here. 15 µL 90 µM N protein in the labeling buffer was diluted 1:3 v/v into either water or labeling buffer. Diluted protein formed droplets in the water dilution, but not in the labeling buffer that contains 300 mM NaCl. Isotopically-coded light (H12) and heavy (D12) BS3 (bis(sulfosuccinimidyl)suberate) crosslinkers (Creative Molecules Inc.) were immediately added to the diluted solution. Final concentration of BS3 was 0.8 mM for 100 mM NaCl dilution and 2.5 mM for 300 mM NaCl dilution. Crosslinking was performed for 30 min at room temperature and then quenched with 1 M Tris pH 8.0 buffer. This experiment was performed in duplicate, once with D12-BS3 crosslinking N protein in droplets and H12-BS3 crosslinking soluble N protein and once with the labels reversed. The crosslinked proteins from both channels were pooled prior to acetone precipitation. Protein was precipitated with acetone overnight at -20 ºC. Protein was pelleted, supernatant was removed, and the samples were air dried for 10 min. The pellets were brought up in 8 M urea, reduced with TCEP, alkylated with iodoacetamide, diluted 4-fold, and digested with two rounds of trypsin. Peptides were desalted on a C18 MacroTrap column (Michrom Bioresources) and fractionated on a Superdex Peptide (GE Life Sciences) size-exclusion column. Fractions enriched in crosslinked peptides were dried and resuspended in 0.1% formic acid for mass spectrometry. Later eluting fractions were used for phosphorylation analysis. Each of the replicates was fractionated by size exclusion chromatography (SEC) and each fraction was injected twice, generating 16 mass spectrometry files for analysis.

Mass spectrometry was acquired on an Orbitrap Fusion Lumos coupled with an Easy-Spray nanoelectrospray ion source, a 15 cm x 75 µm PepMap C_{18} column (Thermo) and an M-class NanoAcuity UPLC system (Waters). Liquid chromatography mass spectrometry (LC-MS) runs were 90 min long. Precursor ions were measured in the Orbitrap at 120 k resolution. Selected precursor ions (triply charged and higher) were isolated and the product ions were measured in the Orbitrap at 30 k resolution. Samples that were analyzed for phosphorylation were run similarly except only HCD product ions spectra were collected and doubly charge precursors were included.

Crosslinked spectra were identified with Protein Prospector 6.2.23 (*66*) using the combination of DSS/DSS:2H12 at uncleaved Lys residues and protein N-terminus as the crosslinking reagents. The corresponding light and heavy dead-end modifications, incorrect monoisotopic peak assignment (+1Da neutral loss), N-terminal pyroglutamate formation, methionine oxidation, and acetylation and loss of the protein N-terminal methionine were set as variable modifications with three variable modifications per peptide allowed. Trypsin with two missed cleavages was the digestion enzyme and the mass tolerances were 20 and 30 ppm for precursor and product ions. The search database comprised the 14 most abundant proteins (sorted by Spectral Abundance Factor) found in the linear peptide SEC fractions, alongside a decoy database that was ten times longer. Crosslink spectral matches were classified at a 1% false discovery rate (FDR) threshold.

Protein Prospector was used to search for phosphopeptides of the linear peptide fractions using the following parameters: tryptic specificity with 2 missed cleavages, 7 and 15 ppm precursor and product ion tolerances, and carbamidomethyl (C) fixed modification. Variable modifications were as above except that Phospho (STY)

was included and only dead-end crosslink modifications were included. Peptides were reported with a maximum expectation value of 0.001, and minimum prospector score of 15. A site localization in peptide (SLIP) score threshold of 6 was used to determine site-localization. All phosphopeptide spectra reported were manually inspected for evidence of correct site assignment. Where S or T residues are adjacent, the location of the phosphosite is ambiguous (see Table S2). All phosphorylation data can be accessed here (or on the MS-Viewer website with search key "jyovxenjny").

A supplemental search of the crosslinked peaklists was made for crosslinks that also contained a phosphorylation site. This search was identical to the previous crosslinking searches except that Phospho (ST) was included as a variable modification, and the search was limited to the N protein sequence. Search results containing a phosphorylation site were manually assessed. Only phosphorylation sites that had been discovered in the phospho search of the linear peptides were considered.

For quantitating the isotopically labeled crosslinks, peak areas were measured from the extracted precursor ion chromatograms (XICs) using the small molecule interface of Skyline (v20.1.0.155). A Skyline transition list was generated containing the elemental composition of each distinct peptide pair with both light and heavy BS3 modification and in each charge state detected in the Prospector search. Retention times were present in the transition list to help with peak detection. Peptide level measurements were imported into R and summarized at the level of unique residue pairs ("crosslinks"). For each precursor ion, the $\log_2$ ratios of heavy to light peptides were calculated and mean normalized to 0 for each of the two biological replicates and then transformed into $\log_2$ (water / salt) ratios. Weighted t-tests were performed in R.

Drug Screening and Image Processing

For the FDA-approved drug screen, 75 mL of N protein at 16 μM was purified from HEK cells. polyC RNA was obtained from Sigma. The FDA-Approved Drug Library (TargetMol) has 1,200 compounds of well-characterized biological activity. 10 mM compounds were stored in 100% dimethyl sulfoxide (DMSO) in 384 well plates. For screening plates, 20 μL wash buffer was aliquoted into 384 well, glass-bottomed plates (Greiner Bio-One) and incubated for 5 min. Buffer was removed, and 0.5 μl of 2 mM compounds were stamped into 384 well plates with an Analytik-Jena Cybio Well Vario liquid handler. Final concentration of 7.8 μM N protein and 50 ng/μL polyC RNA were added to each plate. Each well contained 25 μL mixture and 40 μM compound in the primary screen.

Wells were homogenized with a Bioshake 3000 ELM orbital shaker at 2,400 rpm for 45 s and droplets were allowed to settle for 1 h before imaging. 4 images (each 224 x 167 μm) were taken at 40x with a Molecular Devices ImageXpress Micro High Content Imaging System. Images were analyzed with Metamorph Imaging software calculating droplet count and area. Data was then uploaded to CCDvault for normalization to DMSO vehicle control wells. Wells that exhibit 3 standard deviations from the untreated sample and as well as others that produced qualitative morphological changes in N droplets were selected for dose response screening. The same N protein and RNA were tested against candidate drugs using a 10-point serial dilution starting at 40 μM using the same procedure. Dose response curves were generated by fitting to the equation:

$$y(C) = y_{min} + \frac{(y_{max} - y_{min})}{1 + \left(\frac{C}{EC_{50}}\right)^n}$$

where C is the concentration of the drug, n is the Hill coefficient, EC_{50} is half-maximal inhibition concentration, and y is either the percentage of the area covered by N condensates or the number of N droplets in 0.038 mm^2 imaging area.

Vero E6 Cell Culture and Drug Testing

Vero E6 cells were cultured in phenol-negative DMEM media supplemented with 10% FBS and 1% PS at 37° with 5% CO_2. For drug testing, cells were transferred to glass-bottomed plates (Nunc Lab-Tek, 0.4 mL working volume) in media without PS and allowed to recover for 24 hours. Cells were transfected with N with a C-terminal GFP tag. For each well, 2 µL PEI (1mg/mL, pH 7.0 in PBS) was dissolved in 200 µL DMEM media and 0.66 µg DNA were dissolved in 200 µL DMEM media. The mixture was incubated at room temperature for 15 min before being adding it to the cell culture. Cells were allowed to express N-GFP for 48 h; then nilotinib or DMSO was added to the media and cells were cultured for an additional 24 hours. An hour prior to imaging, DRAQ5 stain at 1:1000 was added to the media. Cells were imaged live using epifluorescence imaging using the same laser lines and camera as described above. Puncta were identified in Fiji using the Niblack function (radius = 30 pixels, parameter 1 = 5). Each measurement of puncta per nucleus represents the number of detected puncta divided by the number of detected nuclei in a single image. Solidity (area/convex area) was measured by Fiji. Significance tests were performed in Origin.

Data Availability and Transparency

Raw data and software used in this study are uploaded to github.

Author Contributions

A.J., L.F., M.J.T., E.W., J.S., and A.Y. conceived of experiments. A.J. cloned the constructs and generated RNA. A.J. and L.F. purified the protein and performed the *in vitro* experiments. M.J.T. performed the mass spectrometry experiments. A.J. cultured the cells and performed live cell imaging. A.J., M.J.T., E.W., and A.N. analyzed the data. K.C. created the model figures. A.J., L.F., M.J.T., E.W., A.N. and A.Y. wrote the manuscript with inputs from all authors.

Funding

A. J. is supported by the NSF GRFP fellowship (DGE-1752814). L.F. is supported by the NIH F32 Fellowship (GM123655). Experiments were supported by NSF (MCB-1954449, A.Y.) and NIGMS (R35GM136414, A.Y.). Experiments at the UCSF Biomedical Mass Spectrometry Resource are supported by the Dr. Miriam & Sheldon G. Adelson Medical Research Foundation (AMRF) and the UCSF Program for Biomedical Breakthrough Research (PBBR). The content is solely the responsibility of the authors and does not represent the official views of funding institutions.

Acknowledgements

We thank John T. Canty and other members of Yildiz laboratory for helpful discussions, Thomas Graham (UC Berkeley) for advice and reagents for in vitro transcription, Xammy Nguyenla, Erik Van Dis and Sarah Stanley (UC Berkeley) for Vero E6 cells, Kathy Li (UCSF) for assisting with CLMS sample preparation, Nevan Krogan (UCSF), Amy Gladfelter (UNC) and Volker Thiel (Univ. Bern) for providing the plasmids, the UC Berkeley MacroLab for TEV protease and competent cells, and the UC Berkeley Cell Culture Facility for HEK GNTI cells.

Declaration of interests

The authors declare no competing interests.

References

1. J. Shang *et al.*, Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221-224 (2020).
2. K. K. Chan *et al.*, Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science*, eabc0870 (2020).
3. M. Hoffmann *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280.e278 (2020).
4. H. Yao *et al.*, Molecular architecture of the SARS-CoV-2 virus. *bioRxiv*, 2020.2007.2008.192104 (2020).
5. S. Klein *et al.*, SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *bioRxiv*, 2020.2006.2023.167064 (2020).
6. G. Wolff *et al.*, A molecular pore spans the double membrane of the coronavirus replication organelle. *bioRxiv*, 2020.2006.2025.171686 (2020).
7. Y. L. Siu *et al.*, The M, E, and N structural proteins of the severe acute respiratory syndrome coronavirus are required for efficient assembly, trafficking, and release of virus-like particles. *J Virol* **82**, 11318-11330 (2008).
8. D. E. Gordon *et al.*, A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv*, 2020.2003.2022.002386 (2020).
9. M. Bouhaddou *et al.*, The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **182**, 685-712.e619 (2020).
10. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434 (2019).
11. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
12. A. A. Hyman, C. A. Weber, F. Julicher, Liquid-liquid phase separation in biology. *Annual review of cell and developmental biology* **30**, 39-58 (2014).
13. A. Jain, R. D. Vale, RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247 (2017).
14. S. DW *et al.*, Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, (2020).
15. J. A. Riback *et al.*, Composition dependent phase separation underlies directional flux through the nucleolus. (2019).
16. R. McBride, M. van Zyl, B. C. Fielding, The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991-3018 (2014).
17. J. Cubuk *et al.*, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *bioRxiv*, 2020.2006.2017.158121 (2020).
18. J. D. Perlmutter, C. Qiao, M. F. Hagan, Viral genome structures are optimal for capsid assembly. *Elife* **2**, e00632 (2013).
19. T. Ukmar-Godec *et al.*, Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat Commun* **10**, 2909 (2019).
20. S. Boeynaems *et al.*, Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. *Proc Natl Acad Sci U S A* **116**, 7889-7898 (2019).
21. O. D. King, A. D. Gitler, J. Shorter, The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res* **1462**, 61-80 (2012).
22. R. J. Wheeler *et al.*, Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. (2019).
23. W. Ma, G. Zhen, W. Xie, C. Mayr, Unstructured mRNAs form multivalent RNA-RNA interactions to generate TIS granule networks. *bioRxiv*, 2020.2002.2014.949503 (2020).
24. W. Ma, C. Mayr, A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3'UTR-Mediated Protein-Protein Interactions. *Cell* **175**, 1492-1506.e1419 (2018).
25. P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, A. A. Deniz, Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. *Angew Chem Int Ed Engl* **56**, 11354-11359 (2017).
26. D. W. Sanders *et al.*, Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, 306-324.e328 (2020).
27. T. Thi Nhu Thao *et al.*, Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature* **582**, 561-565 (2020).

28. C. Iserman *et al.*, Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate. *bioRxiv*, 2020.2006.2011.147199 (2020).
29. A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neuböck, I. L. Hofacker, The Vienna RNA Websuite. *Nucleic Acids Research* **36**, W70-W74 (2008).
30. R. Lorenz *et al.*, ViennaRNA Package 2.0.
31. F. Chu, D. T. Thornton, H. T. Nguyen, Chemical cross-linking in the structural analysis of protein assemblies. *Methods* **144**, 53-63 (2018).
32. T. Taverner, N. E. Hall, R. A. O'Hair, R. J. Simpson, Characterization of an antagonist interleukin-6 dimer by stable isotope labeling, cross-linking, and mass spectrometry. *J Biol Chem* **277**, 46487-46492 (2002).
33. Z. A. Chen, J. Rappsilber, Protein Dynamics in Solution by Quantitative Crosslinking/Mass Spectrometry. *Trends Biochem Sci* **43**, 908-920 (2018).
34. A. D. Davidson *et al.*, Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein.
35. N. T. Supekar *et al.*, SARS-CoV-2 Nucleocapsid protein is decorated with multiple N- and O-glycans. *bioRxiv*, 2020.2008.2026.269043 (2020).
36. A. Savastano, A. I. de Opakua, M. Rankovic, M. Zweckstetter, Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *bioRxiv*, 2020.2006.2018.160648 (2020).
37. Q. Ye, A. M. V. West, S. Silletti, K. D. Corbett, Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *bioRxiv*, (2020).
38. C. R. Carlson *et al.*, Phosphorylation modulates liquid-liquid phase separation of the SARS-CoV-2 N protein. *bioRxiv*, 2020.2006.2028.176248 (2020).
39. T. Y. Peng, K. R. Lee, W. Y. Tarn, 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 (2008).
40. T. M. Perdikari *et al.*, SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins. *bioRxiv*, 2020.2006.2009.141101 (2020).
41. R. He *et al.*, Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem Biophys Res Commun* **316**, 476-483 (2004).
42. R. J. Wheeler *et al.*, Small molecules for modulating protein driven liquid-liquid phase separation in treating neurodegenerative disease. *bioRxiv*, 721001 (2019).
43. F. Ayman, W. Ping, A. Mahmoud, S. Hesham, *Identification of FDA Approved Drugs Targeting COVID-19 Virus by Structure-Based Drug Repositioning*. (2020).
44. D. Kumar, V. Chandel, S. Raj, B. Rathi, In silico identification of potent FDA approved drugs against Coronavirus COVID-19 main protease: A drug repurposing approach. *2020* **7**, 10 (2020).
45. Y. Li *et al.*, Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. *bioRxiv*, 2020.2001.2028.922922 (2020).
46. L. Mittal, A. Kumari, M. Srivastava, M. Singh, S. Asthana, Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach. *J Biomol Struct Dyn*, 1-19 (2020).
47. N. Yamamoto *et al.*, HIV protease inhibitor nelfinavir inhibits replication of SARS-associated coronavirus. *Biochem Biophys Res Commun* **318**, 719-725 (2004).
48. S. Jeon *et al.*, Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs. *Antimicrob Agents Chemother* **64**, (2020).
49. Y. Y. Ke *et al.*, Artificial intelligence approach fighting COVID-19 with repurposing drugs. *Biomed J*, (2020).
50. A. Trezza, D. Iovinelli, A. Santucci, F. Prischi, O. Spiga, An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors.
51. J. Dyall *et al.*, Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. *Antimicrob Agents Chemother* **58**, 4885-4893 (2014).
52. G. Garcia *et al.*, Antiviral Drug Screen of Kinase inhibitors Identifies Cellular Signaling Pathways Critical for SARS-CoV-2 Replication. *bioRxiv*, 2020.2006.2024.150326 (2020).
53. P. J. Davis, H. Y. Lin, A. Hercbergs, K. A. Keating, S. A. Mousa, Coronaviruses and Integrin αvβ3: Does Thyroid Hormone Modify the Relationship? *Endocr Res* **45**, 210-215 (2020).

54. H. Chen *et al.*, in *Cell Res*. (England, 2020).
55. S. Lu *et al.*, The SARS-CoV-2 Nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *bioRxiv*, 2020.2007.2030.228023 (2020).
56. E. M. Langdon *et al.*, mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922-927 (2018).
57. C. H. Wu *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 (2009).
58. C. H. Wu, P. J. Chen, S. H. Yeh, Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* **16**, 462-472 (2014).
59. S. Guseva *et al.*, Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci Adv* **6**, eaaz7095 (2020).
60. B. S. Heinrich, Z. Maliga, D. A. Stein, A. A. Hyman, S. P. J. Whelan, Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. *mBio* **9**, (2018).
61. J. Nikolic *et al.*, Negri bodies are viral factories with properties of liquid organelles. *Nat Commun* **8**, 58 (2017).
62. X. Tian *et al.*, Clinical Pharmacokinetic and Pharmacodynamic Overview of Nilotinib, a Selective Tyrosine Kinase Inhibitor. *J Clin Pharmacol* **58**, 1533-1540 (2018).
63. S. Galimberti *et al.*, The CoV-2 outbreak: how hematologists could help to fight Covid-19. *Pharmacol Res* **157**, 104866 (2020).
64. M. Takeda *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 (2008).
65. P. T. McGilvray *et al.*, An ER translocon for multi-pass membrane protein biogenesis. *Elife* **9**, (2020).
66. M. J. Trnka, P. R. Baker, P. J. Robinson, A. L. Burlingame, R. J. Chalkley, Matching cross-linked peptide spectra: only as good as the worse identification. *Mol Cell Proteomics* **13**, 420-434 (2014).