Biomolecular condensates control and are defined by RNA-RNA interactions that arise in viral replication

Dilimulati Aierken[1,2], Vita Zhang[3,4], Rachel Sealfon[5], John C. Marecki[6], Kevin D. Raney[6], Amy S. Gladfelter[3], Jerelle A. Joseph[1,2,*], and Christine A. Roden[3,7,8,*]
[1]Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey, USA
[2]Omenn–Darling Bioengineering Institute, Princeton University, Princeton, New Jersey, USA
[3]Department of Cell Biology, Duke University, Durham, North Carolina, USA
[4]Department of Biochemistry, Duke University, Durham, North Carolina, USA
[5]Center for Computational Biology, Flatiron Institute, New York, NY, USA
[6]Department of Biochemistry, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
[7]Department of Biochemistry and Molecular Medicine, University of Montreal, Montreal, QC, Canada
[8]Lead contact
*Correspondence: jerellejoseph@princeton.edu (J.A.J), christine.roden@umontreal.ca (C.A.R)

SUMMARY

Cells must limit RNA–RNA interactions to avoid irreversible RNA entanglement. Cells may prevent deleterious RNA-RNA interactions by genome organization to avoid complementarity however, RNA viruses generate long, perfectly complementary antisense RNA during replication. How do viral RNAs avoid irreversible entanglement? One possibility is RNA sequestration into biomolecular condensates. To test this, we reconstituted critical SARS-CoV-2 RNA–RNA interactions in Nucleocapsid condensates. We observed that RNAs with low propensity RNA–RNA interactions resulted in more round, liquid-like condensates while those with high sequence complementarity resulted in more heterogeneous networked morphology independent of RNA structure stability. Residue-resolution molecular simulations and direct sequencing-based detection of RNA–RNA interactions support that these properties arise from degree of trans RNA contacts. We propose that extensive RNA–RNA interactions in cell and viral replication are controlled via a combination of genome organization, timing, RNA sequence content, RNA production ratios, and emergent biomolecular condensate material properties.

KEYWORDS

1 RNA, 2 RNA-RNA intermolecular interactions, 3 biomolecular condensation, 4 RNA granules, 5 RNA viruses, 6 RNA-binding proteins, 7 Molecular dynamics simulations, 8 SARS-CoV-2

INTRODUCTION

Proponents of the RNA world hypothesis for the evolution of life suggest that in the primordial soup of ancient earth, the first self-replicating biological polymer was an RNA molecule[1]. In this model, the ability to synthesize a single RNA-binding protein from this self-replicating RNA template was eventually acquired, followed much later by the switch from RNA to DNA as the primary mode of information storage for life. What events would lead to the switch from the purely self-replicating RNA system to self-replicating RNA that synthesizes its own RNA-binding protein?[2] Two possible models for the selective advantage that could have been conferred by the creation of RNA-binding proteins have been proposed: **(1)** the improved protection of the RNA sequence from the harsh conditions of ancient earth and **(2)** the improved catalytic rate of the self-replicating RNA templates through increased local concentration in the presence of RNA-

binding proteins (RBP). These primordial RBPs may have promoted formation of biomolecular condensates[3], consisting of non-membrane bound assemblies of the self-replicating RNA sequence and its encoded RNA-binding protein. An additional problem in this primordial molecular context is the risk of sense RNA sticking to anti-sense template strand RNA which would inhibit both RNA self-replication and further protein synthesis. We explore here a third possible role for the function of original RBPs in preventing non-productive RNA-RNA interactions.

Any self-replicating RNA polymers of ancient earth are now long degraded, but plus strand RNA viruses provide an extent situation where sense and antisense RNAs must coexist by necessity and at potential peril for RNA entanglement.

Plus strand RNA viruses do not have any DNA phase in their replication cycle and instead generate new copies of their protein-coding capable genomes by creation of a perfectly complementary RNA template sequence consisting of minus strand or anti-sense RNA[4]. Thus, plus strand RNA viruses are thought to resemble the original self-replicating RNAs of ancient earth. As a consequence of this replication strategy, plus strand RNA viruses must produce RNA that is both perfectly complementary and the entire length of the genome. How then do viruses prevent or resolve disastrous genome-long complexes of plus and minus strand RNA, which would prevent protein production and trigger innate immune surveillance via the formation long duplexes that resemble DNA duplexes in length?

To examine how RBPs may chaperone RNA-RNA interactions and mitigate risk of irresolvable and risky RNA duplexes, we study complementary RNA sequences that would coexist and at least transiently interact for SARS-CoV-2 replication. We show that highly stable RNA-RNA interactions produce arrested networks of the RBP Nucleocapsid protein forming condensates. This is mitigated by staggering the addition of the condensing protein (plus, protein, minus), the order observed in the replication strategy of the virus. Thus, RNA viruses may employ an ancient method of coating their genomes in condensing RNA-binding proteins to prevent plus and minus strand RNA–RNA interactions during replication. Our data will show that the formation of biomolecular condensates represents an evolutionary trade off with low affinity interactions supporting beneficial functions and spherical condensate morphologies associated with more dynamic exchange. Conversely, detrimental effects are conferred by high affinity RNA–RNA interactions yielding arrested, gel-like, morphologies. Our data will further demonstrate viruses have likely evolved multiple strategies to mitigate detrimental RNA–RNA interactions and that host eukaryotic cells avoid the problem with DNA genome organization that is depleted in perfectly complementary RNA sequences. Thus, we propose that an original purpose of RNA-binding proteins may have been to prevent spurious RNA–RNA interactions between perfectly complementary strands of RNA and that the innovation of DNA was an attempt to avoid the problem entirely. These data are thus of outsized importance to RNA viruses but likely these principles extend to all RNAs, and RNA encoded biomolecular condensates in cells.

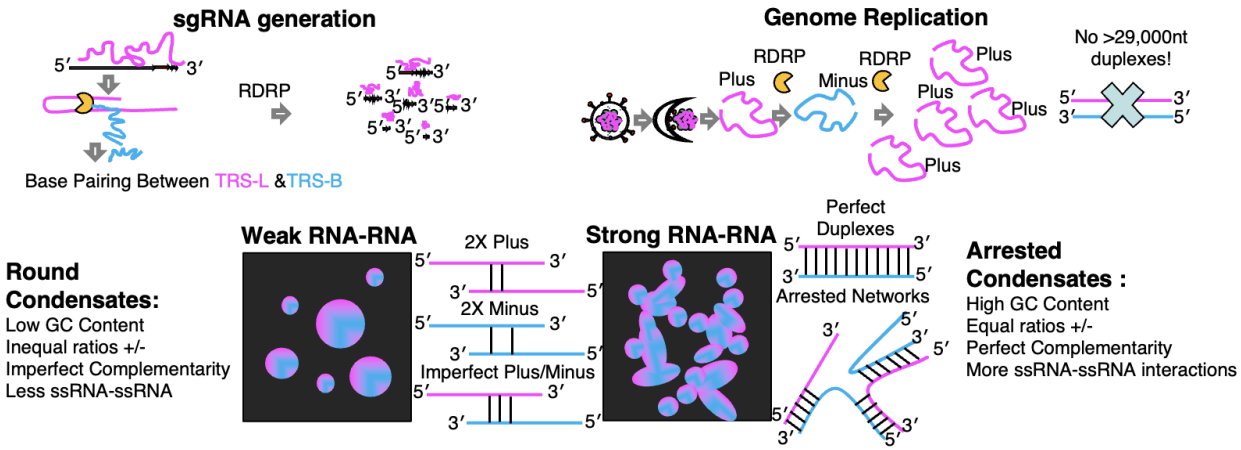
RESULTS

Choice of the model biomolecular condensate: Plus strand viral RNA model system

In this study, we are examining the cause and consequence of RNA–RNA interactions in biomolecular condensates. We are testing the two hypotheses that 1) beneficial low affinity interactions are chaperoned by condensates and 2) detrimental high affinity interactions are avoided by the employment of condensing RNA-binding proteins. Both of these can be tested by using the innate replication needs of viruses where plus genome and minus strand templates must be prevented from sticking to each other. By far the most extreme example of this potential problem is found in nidoviruses, which have the longest described genomes of any RNA virus[5]. Nidovirus is the order that encompasses SARS-CoV-2, the virus responsible for the Covid-19 pandemic[6]. Like most plus strand RNA viruses, SARS-CoV-2 encodes its own >29,000

nucleotide long plus strand RNA genome, an RNA-dependent RNA polymerase complex[7], and an RNA-binding protein capable of undergoing biomolecular condensation with the viral RNA genome, the nucleocapsid protein[8–19]. As a result of extensive pandemic related research, SARS-CoV-2 biomolecular condensates are among the best characterized with respect to the contributions of RNA[11,20–22], where distinct RNA sequence and structure features yield emergent material properties in the resulting biomolecular condensates. These studies lead to the prediction that specific RNA sequences are in part selected for their impact on specific condensate material properties.

Graphical Abstract: SARS-CoV-2 replication cycle employs weak and strong RNA-RNA interactions



Summary:
Detection and quantification of RNA-RNA interaction strength by direct RNA sequencing and simulations
Weak RNA-RNA interactions in imperfectly complementary sequences result in round condensates
Strong RNA-RNA Interactions in perfectly complementary RNA produce arrested networks
Cells and viruses use multiple strategies to avoid strong RNA-RNA interactions

For this manuscript, we will focus on one critical proposed function for SARS-CoV-2 biomolecular condensates, the generation of sub-genomic RNA via RNA–RNA interactions which we and others have proposed could be chaperoned by biomolecular condensate[23–25]s. Utilizing this system, we will first explore whether 5′UTR containing RNA sequences (5′end), plus and minus strand RNA TRS-B sequences, can co-condense.

The name nidoviruses comes from the Latin word for nested, referring to the organizational strategy for the viral RNA genome protein-coding sequences[6,26]. Nidoviruses employ two strategies for protein production. Non-structural proteins (e.g., do not contribute to virion assembly) are generated via a combination of ribosomal frameshifting and protease digestion[27–32]. Structural proteins are generated via the production of sub-genomic RNA (sgRNA)[28,33]. SgRNA production begins during minus-strand RNA production. As the RNA-dependent RNA polymerase synthesizes minus RNA from the 3′ orientation, the synthesis skips from the transcription regulatory sequence body (TRS-B) to the TRS leader (TRS-L) located in the 5′UTR of the genome. The resulting completed minus strand RNA template is used to produce plus strand RNA with one sequence each coding for Spike, ORF3a, Membrane etc., proteins[34]. Productive sgRNA generation is thought to require base pairing between TRS-L (plus) sequences and nascent anti-sense TRS-B sequences (minus)[34–36]. We previously observed that TRS-L/B and TRS-like sequences (YRRRY; where Y = C/U and R = A/G

125 nucleotides) are important drivers of SARS-CoV-2 biomolecular condensation with N protein
126 and are recognized by the N-terminal RNA-binding domain of this protein[25]. In our model, we
127 speculated that the local increased density of N protein condensation promoting motifs on plus
128 strand RNA may help the plus strand RNA genome fold back on itself prior to the skip in minus
129 strand sgRNA production, effectively pinning the complementary sequences of 5′UTR TRS-L
130 and plus TRS-B, nascent minus TRS-B together in the confined volume of the condensate. We
131 further speculated whether such pinning behavior could promote skipping and lead to observed
132 imbalance in sgRNA production correlated with condensation promotion.

133 ***Multiple plus and minus strand TRS-B RNA fragments are capable of driving N protein***
134 ***condensation***

135 To test if biomolecular condensates could promote critical RNA–RNA interactions, we
136 utilized our minimal reconstitution system model of SARS-CoV-2 biomolecular condensates
137 consisting of the structural nucleocapsid protein (N protein) and viral RNA fragments[21]. To this
138 end, we sought to reconstitute the essential TRS-L/B RNA–RNA interaction of nidoviruses.
139 Thus, we synthesized 13 RNA fragments from the virus that are each 483 nucleotides long,
140 including the 5′end of the genome (TRS-L containing), as well as six regions flanking the TRS-
141 B's of the Spike, Envelope, Membrane, ORF7, ORF8, and Nucleocapsid RNA in both the plus
142 and minus strand context roughly 237 nucleotides before and after the TRS-B sequence
143 **(Figure 1A**).

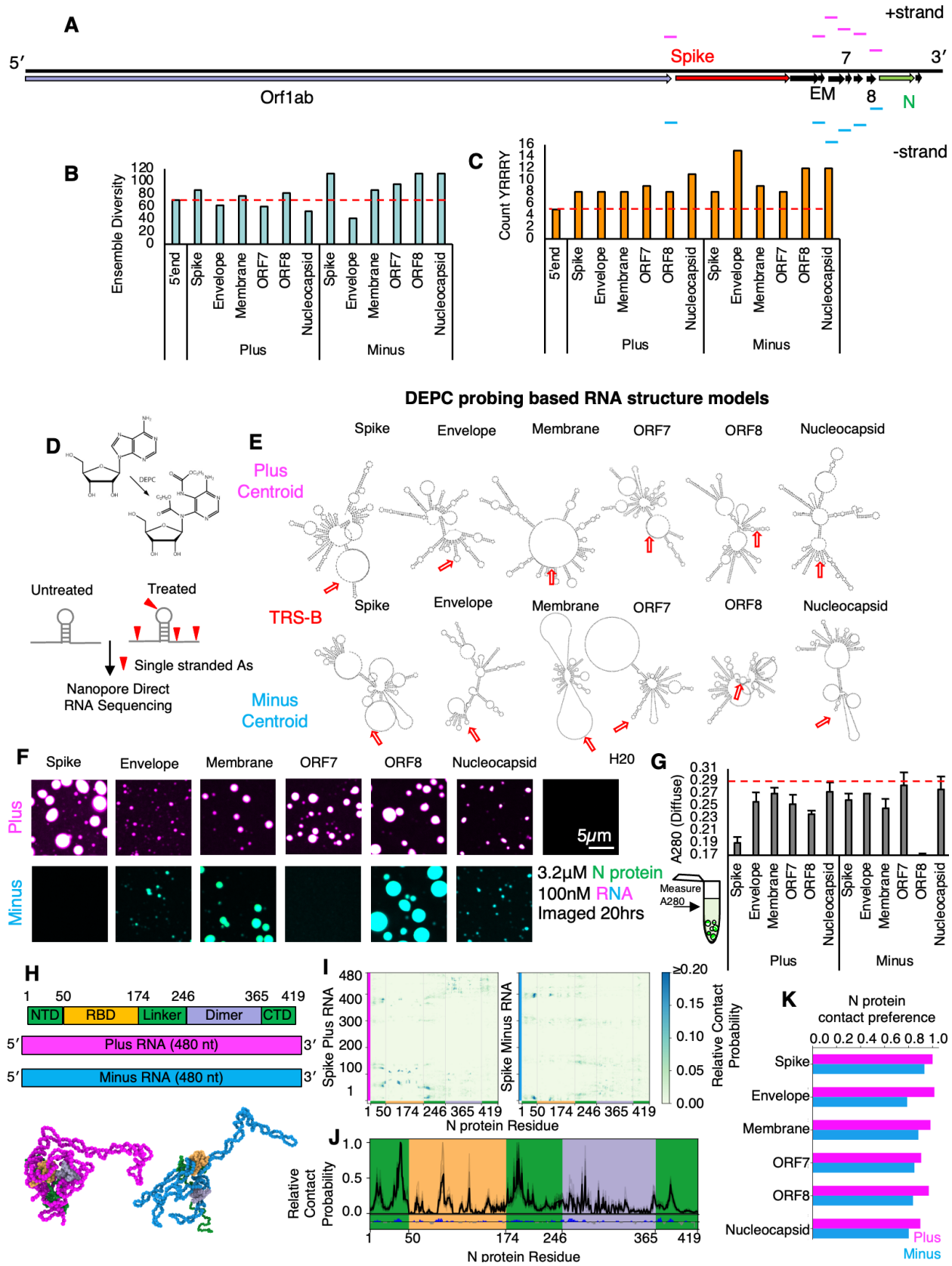


Figure 1. Plus and minus strand TRS-B fragments can undergo condensation with N

protein.

(A) Approximate location of tested fragments in the SARS-CoV-2 genome. Plus strand (magenta) minus (teal), spike (red), N/nucleocapsid (light green). Of note, minus strand RNA position is reflective of the complement rather than the reverse complement for ease of comparison to plus strand RNA positions. **(B)** Ensemble diversity predictions for each tested fragment relative to a previously identified condensate promoting sequence, the 5′end. **(C)** Count of YRRRY motifs (N protein N-terminal RNA-Binding domain target sequence) in tested fragments relative to previously identified condensate promoting sequence the 5′end. Of note, this is the total count and is not reflective of local sequence enrichment around TRS-Bs as depicted in **Supplemental Figure S1A**. **(D)** Cartoon depicting RNA structure probing via direct RNA sequencing protocol. Single stranded A nucleotides are preferentially reactive with DEPC (Diethyl pyrocarbonate) resulting in adducts detectable as mutations in nanopore direct RNA sequencing. **(E)** DEPC based RNA structure models for TRS-B (red arrow) containing fragments in plus (top row) and minus Strand (bottom) context. Plus strand RNAs tend to be more structured, particularly around the TRS-B sequence (red arrow). **(F)** Representative images of condensates formed from 100nM RNA TRS-B sequences and 3.2μM SARS-CoV-2 nucleocapsid protein (N protein). Condensates were imaged 20 hours post mixing at the glass. Images reflect a merged signal of RNA (magenta plus and cyan minus) and protein (green) labeled signals. All tested fragments other than spike minus and ORF7 minus resulted in reproducible observable condensates under these conditions. H2O water only no RNA added control indicative of the requirement of RNA to drive N protein condensation under the tested conditions. **(G)** Diffuse phase absorbance measurements for panel D. A280 signal is reflective of the non-condensate recruited portion of the RNA and protein signal. Plus Spike and minus ORF8 are best able to recruit protein and RNA into the dense phase as evidenced by the reduction of signal into the diffuse phase. Error bars are from 3 technical replicates. **(H)** (Upper Panel) Sequence composition of the N Protein and RNA used in simulations. There are 5 domains for N protein: N terminal domain, RNA-binding domain 1, Linker, RNA-binding domain 2, and C terminal domain. For representation of the protein and RNA, we use a hybrid Mpipi-SIS-RNA model (model 1; see Methods). (Lower Panel) Snapshots of dimers of N protein and Spike RNA plus (magenta) and minus (cyan) strands. **(I)** Contact maps of N protein with Spike plus RNA and minus RNA from dimer simulations. **(J)** (Upper panel) Normalized RNA contact probability distribution along the Nucleocapsid protein sequences from all plus and minus RNA sequences in this study. Each histogram from individual simulations is shown as light gray lines, and the overall average is shown as black lines. (Lower panel) The net charge per residue (NCPR) distribution of N protein. The positive charge distribution is highlighted by blue histograms while negative charge is shown in gray. **(K)** Preferences of binding for N protein and RNA from dimer simulations. Higher frequency RNA contacts in plus strand RNA than minus in simulations.

We confirmed that the synthesized RNA fragments that were chosen had a reasonable enrichment of our previously established N protein condensing features[11]. Specifically, we see low ensemble diversity[37,38] which is indicative of structured/stably folded RNA sequences and promotes the condensation through RNA-binding domain 2 of N protein (**Figure 1B**) and the presence of YRRRY sequences, which promote the condensation through RNA-binding domain 1[25] (**Figure 1C**, **Supplemental Figure S1A**). These RNA structure and sequence features are comparable to the 5′end RNA sequences that have been previously tested (**Supplemental Figure S1B, C**). To confirm RNA structure predictions, we performed RNA structure probing via direct sequencing[39] (**Figure 1D and E**). The majority (10/12) newly tested TRS-B fragments were capable of driving condensation with N protein on their own (**Figure 1F**) with 2 minus strand sequences not eliciting N protein condensates. All tested plus fragments drove condensation and, in most cases, also larger assemblies than the minus fragments, consistent

196 with the higher local density of YRRRY in the plus strand (**Supplemental Figure S1A**) and
197 more stable RNA structures (**Figure 1D and E**).

198 We compared the dilute phase A280 absorbance signal to get "bulk" reaction
199 measurements, following 20 hours of incubation with N protein (**Figure 1G**). Consistent with the
200 observation bulk depletion measurements in **Figure 1G**, minus ORF8 and plus Spike had the
201 largest observed condensate area (**Supplemental Figure S1D**) as well as more protein
202 (**Supplemental Figure S1E**) and RNA signal (**Supplemental Figure S1F**) in the dense phase,
203 these wells had the lowest A280 signal in the diffuse phase (**Figure 1G**) suggesting for these
204 wells, most of the RNA and protein is condensate recruited. To compare the dense phase
205 material properties and the environment, for each RNA system, we quantified the maximum
206 intensity of the protein (**Supplemental Figure S1E**), and the RNA signals (**Supplemental**
207 **Figure S1F**) in the dense phase of the resulting condensates. Maximum intensity was chosen
208 as it avoids non-uniform loss of signal at the condensate interface (smaller condensates have a
209 larger proportion of interface) and thus is more representative of the densest dense phase
210 environment. To examine the chemical environment, we compared the protein to RNA ratio
211 (**Supplemental Figure S1G**). We observed that, although different total RNA and protein signal
212 was observed in the dense phase depending on RNA sequence, the ratio across tested RNA
213 fragments was quite similar suggesting that each tested RNA fragment should have similar
214 dense phase chemical environment despite differences in condensate size.

215 To gain additional biophysical insight into the formation of N protein–plus or –minus
216 strand condensates, we conducted explicit chain molecular dynamics simulations. Here, we
217 have implemented a chemically specific residue-resolution approach to probe sequence-
218 dependent binding between the plus or minus strands with N protein and investigate whether
219 these interactions may be sufficient to sustain condensates. Specifically, we use the Mpipi
220 model[40], a residue-level approach for probing the phase behavior of proteins and RNA. In the
221 original Mpipi implementation, RNA lacks the ability to form base pairs between nucleotides.
222 Thus, we updated the RNA representation with the RNA-SIS model, a consistent nucleotide-
223 resolution model[41,42] that enables explicit RNA base pairing. In the resulting protein–RNA model
224 (referred herein as model 1; see Methods), each amino acid or nucleic acid is represented as a
225 single interaction site, disordered protein regions as fully flexible chains, folded regions of N
226 protein via homology-modeled structures, and nucleotides can explicitly form base-pair
227 interactions (hence, can stabilize secondary structures). Additionally, proteins and RNA interact
228 via both non-charged and charged interactions, where interactions of amino acids with G/A
229 nucleotides are favored over those with C/U based on atomistic predictions. All simulations with
230 model 1 were performed in the NVT ensemble at 300 K and 150 mM NaCl salt. Further model
231 and simulation details are included in the Methods section.

232 We first simulated dimeric systems, composed of N protein and plus or minus strands.
233 As shown in **Figure 1H**, there are five regions in N protein: the N terminal domain (NTD), RNA
234 binding domain (RBD) 1, Linker, RBD 2, and C terminal domain (CTD). Among these regions,
235 NTD, linker and CTD are disordered (green), while the RNA binding domains are mainly
236 folded.[15] In our dimer simulations, all plus and minus strands bind N protein. However, different
237 regions on plus and minus strands show high residue–residue contact frequencies with N
238 protein (**Figure 1I, Supplemental Figure S1H**). As an example, the contact map of N protein
239 with Spike plus and minus RNAs are depicted in **Figure 1I**. To further investigate the binding
240 mechanism, we assess normalized contact probabilities between all RNA strands and N protein
241 (**Figure 1J**). Despite the sequence variations, the nonspecific binding pattern, with respect to N
242 protein, is highly preserved among the RNA strands. The charge pattern analysis of N protein
243 reveals that the observed nonspecific binding patterns are mainly governed by the charge
244 distribution and the structure of N protein (**Figure 1J**). The positive charge distribution highly

correlates with the peaks of contact probability. Interestingly, even though the positive charges are almost evenly distributed throughout the sequence of N protein (**Figure 1J**), the high-probability binding sites are mainly concentrated around RBD 1 and its neighboring disordered regions—suggesting an importance of protein structure for this binding interaction. Moreover, our findings for full-length N protein with long RNAs align well with previous studies, where the dimer of partial N protein (NTD+RBD 1) and short RNAs (poly rU) were studied using Mpipi model.[43,44] Here, we predict that for the full-length N protein with much longer RNA strands, the NTD–RBD emerges as the dominant motif for interaction with RNA. Furthermore, due to size difference between the N protein and RNA, only part of the RNAs bind to N protein. As a result, the unbound nucleotides are free to base-pair with other RNA molecules or interact with additional N proteins. These data suggest that both plus and minus strands bind N protein and that non-specific charge-dominated interactions underlie the formation of condensates observed in the experiments (**Figure 1F**). We examined the preferences of binding for N protein and RNA from dimer simulations by quantifying the contact frequency in tested RNAs (**Figure 1K**). In keeping with our results from **Figure 1F**, plus strand RNA regardless of sequence had higher affinity for simulated N protein. As noted earlier, plus strands are enriched in YRRRY sequences that facilitate condensate formation via interactions with N protein RBD 1. Given that our model encodes stronger protein–RNA interactions for the larger G/A nucleotides, the simulations effectively capture such non-specific interactions between YRRRY sequences and the RBD 1. However, in our simulations we do not observe a dominance interaction between RBD 2 and the more structured plus strands. We believe that this may likely be a limitation of the model in capturing more specific interactions between proteins and RNA.

With this information we next sought to test combinations of plus and minus RNA TRS-B fragments with the 5′end (**Figure 2**) to reconstitute a critical viral RNA–RNA interaction.

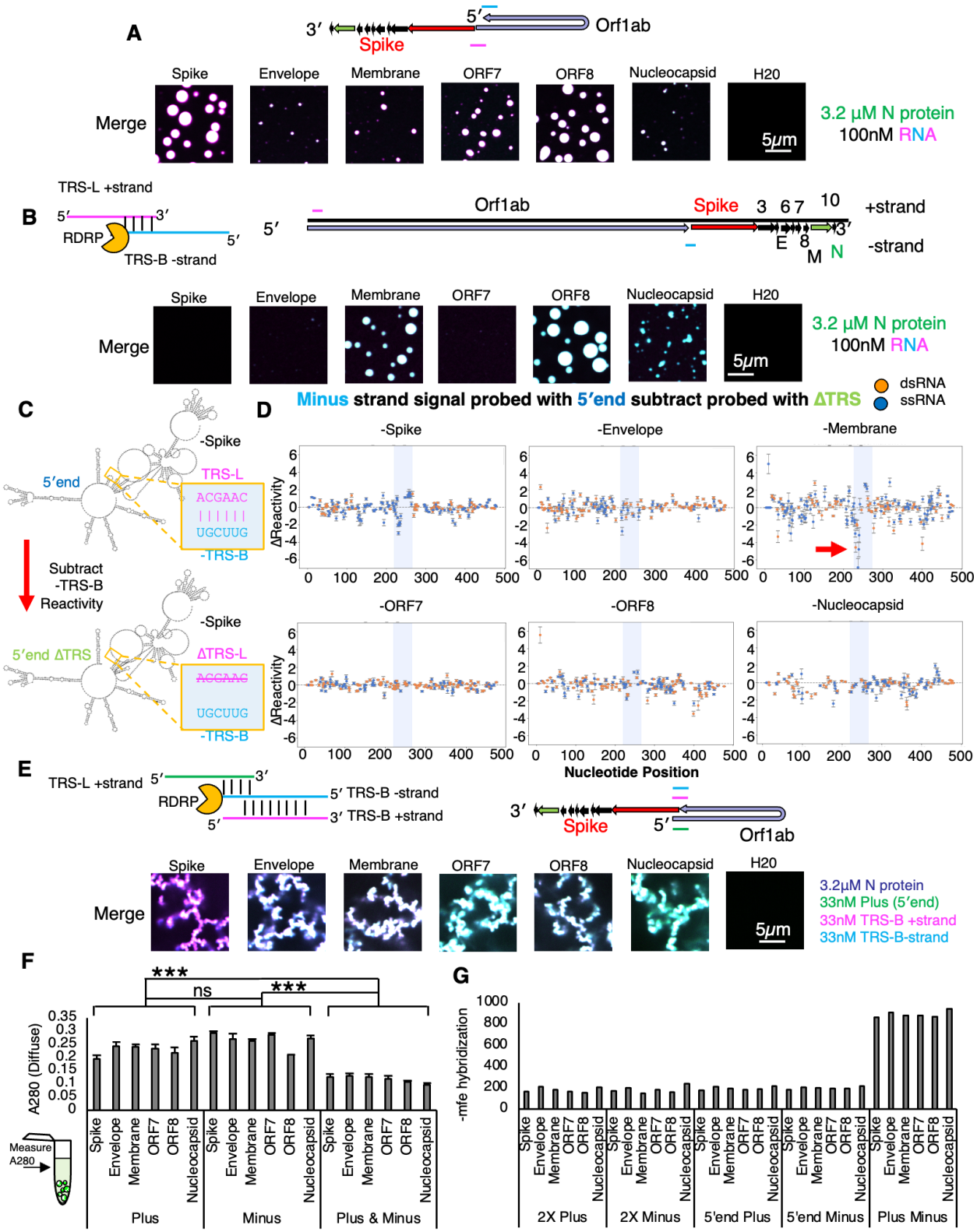


Figure 2. Reconstitution of 5′end RNA looping and sgRNA generation 5′end RNA co-condenses with TRS-B fragments.

For **A, B**, and **E**, all images depict the merged signal of the same total RNA (100nM) and protein (3.2μM) concentrations. As shown in panels **2A** and **2B** combinations of two RNAs are 50nM

each for panel **2E** combinations of RNA are 33.3nM each. All images were acquired 20 hours post mixing at the glass. Scale bar (white line in H_2O only control) is 5 microns. All images in each panel are acquired on the same day and are contrasted the same. **(A)** Cartoon depicts the theoretical arrangement of the plus strand RNA genome to tether the 5′UTR in proximity of the TRS-B of spike (red) prior to minus strand RNA synthesis which may be assisted by condensate formation. Combinations of 5′UTR containing RNA (5′end labeled in cyan) and plus TRS-Bs labeled in magenta. Depicted images are representative of merged signal (N protein in green and RNA in cyan and magenta) of 3 technical replicates and illustrate that 6/6 tested TRS-B containing plus RNAs can co-condense with 5′end in the N protein condensates. Unmerged images are depicted in **Supplemental Figure S2A**. H2O is water only no RNA control. **(B)** Cartoon depicts the theoretical arrangements base pairing between the plus TRS-L containing 5′end and minus strand TRS-Bs prior to sgRNA generation which may be assisted by condensate formation. Base pairing as indicated by black lines is indicated between the plus TRS-L and minus TRS-B. Genome position of minus fragments is reflective of complement rather than reverse complement positions. Combinations of 5′UTR/TRS-L containing RNA (5′end labeled in magenta) and minus TRS-Bs labeled in cyan. For minus TRS-B RNAs capable of driving condensation, 5′end and minus can co-condense. Depicted images are representative of merged signal (N protein in green and RNA in cyan and magenta) of 3 technical replicates and illustrate that RNAs can co-condense in the condensates. Unmerged images are depicted in **Supplemental Figure S2B**. H2O is water only no RNA control. **(C)** Overview of experimental protocol to detect RNA–RNA interactions in trans. Two RNA species are mixed in equal molar ratios (e.g. 5′end and minus strand RNA e.g. -spike) if base pairing occurs (pink and blue sequences) a change in the single stranded reactivity in the minus should be evident particularly when the sequence is absent in the plus (e.g. co-incubation with TRS-Del RNA). **(D)** Δ Reactivity plots for each minus RNA representing the change in signal due to the presence or absence of co-incubated TRS-L containing sequences in the 5′end. Minus membrane RNA shows a large decrease in A reactivity near the TRS-B (blue box/red arrow) when TRS-L is present as opposed to absent ΔTRS-L indicative of base pairing between the TRS-L and B. Change in RNA signal is most easily observed in single stranded A nucleotides (blue circles ssRNA) as opposed to double stranded A nucleotides (orange circles dsRNA). **(E)** Cartoon depicts the theoretical arrangement of plus RNA genome (plus TRS-L and plus TRS-B) as well as minus nascent synthesized RNA genome just prior to sgRNA synthesis. Base pairing as indicated by black lines could take place between TRS-L and minus but should be more extensive between minus and plus TRS-B containing fragments. Combinations of 5′UTR containing RNA (5′end labeled in green) and plus TRS-Bs in magenta minus TRS-Bs labeled in cyan. Of note a difference from all previous protein labels N protein is labeled in blue (Atto405) rather than green. Depicted images are representative of merged signal of 3 technical replicates and illustrate that RNAs can co-condense in the condensates. In contrast to A and B, condensates no longer adopt a rounded morphology. Unmerged images are depicted in **Supplemental Figure S2C**. H2O is water only control (no RNA). **(F)** Diffuse phase absorbance measurements for panels A, B and E. A280 signal is reflective of the non-condensate recruited portion of the RNA and protein signal. Despite containing identical concentrations of RNA and protein, combination which contain 3 RNAs result in more recruitment to the dense phase and lower diffuse phase signal whereas 2 RNAs were not significantly different from each other by this metric ($p<0.001$ ***) **(G)** Prediction of RNA hybridization energy between any 2 RNA species tested thus far (2× plus and 2× minus) (**Figure 1F**), 5′end/plus (**2A**/**E**), 5′end minus (**2B**/**E**), and plus/minus (**2E**). Most RNA combinations have a low -MFE prediction indicative of poor RNA–RNA interaction. Plus/minus has a high predicted -MFE indicative of strong RNA–RNA interaction. Of note, -MFE is predicted rather than MFE as positive numbers can be more intuitive.

Reconstitution of low and high affinity viral RNA–RNA interactions in condensates

325 During sgRNA generation, the RdRp jumps from the TRS-B to the TRS-L sequences
326 skipping ORF1ab sequences in a base pairing and N protein dependent process[27,31,34]. We have
327 postulated that N protein/TRS-RNA condensation may also help promote this skipping by
328 tethering the two RNA sequence in a smaller volume of the condensate[25]. We reasoned that the
329 comparatively higher local density of N protein condensation promoting motifs on the plus RNA
330 as observed in **Figure 1** and **Supplemental Figure S1A** might allow for co-condensation of the
331 5′end and the TRS-Bs effectively tethering the 5′end in place adjacent to the location of the
332 jump in cis. This tether, which could arise by N protein condensate formation, may support
333 sgRNA generation by helping the polymerase skip from the TRS-B to the TRS-L, looping out the
334 intervening plus strand genome sequence. To test if the 5′end and the plus TRS-Bs containing
335 fragments could co-condense, we reconstituted labeled 5′end RNA with labeled TRS-B
336 containing fragments. We were able to observe that, for fragments that resulted in condensates,
337 both labeled RNAs were present in the dense phase (**Figure 2A, Supplemental Figure S2A**).

338 Extensive base pairing between the plus strand 5′end TRS-L sequence and the minus
339 strand TRS-B is thought to be essential for the proper generation of the minus strand sgRNA
340 sequence[33,34]. This is because base pairing may promote skipping of the RNA dependent RNA
341 polymerase from the body to the leader resulting in the needed chimeric sequence. To test if
342 minus strand RNA could also enter co-condensates with the 5′end and thus promote this
343 interaction we tested the labeled 5′end RNA with each of the labeled TRS-B fragments in the
344 minus strand context. Again, we observed under the tested conditions that the primary arbiter of
345 condensation appeared to be the TRS-B minus strand RNA identity and minus strand RNA was
346 poorly able to drive condensation (**Figure 2B, Supplemental Figure S2B**).

347 To confirm that TRS-L and B sequences were capable of pairing in our experimental
348 conditions and without N protein, we performed RNA-RNA interaction mapping for each of the
349 tested minus TRS-B fragments co-incubated with the plus 5′end fragment which contains either
350 the wildtype or TRS-L deleted sequence. We could identify regions in each minus strand RNA
351 that were paired in trans to the TRS-L by comparing the reactivity of nucleotides when probed in
352 the minus TRS-B RNAs in either the presence of 5′end or the 5′end ΔTRS sequence (**Figure**
353 **2C**), By subtracting the probed with ΔTRS-L from the wildtype we observed the greatest change
354 in reactivity was found in the region surrounding the minus TRS-B (light blue box). (**Figure 2D**),
355 suggesting that most of our RNA combinations had successfully reconstituted base pairing
356 between the plus TRS-L and minus TRS-B. Of note, the highest change in reactivity (indicative
357 of stronger base pairing) was observed in the -membrane RNA fragment. Intriguingly, this RNA
358 was the most single stranded (blue dots) in the region surrounding the TRS-B in our RNA
359 structure data (**Figure 1E**) suggesting that the single stranded RNA content in the area
360 surrounding the minus TRS-B may facilitate pairing between the TRS-L and TRS-B.

361 With this information, we reasoned that if our model was correct and N protein co-
362 condensation was able to promote the jump between TRS-B and TRS-L, all three RNA
363 sequences would be present together in the same location[34], perhaps in the dense phase of an
364 N protein biomolecular condensate[25], just after minus strand synthesis. Thus, we differentially
365 labeled all three RNAs and mixed them with labeled N protein. Surprisingly, under the tested
366 conditions, not only were all three RNAs present in the same dense phase, but the morphology
367 of the resulting assemblies also differed significantly from any previously tested condition with
368 condensates consistently forming giant arrested networks (**Figure 2E, Supplemental Figure**
369 **S2C**) Comparison between the diffuse phase of all three tested conditions suggested that
370 presence three RNAs, despite having the same total RNA concentration (100nM in all cases),
371 was much better at recruiting RNA and protein to the dense phase than single RNAs (**Figure**
372 **1F**). We reasoned that the profound morphology change observed in the condensates may be
373 due to a difference in RNA–RNA interaction strength[37,38]. To this end, we predicted the

374 hybridization energy[45] (**Figure 2G**) of all RNA combinations tested so far double the
375 concentration or 2× plus and 2× minus (**Figure 1F**), plus with 5′end (**Figure 2A**), minus with
376 5′end (**Figure 2B**), and plus with minus (**Figure 2E**). We observed that the highest predicted
377 hybridization energy was found in the plus/minus RNA combinations consistent with the perfect
378 complementarity of these RNAs. Very similar lower values were observed for every other
379 combination. Thus, we reasoned that the profound change in morphology was well correlated
380 with the strength of RNA–RNA interaction in the condensate (**Figure 2G**).

381 ***Reconstitution of perfectly complementary RNA–RNA interactions alone is sufficient to***
382 ***yield arrested networks***

383 To test if plus and minus strand RNA–RNA interaction strength alone was sufficient to
384 drive the observed morphology change, we next made condensates that contained only these
385 perfect plus/minus RNA pairs (**Figure 3A**).

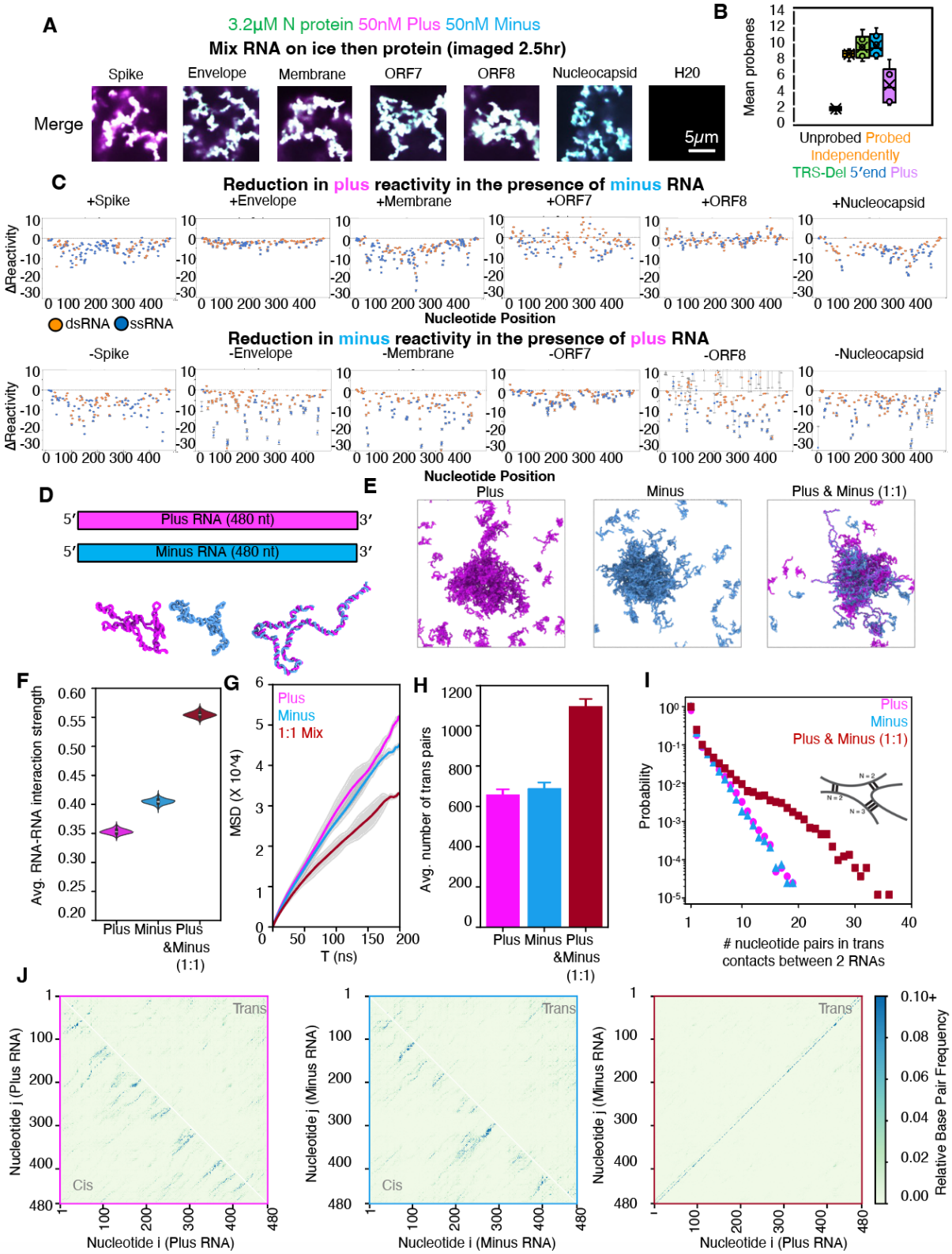


Figure 3. Mixtures of perfectly complementary plus and minus strand RNA result in arrested network formation due to increased RNA-RNA interactions.

389 **(A**) Premixing 50nM each of plus and minus RNA on ice followed by N protein addition (3.2μM)
390 is sufficient to produce arrested networks. For **A**, all images depict the merged signal of the same
391 total RNA (100nM) and N protein (3.2uM) concentrations. Combinations of two RNAs are 50nM
392 each. Scale bar (white line in H_2O only control) is 5 microns. All images in each panel are acquired
393 on the same day and are contrasted the same. **(B)** Mean probeness of minus strand RNA, without
394 probing (black), probed independent of any plus RNA (orange), probed in the presence of
395 imperfectly complementary RNA (green 5′end ΔTRS, blue 5′end wildtype) or probed in the
396 presence of perfectly complementary RNA (purple). Reduction of probed signal, indicative that
397 base pairing is strongest when minus strand RNA is probed in the presence of perfectly
398 complementary plus RNA. These data are indicative of base pairing in trans in this condition. **(C)**
399 Nucleotide resolution Δ reactivity plots (probe plus/minus subtract probed alone) for plus strand
400 RNAs (top panel) and minus strand RNAs (bottom panel). Non-uniform decrease in signal in the
401 presence of complementary RNA as opposed to independently probed. Orange dots are double
402 stranded A nucleotides and blue dots are single stranded A nucleotides. Greater reduction in
403 reactivity is observed for minus strand RNA and single stranded A nucleotides. **(D)** Representative
404 single chains of perfectly complementary plus (magenta) and minus (cyan) RNA base pair in trans
405 in RNA only simulations (model 2; see Methods). **(E)** Representative images depicting Envelope
406 RNA simulations of 2X plus (magenta), 2X minus (cyan), or 1:1 plus and minus RNA. The analysis
407 for these simulations follows: **(F)** Average RNA–RNA interaction strength. This is obtained by
408 normalizing the potential energy of each system by the total number of nucleotides. **(G)**
409 Comparison of mean-squared displacement (MSD) of center of mass. Here, the 1:1 mixture
410 exhibits the slowest dynamics. **(H)** Average number of nucleotides pairs in trans RNA contacts.
411 **(I)** Probability distribution of number of nucleotide pairs in trans contacts. **(J)** Cis and trans
412 contacts for plus-only (left panel) and minus-only (middle panel) systems, as well as trans
413 contacts for 1:1 mixture (right panel).

414 We observed that for all tested RNA pairs (6/6) mixing perfectly complementary plus and
415 minus strand RNA were all capable of driving the networked condensates phenotype (**Figure**
416 **3A**). To determine if the morphology of the condensates was due to RNA–RNA interaction
417 strength, we measured RNA–RNA interaction strength directly using DEPC probing and
418 sequencing[39] (**Figure 3B**). We observed unprobed minus RNA, regardless of sequence, had
419 low level of probing as evidenced by a lack of mutant Adenosines in the raw sequencing data.
420 Minus RNA that was probed independently of any plus RNA had much higher level of probing.
421 Minus strand RNA that was probed in the presence of plus strand RNA that was not perfectly
422 complementary (5′end or TRS-del) had on average comparable levels of reactivity to minus
423 strand RNA that was probed independently of any other sequence being present. In striking
424 contrast, minus strand RNA that was probed in the presence of perfectly complementary plus
425 strand RNA showed a profound reduction in probing signal, perhaps due to trans base pairing of
426 the previously single stranded Adenosines in the minus strand to the complementary Uracils in
427 the plus strand rendering them inaccessible to chemical probing by DEPC (**Figure 3B**). To
428 examine the nucleotide level changes in reactivity, we plotted the Δ Reactivity (the subtraction
429 of the signal for the RNA probed independently from the signal for RNA probed in the presence
430 of its complement) (**Figure 3C**). We observed that individual nucleotide reactivity had differential
431 sensitivity and that plus stranded RNA signal change was far less uniformly reduced than
432 minus. This variability in the sensitivity may be due to differences in single strandedness of the
433 nucleotides in plus versus minus strand RNA. As plus RNA tends to be more structured than
434 minus RNA[46] (**Figure 1B and 1E**), changes in individual nucleotides are more variable (**Figure**
435 **3C**). Of note, and in keeping with previous observations from Trcek lab[47], single stranded
436 RNA/ssRNA reactivity signal (blue dots) was more altered than double stranded RNA/dsRNA
437 indicative of more base pairing for single stranded RNA in the presence of its RNA complement.
438 RNA–RNA interaction mapping by sequencing gives us nucleotide and single molecule

resolution of trans RNA–RNA contacts. It does not however give us a 3-dimensional arrangement of the RNA molecules in contact in the dense phase. To examine the arrangement of RNA molecules, we turned to molecular dynamics simulations of the RNA alone. This approach aimed to better understand the molecular origins for the arrested phenotype observed for N protein–plus–minus systems, as all our data up to this point indicated that the condensate morphology change was primarily RNA–RNA interaction strength dependent.

We hypothesized that an enhanced interaction strength from perfectly complementary RNA–RNA interactions could lead to kinetically arrested states, independent of N protein–RNA binding (and consistent with our RNA only sequencing data from (**Figure 3B and C**). To test this, we adopted a minimal RNA model[41,42] (model 2; see Methods) that describes RNA–RNA binding based on canonical and wobble base-pairing propensities. The model is trained to capture both cis and trans RNA–RNA contacts in the presence of high salt (i.e., effectively neutralized backbone charges) and, therefore, captures RNA clustering in the absence of protein (**Figure 3D, E**). Employing this approach, we investigate the clustering of plus and minus strands individually, as well as mixtures of the two strands (**Figure 3E, Supplementary Figure S3A-C**). In all cases, our simulations predict stronger RNA–RNA interaction strength for RNA clusters with perfectly complementary RNA–RNA interactions compared to their plus/minus strand analogues (**Figure 3F, Supplemental Figure S3G)**. Additionally, the degree of RNA–RNA interactions was directly but not perfectly proportional to the GC content of the RNA (R^2 = 0.61) (**Supplementary Figure S3K and L**), suggestive of sequence and structure mediated effects on RNA–RNA interaction. Another common trend (except ORF7) is that plus-strand clusters in general have lower RNA–RNA interaction strength than the minus strands. This also might be inherent to plus strands, which need to be less prone to form aggregates for a functioning virus replication cycle. Furthermore, as mentioned above, the plus strands are more structured than their minus strand counterparts (**Figure 1E**). This suggests that plus strands are overall less free to base-pair, which is consistent with their lower RNA–RNA interaction strength in our simulations (**Figure 3F**). Interestingly, the reactivity for ORF7 shows an opposite trend to the other sequences (**Figure 3C**), which is also mirrored by the simulations (**Supplemental Figure S3G**).

Next, we quantified the diffusivity of RNA strands in each of the systems to investigate if perfectly complementary RNA–RNA interactions lead to RNA clusters with reduced dynamics. Our analysis revealed that, on average, RNA strands are less dynamic in systems with perfectly complementary RNA–RNA interactions (**Figure 3G, Supplemental Figure S3H**). Notably, the diffusion exponent here is less than 1 for all simulated systems, indicating sub-diffusive behavior of RNAs without the presence of proteins. Therefore, N protein in protein–RNA-condensates maintains liquid-like properties by mediating extensive RNA pairing by its binding to RNA (**Figure 1H-J**) or by other mechanisms.

To gain microscopic insight into the trends in RNA–RNA interaction strengths and RNA strand dynamics, we quantified the degree of trans RNA contacts in our simulations (**Figure 3H**). Indeed, the 1:1 mixture with perfectly complementary RNA–RNA interactions have roughly two-fold greater trans RNA contacts than plus- or minus-strand only systems (**Figure 3H, Supplemental Figure S3I**). Additionally, the probability distribution of nucleotide pairs in trans RNA contacts is right-shifted for the systems with perfectly complementary RNA–RNA interactions (**Figure 3I, Supplemental Figure S3J**), indicative of extensive trans RNA contacts. Furthermore, plus- and minus-strand only systems reveal more randomized contact pattern at the nucleotide level (**Figure 3J**). In contrast, the 1:1 mixture exhibits a highly ordered nucleotide contact pattern, consistent with the perfectly complementary nature of the plus and minus strands. Thus, these results strongly suggest that the arrested networked condensates primarily arise from extensive trans RNA contacts rather than random RNA entanglement.

Together, our experiments and simulations suggest that perfectly complementary RNA–RNA interactions could lead to an arrested phenotype during viral replication.

Coating with N protein reduces trans RNA-RNA interactions

Given the strong impact duplex RNA has on arresting condensates, how do viruses resolve the RNA duplexes that are an unavoidable step in the replication cycle of RNA viruses (in the case of SARS-CoV-2 >29,000 nucleotides long)? We hypothesized that separation in timing between protein binding and replication of the genome may be a mechanism to prevent the potentially deleterious duplexes. In the case of SARS-CoV-2, one complete plus strand copy of the genome enters the cell which could be bound by the ~1000 N proteins which also enter with the virion[48]. From this single plus strand RNA, a minus strand is synthesized (**Figure 4A**). Thus, binding/condensation with N protein or other RNA binding proteins might prevent the sticking of the plus strand RNA to the minus strand RNA pre-coating plus strand RNA with protein. We reasoned we could mimic this experimentally by altered timing of addition of RNA to protein.

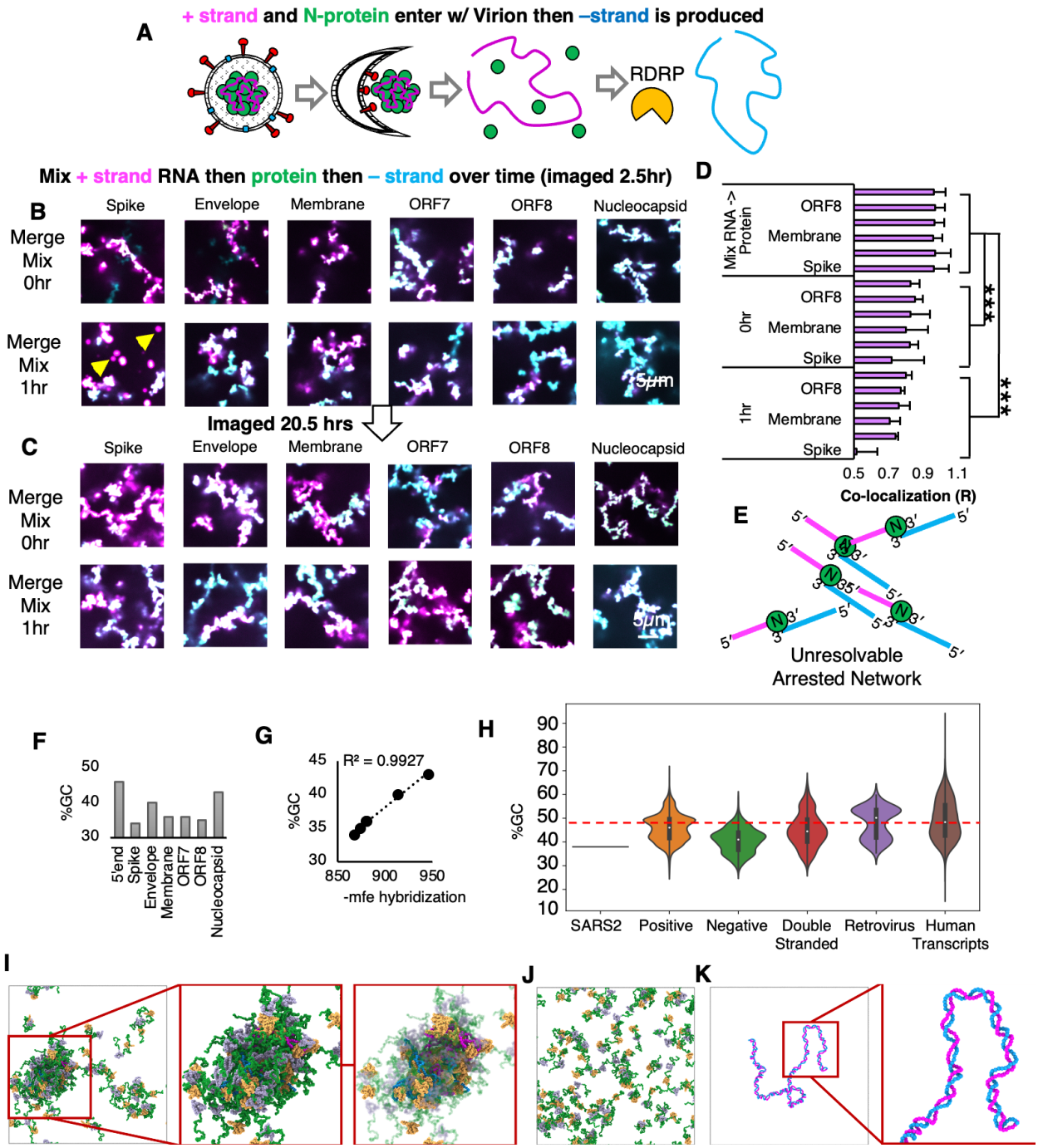


Figure 4. RNA addition timing and GC content modulate RNA-RNA interactions

(**A**) Cartoon depicts the genome replication cycle of SARS-CoV-2, applicable to many plus strand RNA viruses. Plus stranded RNA and N protein precedes minus strand RNA production. Plus strand RNA (pink squiggle) enters the cell, bound by N protein (green dots) in ribonucleoprotein (RNP) complexes (30-35 per virion for SARS-CoV-2). Upon entry into the cytoplasm, N protein uncoating (likely by ribosome read through) allows for the translation of multiple proteins including the RNA-dependent RNA polymerase (RDRP orange Pacman). RDRP uses the plus strand RNA genome as a template to produce the minus strand RNA (teal squiggle). For **Figure 4B** and **C**, all images depict the merged signal of the same total RNA (100nM) and N protein (3.2μM)

concentrations. As shown in panel **Figure 3A** combinations of two RNAs are 50nM each. Scale bar (white line in H_2O only control) is 5 microns. All images in a given panel are acquired on the same day and are contrasted the same. **(B)** Staggering the addition of minus RNA to plus RNA and protein by 0, and 1 hour prevents complete mixing of plus (magenta) and minus (cyan) RNA signals. Of note at this timepoint for 1 hour post preincubation, some amount of plus Spike RNA is still in spherical assemblies (magenta circles as marked by yellow arrows). **(C)** Related to **B** additional incubation following preincubation does not lead to increased co-localization in the RNA channel (plus magenta minus cyan mixing for any tested RNA combination. Quantification in **Supplemental Figure S4C**.) **(D)** Quantification of the reduced co-localization of plus and minus RNA signals following preincubation of plus RNA (**Figure 4A**) with N protein as compared to pre-mixing of RNA on ice (**Figure 3A**). In all tested RNA combinations, for all 3 technical replicates of each condition (error bars), preincubation results in a reduction of colocalization (R) whereas premixing, R values are close to 1 indicating near perfect co-localization of plus and minus RNA signals. **(E)** Model of the arrangement of plus and minus strand RNA molecules with N protein condensates which could result in an unresolvable arrested network morphology. **(F)** GC content of tested TRS-B fragments with Spike TRS-Bs having the lowest %GC and Nucleocapsid having the highest. (Of note plus and minus strand RNA of the same type e.g. spike has identical GC content). **(G)** %GC content is correlated with -mfe of RNA–RNA interaction strength, indicating that the strength of RNA–RNA Interaction is proportional to the GC content with plus/minus spike being the least “sticky” and plus/minus nucleocapsid being the most “sticky”. Consistent with this, predicted spike RNA has the least propensity to be mixed at 2.5 hours panel **4C** and **4D**. **(H)** Genomes of RNA viruses are depleted in GC content compared to the human transcripts. SARS-CoV-2 has very low fraction of GC content compared to human transcripts and plus strand RNA viruses in general. Negative stranded RNA viruses have the lowest GC content on average. Red dotted line depicts 50% GC content.) **(I)** Coarse-grained simulation of an N protein–RNA system (64 N protein; one plus and one minus Envelope RNA). Simulations are performed using model 1 (see Methods). Zoomed in snapshots of the protein–RNA cluster are shown. For clarity the transparency of portions of proteins in the rightmost snapshot are enhanced to reveal the protein coating on RNA. **(J)** Coarse-grained simulation (model 1) of a pure N protein system (no RNA). Without RNA, N protein does not form clusters. **(K)** Coarse-grained simulation of one plus and one minus Envelope RNA strand, using model 1. In this case, two RNA strands form a perfectly complementary dimer that is sustained over a one microsecond simulation.

To test the hypothesis that timing of minus strand RNA addition relative to protein might mitigate strong RNA–RNA interactions we altered the timing of addition of RNA with plus RNA, N protein, and minus RNA—representative of the proposed pre-binding of N protein to the plus strand prior to the production of minus by the RNA-dependent RNA of polymerase. To this end, we first mixed N protein to plus strand RNA and then staggered the addition of minus strand to this mix by 0, and 1 hours (**Figure 4B**). We assessed the ability of N protein to mitigate RNA-RNA entanglement by measuring the degree of colocalization[49] of the plus and minus strand RNAs in the same pixel (**Figure 4D**). The presence of protein prior to the plus and minus strands encountering one another was sufficient to prevent the plus and minus RNAs from mixing well for all the tested RNA sequences (**Figure 4B-D**). Remarkably even the almost immediate addition (0 hour) of minus RNA did not mix well with plus RNA that has been premixed with N protein, even after 2.5 hours of incubation of all three components together. We quantified the difference between premixing RNA on ice (**Figure 3A**) and staggering minus strand addition after plus/N protein mixing by measuring the co-localization (R) of the plus and the minus strand RNA signals (**Figure 4D**). Diffuse phase RNA measurement suggested that the dense phase recruited different levels of RNA and protein depending on staggering consistent with longer incubation with the higher condensate promoting sense RNAs better able to recruit RNA to the dense phase (**Supplemental Figure S4A**). We observed that for all tested RNA combinations staggering the

568 minus strand RNA addition resulted in a statistically significant reduction in RNA co-localization
569 indicating the prevention of complete mixing between the opposed RNA strands (**Figure 4D**).
570 Many condensates display liquid-like material properties[50–58]. If these properties were
571 present in our arrested networks, we would predict that additional incubation would allow for the
572 complete or more complete mixing of RNA, with continued failure to mix indicative of more solid
573 or aggregate-like material properties[59]. To determine if continued incubation would allow for
574 additional mixing, we again imaged the condensates 18 hours later (approximately 20.5 hours
575 from the mixing of plus RNA and N protein). We observed that the plus and minus strand RNA
576 signals continued to be not co-localized even 18 hours later (**Figure 4C**) with no significant
577 difference observed between 2.5-hour observation timepoint and 20.5 hours observation
578 timepoint (**Supplemental Figure S4B and C**).
579 The lack of complete mixing of plus and minus strands suggests that the coating of RNA
580 by N protein, mediated by favorable N protein–RNA interactions, aids in suppressing trans RNA
581 contacts. Interestingly, recent work has predicted a strong interaction motif between N protein
582 and RNA sequences—with the NTD and RBD 1 of N protein forming a positive groove that binds
583 RNA[60]. Thus, we also hypothesized that such an interaction motif could out-compete RNA–RNA
584 binding, effectively reducing sticking of RNAs to one another. To probe this at submolecular
585 resolutions, we simulated systems composed of 64 N proteins and two complementary RNA
586 strands (Envelope), which represents a system with excess N proteins, which is consistent with
587 the ratio of protein-to-RNA in our in vitro experiments. We then assess the formation of protein–
588 RNA clusters in our simulations. Here a cluster is defined using a distance-based metric (see
589 Methods). There are about 37 N proteins and 2 RNAs in the formed RNA–protein clusters, which
590 is a charge balanced case. Interestingly, in early viral infection, it is estimated that approximately
591 1,000 N proteins colocalize with a 30-kb RNA strand[48], which is consistent the charge ratio in our
592 predicted protein–RNA cluster. Surprisingly, despite being in proximity, the two RNA strands form
593 no stable trans RNA contacts due to excessive N protein coating (**Figure 4I**). Multiple RBD
594 domains, along with regions of N protein, can coat the same RNA strand, thereby reducing the
595 probability of extensive base pairing. In **Figure 4I**, the RBD 1 coating is easier to detect by making
596 the remaining domains of the proteins semi-transparent. As a reference, in these simulation
597 conditions N protein is not able to form clusters without RNA (**Figure 4J**), and the two
598 complementary RNA strands are able to sustain extensive base pairing in the absence of N
599 protein (**Figure 4K**). Collectively, these simulations support the model that the coating of plus
600 strand RNA by N protein can prevent mixing of the plus and minus strand RNAs. However,
601 exposed non-coated regions in plus might still be able to base pair with perfectly complementary
602 minus strand RNA resulting in the arrested networks observed experimentally (**Figure 4E**). Thus,
603 additional strategies other than coating of plus RNA by N protein prior to minus strand RNA
604 production must be employed to mitigate spurious and potentially strong complementary RNA
605 interactions between the plus and minus strands.
606 Our simulations and wet lab experiments revealed that coating RNA in N protein reduced
607 spurious interactions and co-localization. We next asked if there were any RNA sequence specific
608 differences in co-localization in our 6 tested combinations. Condensates with longer plus strand
609 only incubation time recruiting more protein to the dense phase (**Supplemental Figure S4A**),
610 consistent with the higher propensity of plus strand RNA for driving N protein condensation but
611 there was no obvious consistent trends with plus strand RNA alone condensates **(Figure 1F and**
612 **G)** suggesting plus strand affinity to N protein did not play a significant factor in suppression of
613 RNA-RNA interactions by pre-coating with N-protein. Examination of the RNA signal reveal that
614 RNA signals failed to mix upon staggered addition, and the degree of colocalization was not
615 uniform with the lowest degree of mixing as quantified by plus/minus RNA low colocalization score
616 observed in Spike TRS-B fragments (**Figure 4D**), (**Supplemental Figure S4D**). Further,
617 colocalization of RNA signal was not altered with longer incubations times (2 vs. 20hr) suggesting
618 that once arrested networks form they are irresolvable under tested parameters (**Supplemental**

619 **Figure 4C**) Additionally, when the normalized colocalization signal was compared for the 0-, and
620 1-hour conditions, only Spike was significantly less mixed (**Supplemental Figure S4D**) following
621 incubation time and at the 2.5-hour timepoint in **Figure 4B** Spike still had rounded plus strand
622 only condensates as indicated by yellow arrows. We thus reasoned that there must be some
623 difference in the affinity of plus and minus strand RNA duplexes for Spike as compared to the
624 others and Spike might also be the least "sticky" duplex. Of note, we previously observed that
625 SARS-CoV-2 had higher GC content at viral ends (e.g. Nucleocapsid RNA) then middles (e.g.
626 Spike RNA)[21]. This arrangement was not consistently observed in other viruses regardless of type
627 (**Supplemental Figure S4E**). Thus, our results suggested that a comparison of Spike RNA to
628 other tested RNA features may provide insight into the sequence specificity controlling trans RNA-
629 RNA interaction.
630 To assess if Spike plus/minus RNA is different from other tested RNA duplexes we began
631 with an examination of the primary sequence content. We observed that of our tested TRS-B
632 fragments, Spike had the lowest GC content (**Figure 4F**) and further, the degree of predicted -
633 MFE energy of hybridization[45] was perfectly correlated with the GC content suggesting this is the
634 least sticky RNA sequence (**Figure 4G**). Similar results were also observed in RNA only
635 simulations in **Supplemental Figure S3**.
636 We reasoned if depletion in GC content was an RNA viral strategy for reducing RNA–RNA
637 interactions then depletion in GC content should be a near universal feature of RNA viral
638 genomes. To this end, we computed the GC content by RNA virus type and, consistent with
639 previous publications[61], we observed that RNA viruses tended to have lower GC content than that
640 of the human transcriptome (**Figure 4H**). Intriguingly, the RNA virus type with the lowest GC
641 content was negative strand RNA virus which must first replicate a plus strand copy prior to protein
642 production. Negative strand viruses must arguably have the greatest problem with stickiness as
643 no translation occurs prior to plus strand genome synthesis[62] event to potentially prevent RNA–
644 RNA interaction. We examined the patterning of the GC content to see how common GC content
645 depletion was in RNA viruses. Intriguingly, only in a few viral orders including nidoviruses was the
646 GC content in the at the end of the plus strand greater than that at the middle; the majority of
647 viruses displayed the opposite configuration (**Supplemental Figure S4E**). These data collectively
648 suggest that nidovirus genomes are particularly depleted in GC content at the middle and that
649 this may help mitigate RNA–RNA interaction between the plus and minus strand genomes during
650 replication.
651 Our data from **Figure 3 and 4** suggest that RNA viral sequences may be under selective
652 pressure to maintain low GC content to mitigate RNA–RNA interactions between plus and minus
653 strand genomes during replication. Thus, we reasoned that as viruses may be under selective
654 pressure to maintain low GC content this should lead to a commensurate decrease in the
655 propensity to form g-quadruplexes. To examine this possibility further we predicted g-
656 quadruplexes sequences[63] in the SARS-CoV-2 genome. In keeping with previous observations
657 of acute RNA viruses[64], we observed that SARS-CoV-2 is overall has both few and low-quality
658 predicted g-quadruplex sequences[65] (**Supplemental Figure S4F, G and H**) with no predicted
659 sequence remotely approximating the score of a bona fide RNA g-quadruplex, Orn-1[66].
660 Additionally, only one tested RNA fragment, the 5′end plus strand, had a predicted g-quadruplex
661 sequence (**Supplemental Figure S4G and S4H**). Addition of non-physiological excessive[66]
662 magnesium ion concentrations (10mM) had no significant impact on the degree of co-localization
663 under conditions of staggered minus strand RNA addition. (**Supplemental Figure S4I and S4J**).
664 Thus, we concluded that g-quadruplexes do not contribute significantly either to the arrested
665 morphology of the condensates or the degree of RNA–RNA interactions.
666
667 ***Control of RNA-RNA interactions by altering the ratio of plus and minus strand RNA***

668 Although low-GC content was correlated with a reduction in mixing of RNAs and a
669 reduction in RNA–RNA interaction strength, low GC content and N protein coating are in
670 combination still insufficient to block the formation of arrested networks (**Figure 3 and 4**). We
671 wondered if there might be other strategies employed by RNA viruses to limit the spurious
672 interaction between the plus and minus strand RNA genomes during viral replication. Plus-
673 strand RNA viruses like SARS-CoV-2 begin with a single complete plus strand genome carried
674 by the virion[48]. From this plus strand RNA, at least one minus strand template is produced[67].
675 Therefore, at early times in infection, plus and minus RNA are perhaps present at equal molar
676 ratios, however this ratio does not persist over the duration of infection as the virus uses the
677 minus strand RNA to produce numerous copies of plus strand RNA. Thus, as infection
678 progresses, a higher proportion of plus strand is produced relative to minus strand. We
679 hypothesized if the change in this ratio might also mitigate the strong RNA–RNA interactions by
680 effectively spreading plus/minus RNA–RNA interactions over an increasing higher number of
681 plus strand molecules such that eventually the contributions of the minus strand become
682 effectively saturated.
683 To test the hypothesis that the ratio of plus/minus RNA strands can influence the
684 morphology and arrest of condensates, we produced condensates with 1:1 plus/minus, 10:1
685 plus/minus and 100:1 plus/minus RNAs. Of note, in all cases 100nM of total RNA and 3.2μM of
686 protein were used. We observed that in all tested TRS-B fragments, higher ratios of plus strand
687 RNA relative to minus resulted in a suppression of the arrested network formation with a 10:1
688 partially and 100:1 almost completely suppressing the amorphous shape/ lack of circularity
689 observed when 1:1 ratio are present (**Figure 5A**).

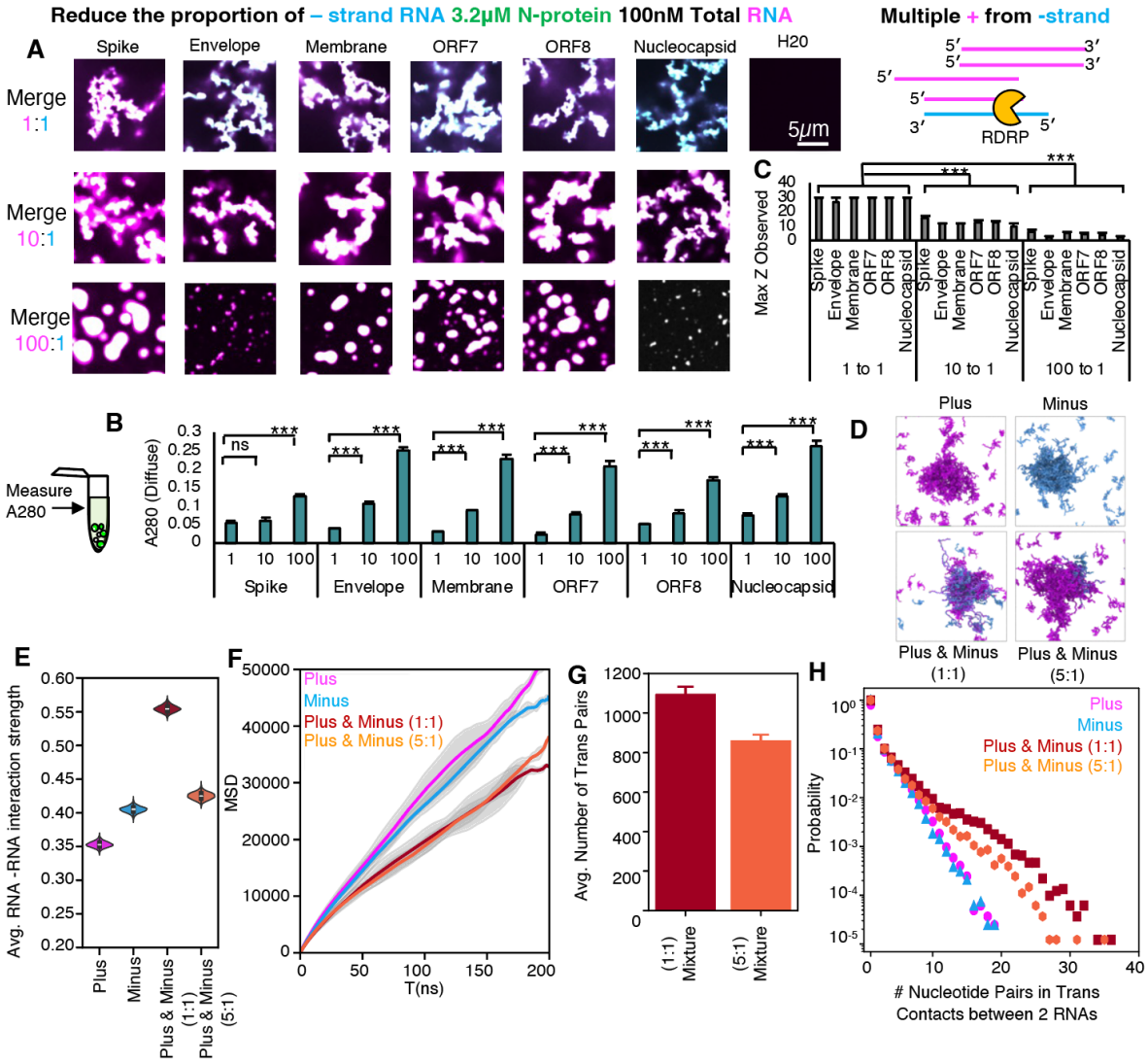


Figure 5. Imbalance in complementary RNA sequence concentrations present in late-stage viral replication attenuates RNA-RNA interactions

(A) Cartoon depicts the RDRP (orange Pacman) synthesizing multiple plus strand RNA genome copies from a single minus strand RNA indicating plus and minus strand RNA genome ratios are not uniform over the course of infection. 1:1 ratio mix (early infection) of plus and minus strand RNA result in arrested networks and 100 plus (magenta) to minus (cyan), consistent with later in infection, suppress arrested network formation resulting in more liquid-like condensates. 10:1 plus to minus result in intermediate phenotypes. **(B)** Diffuse phase RNA measurements related to panel A. Higher plus strand to minus strand RNA ratio results in a commensurate increase in RNA and protein signal (A280) absorbance in the diffuse phase indicating the lower ratios are better able to drive condensation. Error bars are technical replicates. NS not significant (p<0.001 ***) **(C)** Related to **Supplemental Figure S5**. Observed Z height for condensates formed in A. 1:1 ratios of plus and minus RNA extend at least 30 microns into Z (arrested networks) whereas 100:1 extend only a few microns. 10:1 is intermediate. Error bars are 3 technical replicates. (p<0.001 ***). **(D)** Comparison of 2X plus, 2X minus, 1:1 plus/minus, and 5:1 plus minus in RNA only simulations of envelope RNA (model 2). **(E)** Average RNA-RNA interaction strength for systems depicted in (**D**). Consistent with experiments, 1:1 plus/minus

(maroon) results in a stronger interaction (i.e., most negative) than the 5:1 (orange) system. **(F)** 5:1 plus/minus simulated RNA system (orange) shows intermediate MSD between 2X plus (magenta) or 2X minus (cyan) and 1:1 plus/minus (maroon). **(G)** Reduction in trans RNA–RNA pairs between 5:1 plus/minus and 1:1 plus/minus in RNA simulations. **(H)** Probability of long trans RNA-RNA interactions decreases in 5:1 as compared to 1:1.

Consistent with the morphology change, the plus/minus RNA ratio alters the amount of RNA in the dilute phase (**Figure 5B**), with excess plus strand associated with more RNA in dilute phase, despite total RNA in the reactions being the same. To quantify the degree of network formation, we measured the height of a continuous condensate from the cover glass up into solution. We found in 1:1 mixes that the arrested network extends from the glass up to 30 microns into Z **(Figure 5C and Supplemental Figure S5A, S5B, and S5C**) with 10:1 extending an intermediate height into solution as compared to 1:1 and 100:1 indicating a suppression of arrested network formation with increasing plus/minus RNA ratios. We were curious as to how changing RNA ratio altered the arrangement of molecules in the dense phase. To characterize the impact of the mixing ratio on the strength of RNA–RNA interactions and RNA strand dynamics, at the nucleotide level, we simulated a system composed of a 5:1 plus/minus strand ratio (via model 2, see Methods, **Figure 5D**). Indeed, an excess of plus strands yields reduced RNA–RNA interaction strength (**Figure 5E**) and exhibits slightly faster dynamics on long timescales (**Figure 5F)** compared to the 1:1 plus/minus case. These effects can be attributed to a significant reduction in trans RNA contacts when plus strands are in excess (**Figure 5G and H**). These data collectively indicate that the degree of RNA–RNA interactions contribute to arrested network formation and that altering the ratio of plus and minus RNA may be an additional strategy used by RNA viruses to mitigate complementarity-induced arrested network formation.

Collectively, our data thus far suggest that perfect complementarity between the plus and the minus strand leads to arrested network formation proportional to the RNA–RNA interaction strength as governed by N protein interactions (**Figure 4**), RNA sequence and in particular GC content (**Figure 4**) and the ratio of RNAs present (**Figure 5**). These combined features serve to minimize entanglements by reducing hydrogen bonding between complementary RNAs and spreading of RNA interactions over increased number of plus strand RNA molecules relative to minus. We wondered if other strategies might be employed to reduce complementarity perhaps via a change in RNA sequence for one of the pairs of RNA as each of the above strategies was not completely sufficient for preventing arrested condensate morphology.

Reducing RNA-RNA interaction strength with mutations or staggering RNA overlap reduces arrested network formation

As RNA viruses replicate, they accumulate mutations in their genomes due to the high error rate of viral RNA dependent RNA polymerases (e.g. average of 2.68×10^{-5} de novo errors per cycle with a C > T bias for SARS-CoV-2 RdRp[68]) and the action of RNA modifications such as inosine (ADAR[69]), methylation (METTL3/14)[69], and 2'OME[70] (NSP16)[71]. Notably, all these modifications have been detected on SARS-CoV-2 viral RNA. These mutations/modifications may offer a selective advantage to RNA viruses by increasing the diversity of the genome allowing for escape from immune surveillance. In keeping with this theory, most recent pandemics were caused by RNA viruses[72] with the single most recent DNA virus pandemic, Mpox, arising due to reduced frequency of individuals vaccinated for smallpox in the population[73]. Thus, increasing mutational burden represents a positive effect for RNA viruses with a tradeoff between increased diversity and continued functionality of protein and RNA elements but potentially a secondary benefit of reducing complementarity between the plus and minus strands of the RNA genome.

759 Since RNA viruses accumulate multiple mutations/modifications during replication, we
760 wondered if these changes might also be a strategy employed by RNA viruses to reduce RNA–
761 RNA interactions between the plus and the minus strand RNAs. Thus, we sought to ask, can
762 mutations on minus RNA strand reduce complementarity to the plus RNA strand and if so, how
763 many mutations are required?
764 To this end, we chose to reduce complementarity to the plus strand through
765 modifications of minus strand RNA in two ways, **(1)** increased inosine content and **(2)** increased
766 mutation in the template sequence. We chose the minus strand rather than the plus as this RNA
767 had less propensity to drive condensation with N protein on its own (**Figure 1**).
768
769

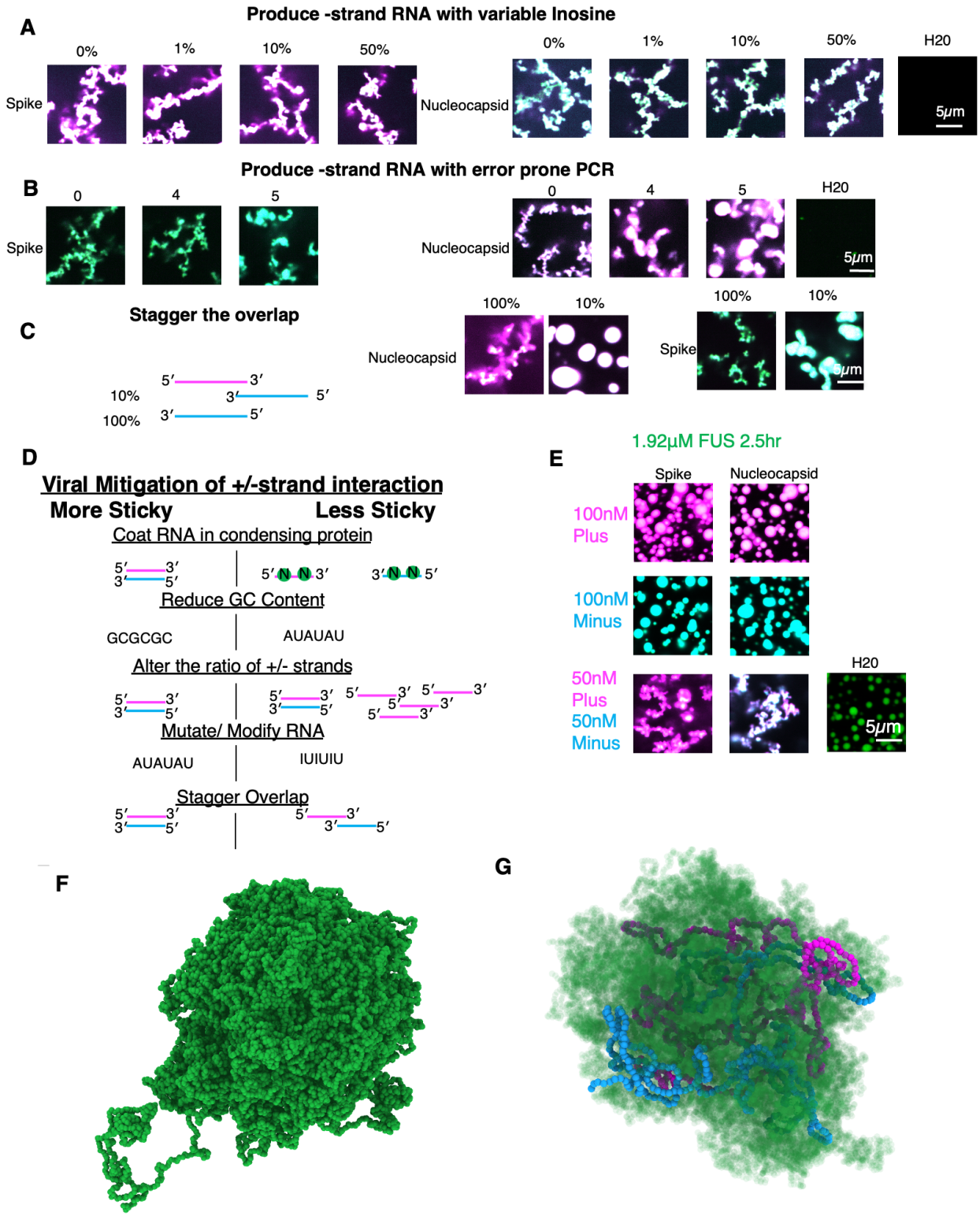


Figure 6. Protein sequence independent role for RNA–RNA interactions in driving biomolecular condensate morphology.

(A) No obvious effect of increasing inosine concentration on arrested network morphology. Minus strand RNA (cyan) was produced with 0, 1, 10, and 50% inosine to replace increasing amounts of A nucleotides with I. Combinations of two RNAs are 50nM each. Scale bar (white line in H2O

water only control) is 5 microns. All images in each panel are acquired on the same day and are contrasted the same. **(B)** 4-5 rounds of error prone PCR are sufficient to reduce the arrested network formation. Minus strand RNA (cyan) was produced with DNA templates made from 0 (I proof) or multiple rounds of error prone PCR. 1µL of error prone PCR reaction one was used as template for error prone PCR 2. Of note cycles 1-3 were also tested and resulted in no obvious change in arrested network formation. **(C)** Reducing the overlap of the plus and minus RNAs by shifting the position of the minus strand RNA template results in condensates with more circular morphology. **(D)** Summary of multiple viral strategies used to avoid extensive RNA–RNA interactions. **(E)** FUS protein also yields arrested networks when perfectly complementary plus and minus strand RNA is present. 100nM of plus (magenta, top row) and minus (cyan, middle row) strand RNA from Spike and Nucleocapsid TRS-Bs yielded rounded co-condensates with FUS protein (green). 50nM each plus (magenta) and minus (cyan) yield arrested networks (bottom panel) with FUS. All images depict the merged signal of the same total RNA (100nM) and N protein (3.2µM) concentrations. Scale bar (white line in H2O water only control) is 5 microns. All images in each panel are acquired on the same day and are contrasted the same. Simulations of FUS and RNA. **(F)** Simulation of full 64xFUS proteins without RNA. The proteins can form condensates. **(G)** Simulations of 64xFUS proteins (semitransparent green) with two Envelope RNA plus (magenta) and minus (blue) strands. The FUS proteins can colocalize with the protein.

We observed that increasing amounts of inosine (**Figure 6A**) did little to mitigate the arrested network morphology of the condensates but, following 4-5 rounds of error prone PCR, arrested network formation was partially mitigated (**Figure 6B**). Of note, moving forward we limited our experiments to two RNA pairs as Spike and Nucleocapsid RNA TRS-Bs represent the least and most “sticky” RNA combinations, respectively as measured by GC content. (**Figure 4F**) Thus, these results should be representative of the possible RNA conditions for any perfectly complementary RNA pair.

Our data thus far revealed that the presence of perfectly complementary stretches of 480 nucleotides results in arrested network configurations in condensates. As such long stretches of complementarity are probably actively avoided by host cells by careful organization of the genome (e.g., splicing, non-overlapping genes in Watson and Crick DNA sequences, staggered expression timing, RNA export etc.),we reasoned it would be informative to reduce the complementarity by staggering the overlap of the sequence to be 10 percent of the previous conditions (48 nucleotides of perfect complementarity sequences). We did this by altering the positioning of the template for the minus strand in the genome shifting it 3 prime to the plus strand RNA template. We observed that the 10 percent overlap condition almost completely suppressed the arrested network phenotype, suggesting once again that the degree of RNA–RNA interaction strength is proportional to the arrested morphology (**Figure 6C**). In keeping with this finding, we would expect that producing sub-genomic RNA fragments[34] (sgRNA), a strategy employed by betacoronavirus might be yet another strategy to reduce interaction between plus and minus strand RNA sequences as these shorter sequences would have less propensity to base pair than full-length genomes. Thus, multiple strategies to mitigate RNA–RNA interactions between perfectly complementary RNA sequences are accessible to viruses (**Figure 6D**).

Protein content is dispensable for morphology regulation by extensive RNA-RNA interaction

Given that 48 nucleotides had only modest degree of conglomeration, we reasoned that in any given eukaryotic cell, perfect sense and antisense RNA should only be problematic if the number of identical nucleotides is greater than 48 nucleotides, particularly when both sequences are simultaneously transcribed in the same nucleus. Viruses replicate not only in the presence of their own proteins but in the milieu of many host factors including an array of host-RBPs.

827 As Fused in sarcoma (FUS) protein is the best characterized condensing protein which
828 can bind RNA sequences with high nuclear protein concentrations[74], we reasoned this would be
829 an ideal candidate to test whether strong RNA–RNA interactions drive arrested network
830 formation regardless of the identity of the co-condensing protein. To this end we tested our two
831 representative RNA duplexes plus/minus Spike and Nucleocapsid RNA TRS-B and plus/minus
832 with FUS protein rather than N protein and observed that, in keeping with the results for N
833 protein, FUS was able to make round condensates with either plus or minus Spike or
834 Nucleocapsid RNA but formed arrested networks with the combination of the plus and minus
835 RNA (**Figure 6E**). These data suggest that RNA–RNA interaction network duplex formation is
836 not dependent on condensing protein identity and thus should be a universal feature of
837 biomolecular condensates. These data suggest condensate morphology is due to RNA–RNA
838 rather than RNA–protein or RNA–N protein interactions. Simulations of FUS protein without
839 (**Figure 6F**) or with RNA (**Figure 6G**) further illustrated that FUS protein could reduce RNA–
840 RNA interactions when complementary sequences were present.
841
842 ***Role of cytoplasmic proteins in modulating RNA-RNA interactions***
843
844 Lastly, we wanted to know if the observation we saw cell free also held true in cells which
845 contain many other host proteins (including FUS) and RNAs. To this end, we introduced N protein
846 and our RNA duplexes into the cytoplasm of Vero E6 cells utilizing acid washed beads[75] to deform
847 the membrane allowing for the penetration of RNA and protein (**Figure 7A**).

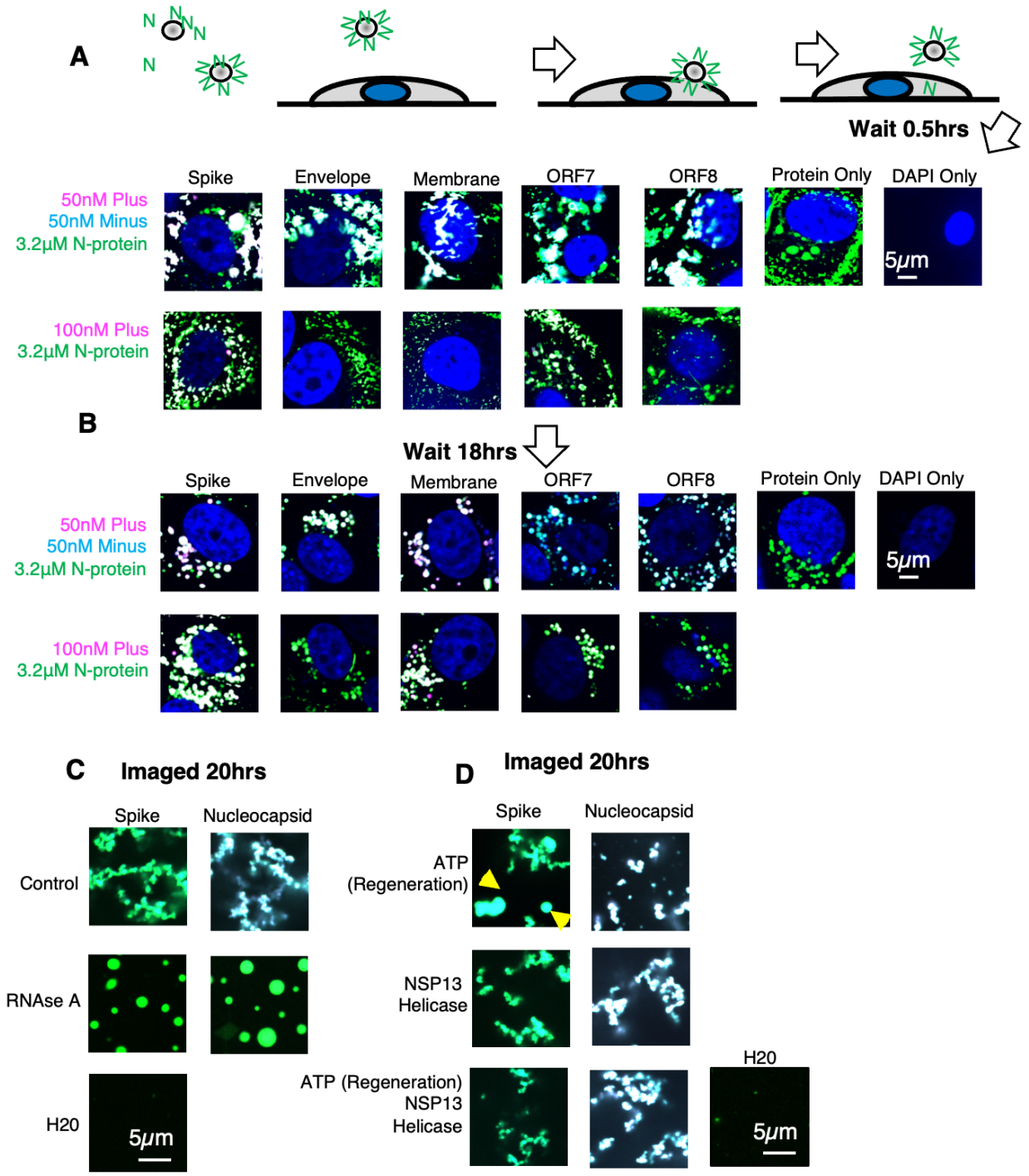


Figure 7. Cytoplasmic RNAses and Helicases may help to resolve arrested networks. (**A**) Cartoon depicts acid washed bead-based strategy for the delivery of RNA and protein to the cytoplasm of cultured, adherent, mammalian cells. Beads are incubated with protein (green Ns) and RNA solution in physiological buffer. They are then rolled over the membrane of adherent mammalian cells. Contact between the bead and the membrane surface disrupts the membrane such that protein and RNA can enter. Resulting protein is deposited in the cytoplasm (grey semi-circle) not the nucleus (blue circle). Representative images showing the merged signal (plus RNA in magenta, minus RNA in cyan, N protein in green, nucleus in blue). From 5 tested

857 RNA duplexes (top panel) or just plus strand RNA (bottom panel). Imaged at 0.5 hours after
858 bead treatment. All duplexes tested result in arrested networks (top panel) whereas plus alone
859 does not (bottom panel). **(B)** Representative images showing the merged signal (plus RNA in
860 magenta, minus RNA in cyan, N protein in green, nucleus in blue). From 5 tested RNA duplexes
861 (top panel) or just plus strand RNA (bottom panel). Imaged at 18.5 hours after bead treatment.
862 Compared to the 0.5-hour timepoint (A) condensates appear to relax. **(C)** Treatment with
863 RNAse A resolves arrested networks. Representative images showing the merged signal (plus
864 RNA in magenta, minus RNA in cyan, N protein in green), **(D)** Presence of ATP in the absence
865 of SARS-CoV-2 RNA helicase NSP13 results in partial mitigation of arrested network formation
866 for spike (yellow arrows).
867
868 We observed that 1 hour following incubation with beads, protein and RNA, all tested
869 RNA duplexes had observable arrested networks formed in the cytoplasm of cells containing
870 both protein and RNA signals. Of note, we did not test Nucleocapsid RNA to avoid any potential
871 contributions from the additional protein translation (our protein was produced with the native
872 SARS-CoV-2 sequence and therefore may contain small amounts of contaminating plus RNA
873 sequence from bacterial production of Nucleocapsid RNA and in cells this could be translated to
874 additional N protein fragments). In comparison, addition of only plus strand RNA and N protein
875 resulted in more rounded condensates. The cytoplasm of eukaryotic cells contains numerous
876 RNA transcripts, ribosomes, and proteins (RNA-binding or otherwise). To address if incubation
877 in the cytoplasmic milieu could resolve the arrested network we again imaged cells 18 hours
878 later (**Figure 7B**). We observed that, following 18-hour incubation, N protein and RNA signal no
879 longer resembled an arrested network but rather more rounded condensates. Similar rounded
880 condensates were observed in plus strand RNA only conditions. Collectively these data
881 suggested that one or more cellular factors may be acting on the RNA and protein arrested
882 networks at long time scales to resolve them.
883 To address what mechanisms could be used by the cell to resolve an arrested network
884 of RNA–RNA interactions we returned to our cell free system. As double stranded RNA present
885 in our RNA duplexes of a viral origin should be an excellent target for the host native immune
886 system[76], we first reasoned that one pathway for resolving arrested networks may be to simply
887 degrade the RNA through the action of endogenous RNAases such as the interferon induced
888 RNAse L etc[77]. In SARS-CoV-2 infections, this activity could be provided by NSP15 which
889 preferentially degrades AU rich dsRNA[78]. To test this possibility, we added RNAse A to our
890 preincubated duplexes and N protein (0.5 hours) (**Figure 7C**). We observed that addition of
891 RNAse A was sufficient to completely abolish the arrested network phenotype, indicating that an
892 additional way that cells may resolve an arrested network is to degrade the RNA holding it
893 together via its extensive network of RNA–RNA interactions.
894 We reasoned another likely candidate protein family that could act to resolve the
895 arrested networks was "helicases" which are thought to function by unwinding RNA duplexes in
896 some cases[79]. Contained within the genome of SARS-CoV-2 is the RNA helicase NSP13[80]. This
897 protein is essential for viral RNA replication. Although not present in uninfected Vero cells, we
898 reasoned that this protein was the most likely candidate for resolving SARS-CoV-2 RNA–RNA
899 interactions during infection. To test if NSP13 could resolve an RNA duplex of 480 nucleotides
900 long (far longer than any duplex tested for any helicase activity to our knowledge), we incubated
901 our Spike and Nucleocapsid TRS-B plus/minus RNA duplexes with N protein and NSP13. As
902 NSP13 activity is dependent on ATP and ATP is a reported hydrotrope[81] which can promote
903 solubilization of biomolecular condensates we also included +/- ATP/creatine kinase ATP
904 regeneration as controls (**Figure 7D**). Intriguingly, NSP13 alone had no noticeable effect on the
905 arrested network formation nor did NSP13 with the addition of ATP. However, the addition of
906 ATP alone did have a modest effect on the morphology of Spike TRS-B plus/minus arrested
907 networks in keeping with the published observation of ATP's hydrotrope activity. This effect of

908 ATP was not readily observed in Nucleocapsid TRS-B RNA conditions consistent with Spike
909 RNA being the "least sticky" with respect to GC content (**Figure 4F**). We interpret these results
910 as indicating that ATP's hydrotrope activity can help limit weaker RNA–RNA arrested network
911 formation only if it has not been consumed by NSP13 which has reported tremendous ATPase
912 activity[82]. Thus, high intracellular ATP may help solubilize arrested networks but once a 480-
913 nucleotide duplex is formed it cannot be resolved under tested conditions by NSP13.
914

915 ## DISCUSSION

916 Beyond storing information, RNA provides an essential architectural role in the formation
917 of biomolecular condensations[24]. Viruses risk RNA-RNA interactions that could lead to potentially
918 deleterious states, these are the strategies we have uncovered that viruses can use to avoid the
919 risks inherit in complementary RNA by (1) pre-coating RNA genomes in condensing protein, (2)
920 reducing genome GC content, (3) altering the ratio of plus and minus strands, and (4) Mutating
921 or modifying RNA. This concept likely extends to all eukaryotic condensates which can
922 additionally (5) tune native RNA-RNA interactions in condensates by stagging overlapping gene
923 sequences in DNA. Thus, in our manuscript, we present a likely universal feature of RNA
924 containing biomolecular condensates, uncovering how RNA–RNA interaction complementarity
925 confers the morphology of the resulting condensate. We uncovered this central axiom of
926 biomolecular condensates by reconstituting representative molecules from two essential SARS-
927 CoV-2 replication steps: subgenomic RNA generation and genome replication. We observed that
928 low affinity interactions which arise because of low sequence complementarity (e.g. pairing
929 between the plus TRS-L and minus TRS-B) yield rounded condensates whereas high affinity
930 interactions (e.g. pairing between the plus TRS-B and minus TRS-B) due to longer, stronger
931 complementarity RNA sequences yield arrested networks. Condensate morphology correlates
932 with experimental measurements of RNA–RNA interaction strength via direct sequencing of
933 reduced DEPC reactivity and via simulations.

934 We believe this feature of RNA and condensate biology was not yet discovered due to (1)
935 the overall reliance on the field on shorter and homotypic RNA polymers (2) the preference for
936 biochemical reconstitution experiments to be undertaken with a single RNA species for ease of
937 interpretation. These previously implemented conditions would not be predicted to yield arrested
938 networks unless RNA structure is extraordinarily low[37,38] or multiple RNAs are present as is the
939 case for reconstituted stress granules[83,84] from cellular RNA extract. Our work provides a unique
940 insight into this RNA dependent process as it explores the RNA parameters required to yield
941 entangled RNA. RNA–RNA interactions are present in all condensates but, according to our
942 model interesting changes to condensate morphology happen when a high degree of
943 complementarity is present. This in principle can occur when 2 or more RNAs are present (as with
944 our experiments in this paper) or when longer sequences of RNA are present (the longer the RNA
945 the more chances for complementarity). Our experimental and modeling data suggest that
946 arrested network morphology is due to contributions from trans RNA contacts and long runs of
947 RNA contact rather than protein–RNA or protein–protein interactions or random RNA
948 entanglement making this a universal feature of RNA containing biomolecular condensates. Our
949 simulation data suggests that these networks arise predominately from the daisy-chaining of
950 incomplete base pairing between the complementary RNA strands, favoring arrested networks
951 over perfect duplexes. It is likely that both RNA structure and protein binding also serve to coat
952 regions of RNA blocking perfect complementarity to promote arrested network formation.

953 Our results suggest that lowered GC content confers reduced RNA–RNA interaction
954 strength. RNA viral genomes are dramatically depleted in GC content. Of note, the most GC
955 depleted order of plus strand RNA viruses following nidovirales like SARS-CoV-2 is
956 Cryppavirales. This is intriguing as many constituent species from Cryppavirales have lost all

957 packaging machinery (e.g. capsid/nucleocapsid proteins)[85]. These viruses may instead rely solely
958 on reduced GC content to prevent sticking during replication. Further evidence of the selective
959 advantage of low GC content has been observed in RNA sequencing data collected over the
960 course of the pandemic. These data reveal that SARS-CoV-2 has accumulated numerous coding
961 and non-coding mutations over time. Intriguingly, the virus has undergone excessive C–U
962 transitions[86], likely mediated by host-cell RNA editing machinery, specifically mediated by
963 APOBEC proteins[87].

964 Our results suggest that SARS-CoV-2 and other viruses may benefit from the lower GC
965 content during replication with edited genomes providing a selective advantage to the viruses
966 during replication beyond the increase in genetic diversity by reducing extensive RNA–RNA
967 interactions. In keeping with this observation, we observed that addition of inosine, the product of
968 ADAR mediated editing) to minus RNA had limited impact on the propensity to form arrested
969 networks. We would speculate this is due several factors; **(1)** inosine chemically resembles
970 guanosine which can pair with both C and U[88], **(2)** the addition of inosine could disrupt minus
971 strand RNA structure, rendering RNA more single stranded and prone to base pairing with plus,
972 **(3)** error prone PCR results suggest that more than 60 mutations or 12.5 percent (50 percent of
973 average A content) of nucleotides need to be mutated before there is a noticeable impact on A
974 content. Of note, for our error prone PCR conditions, 5 cycles should result in approximately 200
975 mutations per RNA molecule, not confined to A nucleotides. This may explain the observed net
976 loss of C bases[86] as opposed to A bases as our results suggest that ADAR editing may increase
977 RNA–RNA interactions strength whereas APOBEC mediated C–U transitions may act to reduce
978 RNA-RNA interaction strength.

979 Viruses including coronaviruses actively maintain the structure of the RNA in defined
980 configurations when mutations arise through compensatory mutations to restore base pairing
981 (such as the replacement of a GC pairs with an AU or GU wobble)[89]. Viruses thus retain RNA
982 structure despite it being a trigger for innate immune surveillance. It is intriguing to speculate that
983 viruses employ RNA structure as one part of their multi-pronged strategy of mitigating extensive
984 RNA–RNA interactions, burying the complementary sequence in cis RNA structure to avoid
985 pairing in trans.

986 Other parts of the viral strategy to mitigate strong RNA–RNA interactions involve coating
987 RNA in condensing protein and altering the ratio of plus and minus strands and we have presented
988 evidence for both in reducing arrested network formation. We postulate that these strategies are
989 also likely applicable to non-plus stranded RNA viruses. Our results suggest that the replication
990 strategy employed by all RNA viruses to produce proportionally fewer template RNA molecules
991 relative to packaged RNAs may have an additional advantage beyond economy, as this strategy
992 limits the damaging effects of extensive RNA–RNA interactions, spreading perfect
993 complementarity over an increasingly large number of molecules. Additionally, our results suggest
994 that a reason for the enrichment of proteins predicted to be prone to condensation in viral
995 genomes[90,91], may be to prevent extensive RNA–RNA interactions during genome replication by
996 separately coating the genome and its template. Minus strand RNA viruses may rely more heavily
997 on genome condensation and condensing proteins to block extensive RNA-RNA interactions
998 whereas plus and double stranded RNA viruses may employ ribosome mediated translation to
999 separate their genomes. This may explain why minus strand RNA viruses are more likely to
1000 employ inclusion bodies (a biomolecular condensate formed by the virus during genome
1001 production) in their replication cycle[92].

1002 Our helicase experiments suggest that preformed genome-long duplexes may be difficult
1003 to resolve by even viral machinery. How then can the replication of double-stranded RNA viruses
1004 be explained, which contain perfect genome-long double-stranded RNA duplexes packaged in
1005 the virion? We believe the answer to this may be found in the packaging and genome organization

1006 strategies employed by double-stranded RNA viruses. Double stranded RNA virus genome length
1007 tends to be shorter and split over multiple molecules, both features which we would predict would
1008 reduce RNA-RNA interactions. As to packaging, dsRNA utilizes a unique packaging strategy as
1009 compared to plus and minus strand RNA viruses. Chronologically, plus strand RNA viruses are
1010 thought to have evolved first, followed by double-stranded RNA viruses which then give rise to
1011 minus strand RNA viruses[93]. In keeping with this model, double-stranded RNA viruses are initially
1012 packaged in a manner like plus strand RNA viruses but following packaging of single copies of
1013 plus strand RNA, complementary minus strand RNA is replicated in the virion by pre-loaded
1014 RDRP[94]. Thus, +/- strand genome entanglement may be limited by reducing the number of
1015 molecules, restricting minus strand synthesis to the virion, and by near simultaneous synthesis of
1016 all RNA duplexes.

1017 In cells and during infection helicases/RNAses as well as endogenous ATP may help to
1018 resolve duplexes. We will briefly speculate on the implications of these findings in the following
1019 sections. RNAses: We would speculate that endogenous and virally encoded (e.g. NSP15 in
1020 SARS-CoV-2[78]) RNAses may be a strategy of last resort to resolve RNA tangles with the negative
1021 consequence of degradation of RNA. In keeping with this theory, NSP15 is an essential protein
1022 for SARS-CoV-2 viral replication, and RNAseL mediated degradation can dissolve most RNA-
1023 dependent condensates in cells[77]. Helicases: Under our tested conditions, NSP13 was unable to
1024 resolve the arrested network of RNA. We do not preclude the possibility that under some
1025 conditions NSP13 could prevent or reduce extensive RNA-RNA interaction as suggested by its
1026 essential role in viral replication, rather our data indicate the possibility that there are some
1027 conditions which might be impossible to resolve. It is also possible that in our reconstitution
1028 experiments NSP13 is missing one or more required co-factor required for rapid processivity.

1029 Eukaryotic cells may partially avoid the problem of long RNA-RNA interactions in native
1030 RNA sequences by organizing their genomes in such a way as to avoid perfectly
1031 complementary stretches of RNA, but a nuclear protein could further help to limit RNA–RNA
1032 interaction. We postulated that a high-concentration, nuclear-localized, condensing RNA-binding
1033 protein might perform a similar role endogenously in eukaryotic transcription as SARS-CoV-2 N
1034 protein does for viral RNA production. Following this logic, we reasoned that host condensate
1035 proteins may also result in arrested networks upon binding to perfectly complementary RNA.
1036 Thus, we sought to test the propensity of such a protein to form arrested networks when
1037 perfectly complementary RNA is present and show that FUS may be partially able to mitigate
1038 extensive RNA-RNA interaction in a manner like N. These data collectively suggest a protein
1039 sequence independent role for condensing proteins in suppression of strong trans RNA-RNA
1040 interactions.
1041 In summation, our data collectively reveal the importance of RNA-RNA interaction strength
1042 in regulating condensate morphology. Weak RNA-RNA interactions conferred by low
1043 complementarity in RNA sequences results in round condensates whereas strong RNA-RNA
1044 interactions conferred by high complementarity in RNA sequences resulted in arrested networks.
1045 Our data indicate this phenomenon is largely RNA and protein sequence independent and thus
1046 broadly applicable across the biomolecular condensate field whenever RNA is present.
1047 Additionally, our results suggest explanations for the enrichment for specific RNA sequence and
1048 structure features in RNA viral genomes due to reliance on RdRp proteins and RNA templates for
1049 replication.

1050

1051 **ACKNOWLEDGEMENTS**

1052

We thank Professor Yifan Dai for the use of his microscope and his lab in the collecting the preliminary data for this manuscript. We thank Dr. Wilton Snead for providing FUS and TEV protein. We thank Professor Zhao Zhang in the Department of Pharmacology and Cancer Biology for use of his Nanopore Instrumentation GridIon. This study was supported in part by the Center for Microbial Pathogenesis and Host Inflammatory Responses grant P20GM103625 through the NIH National Institute of General Medical Sciences Centers of Biomedical Research Excellence. J.C.M and K.D.R were supported by NIGMS MIRA (R35GM122601) and a kind gift from the Palade Family. J.A.J. acknowledges start-up funds provided by the Department of Chemical and Biological Engineering and the Omenn–Darling Bioengineering Institute at Princeton University. J.A.J. also acknowledges research support from the Chan Zuckerberg Initiative DAF (an advised fund of Silicon Valley Community Foundation; grant 2023-332391), the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM155259, and the National Science Foundation (NSF) through the Princeton University (PCCM) Materials Research Science and Engineering Center DMR-2011750. The simulations reported on in this manuscript was substantially performed using the Princeton Research Computing resources at Princeton University, which is a consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology. C.A.R .acknowledges start-up funds provided by the Université de Montréal. C.A.R. was additionally supported by NIH T32CA9156-43, F32GM136164, K99AI173439-01A1 and L'OREAL USA for Women in Science Fellowship. V.Z. and A.S.G were supported by 7R01GM081506-13. V.Z. was supported by Duke University School of Medicine International Chancellor's scholarship.

COMPETEING INTERESTS

All other authors declare that they have no competing interests.

DATA AND MATERIALS AVAILABILITY

All data and/or materials are available upon request from C.A.R. or J.A.J.

CONTRIBUTIONS

J.A.J, C.A.R., and A.S.G. conceptualized the project, designed experiments, prepared figures, drafted and edited the manuscript. C.A.R. and V.Z. performed experiments, prepared figures and analyzed data. D.A. and R.S. designed and performed experiments and computational analyses, analyzed data, prepared figures. J.C.M and K.D.R designed experiments and analyzed data. All authors contributed to the editing of the manuscript.

METHODS

Protein production: SARS-CoV-2 nucleocapsid protein was produced occurring to our established protocols[21]. RNAse A was purchased from NEB (T3018L).

FUS protein production and labeling

We produced the following sequence.

6xHis-MBP-FUS FL WT.

MGSDKIHHHHHHSSGTKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFP
QVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEA

LSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKY
DIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKV
NYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVAL
KSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTN
SGSDITSLYKKAEGGTENLYFQGHMASNDYTQQATQSYGAYPTQPGQGYSQQSSQPYGQQS
YSGYSQSTDTSGYGQSSYSSYGQSQNTGYGTQSTPQGYGSTGGYGSSQSSQSSYGQQSSY
PGYGQQPAPSSTSGSYGSSSQSSSYGQPQSGSYSQQPSYGGQQQSYGQQQSYNPPQGYG
QQNQYNSSSGGGGGGGGGGNYGQDQSSMSSGGGSGGGYGNQDQSGGGGSGGYGQQDR
GGRGRGGSGGGGGGGGGGYNRSSGGYEPRGRGGGRGGRGGMGGSDRGGFNKFGGPRD
QGSRHDSEQDNSDNNTIFVQGLGENVTIESVADYFKQIGIIKTNKKTGQPMINLYTDRETGKLKG
EATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFATRRADFNRGGGNGRGGRGRGGPMGRGG
YGGGGSGGGGRGGFPSGGGGGGGQQRAGDWKCPNPTCENMNFSWRNECNQCKAPKPDG
PGGGPGGSHMGGNYGDDRRGGRGGYDRGGYRGRGGDRGGFRGGRGGGDRGGFGPGKM
DSRGEHRQDRRERPY

The 6xHis-MBP-FUS FL WT (MBP-FUS) and MBP-7xHis-TEV protease (MBP-TEV) plasmids were obtained from Addgene (plasmids # 98651 and # 8827, respectively). After transforming the plasmids into BL21 E. coli, cells were grown at 37°C until reaching OD600 0.6-0.8 and proteins were induced overnight at 18°C following addition of IPTG (0.5 mM and 0.4 mM for MBP-FUS and MBP-TEV, respectively). Cells were harvested by centrifugation at 14,000 rcf at 4degC the following day and resuspended in lysis buffer. For MBP-FUS, lysis buffer consisted of 50 mM HEPES pH 7.4, 1.5 M NaCl, 10 percent glycerol, 20 mM imidazole, 5 mM beta-mercaptoethanol, and EDTA-free protease inhibitor cocktail (Pierce A32965); for MBP-TEV, lysis buffer consisted of 25 mM HEPES pH 7.4, 300 mM KCl, 10 percent glycerol, and 20 mM imidazole. Cells were lysed using probe sonication (QSonica Q500 with 20 kHz converter and 1/4" microtip), and lysate was clarified by centrifugation at 27,000 rcf for 30 min at 4°C. Clarified lysate was mixed with 0.4-0.5 mL of washed, packed HisPur cobalt resin (Thermo Scientific 89965) per 1 L of cells for 1 hour at 4°C. After protein binding, resin was transferred into a gravity flow column and washed with approximately 80-100 resin bed volumes of lysis buffer without protease inhibitor cocktail. MBP-TEV was further washed using approximately 100 resin bed volumes of lysis buffer containing 10 mM ATP-MgCl2, followed by an additional 80-100 resin bed volumes of lysis buffer without ATP-MgCl2. Proteins were then eluted using elution buffer. For MBP-FUS, elution buffer consisted of 50 mM HEPES pH 7.4, 150 mM NaCl, 10 percent glycerol, 200 mM imidazole, and 5 mM beta-mercaptoethanol; for MBP-TEV, elution buffer consisted of 25 mM HEPES pH 7.4, 150 mM KCl, 10 percent glycerol, and 250 mM imidazole. Proteins were dialyzed overnight using 3 mL Slide-A-Lyzer 20K and 10K MWCO dialysis cassettes for MBP-FUS and MBP-TEV, respectively, (Thermo Scientific 66003 and 66455) with two rounds of 1 L dialysis buffer. For MBP-FUS, dialysis buffer consisted of 20 mM HEPES pH 7.4, 150 mM NaCl, 5 percent glycerol, and 5 mM beta-mercaptoethanol; for MBP-TEV, dialysis buffer consisted of 25 mM HEPES pH 7.4, 150 mM KCl, 10 percent glycerol, and 5 mM beta-mercaptoethanol.

Approximately 0.5 mL of purified MBP-FUS protein was fluorescently labeled with amine-reactive Atto 488-NHS ester (Sigma-Aldrich 41698-1MG-F). After resuspending the reactive dye in anhydrous DMSO to a stock concentration of 10 mM, dye was added to the protein at a dye:protein molar ratio of 1:1. The dye was allowed to conjugate to the protein for 20 minutes at room temperature, and the mixture was transferred to a 0.5 mL Slide-A-Lyzer 20K MWCO dialysis cassette (Thermo Scientific 66005) and dialyzed with two round of 0.5 L MBP-FUS dialysis buffer to remove unconjugated dye.

Protein and conjugated dye concentrations were measured using a Nanodrop spectrophotometer. The A260/A280 absorbance ratios were found to be in the range of 0.60-0.65, indicating relatively

minor absorbance contributions from residual nucleic acids. Small aliquots of labeled and unlabeled proteins were snap frozen in liquid nitrogen and stored at -80°C.

NSP13 protein production:

Plasmid construction:

The coding sequence for NSP13 from the SARS CoV-2 Washington isolate (Genbank MN985325) was synthesized as an E. coli codon-optimized fragment (GenScript) and cloned into the pUC57 vector by the manufacturer. The NSP13 sequence was PCR-amplified with Pfu polymerase, and the resulting fragment cloned into the BsaI site of the pSUMO (LifeSensors) plasmid using the HiFi Assembly kit (NEB) to produce an N-terminal six histidine-tagged SUMO-NSP13 fusion cassette (6XHis-SUMO-NSP13). The ULP-1 cleavage site encoded in the SUMO tag was located between the 6X-His-SUMO fragment and the NSP13 sequence to allow for the production of authentic N-terminal SARS Co-V2 NSP13 protein. Samples were sequenced at the UAMS Sequencing Core Facility using a 3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA).

Sequence of NSP13:

SUMO-NSP13 that is expressed, and the NSP13 WT in bold:

MGHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQ
GKEMDSLRFLYDGIRIQADQAPEDLDMEDNDIIEAHREQIGG**AVGACVLCNSQTSLRCGACIRR**
PFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDVTQLYLGGMSYYCKSHKPPISFPLCA
NGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAGDYILANTCTERLKLFAAETLKATEETFKLS
YGIATVREVLSDRELHLSWEVGKPRPPLNRNYVFTGYRVTKNSKVQIGEYTFEKGDYGDAVV
YRGTTTYKLNVGDYFVLTSHTVMPLSAPTLVPQEHYVRITGLYPTLNISDEFSSNVANYQKVG
MQKYSTLQGPPGTGKSHFAIGLALYYPSARIVYTACSHAAVDALCEKALKYLPIDKCSRIIPAR
ARVECFDKFKVNSTLEQYVFCTVNALPETTADIVVFDEISMATNYDLSVVNARLRAKHYVYIGD
PAQLPAPRTLLTKGTLEPEYFNSVCRLMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAH
KDKSAQCFKMFYKGVITHDVSSAINRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASKILG
LPTQTVDSSQGSEYDYVIFTQTTETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEI
PRRNVATLQ

Protein Expression and Purification:

Wild-type plasmids were transformed into Rosetta2 cells, and colonies were grown to saturation overnight at 37°C in NZCYM (Research Products International) supplemented with kanamycin (50 µg/ml) and chloramphenicol (25 µg/ml). The cultures were diluted 1:100 in fresh antibiotic-containing NZCYM media and allowed to grow until reaching an OD600 nm of 0.8-1. The cultures were supplemented with 0.1 mM ZnSO4 and 0.2% dextrose and cooled on ice for 10 minutes. Expression of the proteins was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18°C for 12-16 hours. The cells were harvested by centrifugation at 4,000 x g for 15 minutes at 4°C, and the cell pellets were stored at -80°C until purification.

All purifications steps were carried out on ice or at 4°C. Pellets were fully resuspended in lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 1 mM β-mercaptoethanol, 10% glycerol and 20 mM imidazole) supplemented with 2 mM phenylmethylsulfonyl fluoride (PMSF) and 1X EDTA-free protease inhibitor cocktail (Pierce), lysed by microfluidization and clarified by centrifugation at 17,000 x g for 1 hour at 4°C. The His-tagged SUMO-NSP13 was initially isolated from the crude supernatant using immobilized metal ion affinity chromatography (IMAC). The sample was passed through a HisTrap FF column (Cytiva) at 1 ml/minute. The Ni Sepharose affinity resin was washed extensively with 20 column volumes of buffer, and the protein eluted with 10 column volumes of lysis buffer containing 200 mM imidazole. After dialysis of the pooled

SUMO-NSP13-containing fractions into low imidazole lysis buffer, the 6XHis-SUMO tag was cleaved with 6XHis-ULP-1 for 4 hours at 4°C, and complete digestion was confirmed by SDS-PAGE analysis. The ULP-1 and SUMO tags were separated from the NSP13 proteins by subjecting the sample to a second round of Ni2+-affinity chromatography as before. NSP13-containing fractions were pooled, dialyzed against low salt buffer (50 mM sodium phosphate, pH 6.8, 150 mM NaCl, 4 mM β-mercaptoethanol, 0.5 mM EDTA and 10% glycerol) and passed through a HighTrap SP (Cytiva) ion exchange column. Under these conditions, NSP13 and mutants did not adhere to the SP column and the flow-thru fractions were collected. The NSP13-containing fractions were pooled and concentrated with an Amicon Ultra-15 centrifugation filter unit. The NSP13 protein sample was passed through a Sephacyl S200-HR HiPrep 26/60 (Cytiva) column equilibrated in NSP13 Storage Buffer (25 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM TCEP and 20% glycerol). Purified NSP13 was quantified by UV spectrophotometry at 280 nm using the expected extinction coefficient of 68,785 M-1 cm-1 and confirmed using the BCA Protein Assay (Pierce). Protein samples were aliquoted and stored at -80°C.

RNA template production: RNAs representing the TRS-L (plus) and 6 of the TRS-Bs (plus and minus) of SARS-CoV-2 were produced by PCR amplification with T7 containing primers (iProof) from SARS-CoV-2 RT-PCR genome fragments 1, 4, and 5, a kind gift from Ahmet Yildez lab[22]. Templates were chosen to be equal length (480+ GGG) and have equal flanking sequences 5′ and 3′ of the TRS-B. Ensemble diversity was predicted using Vienna fold[95]. RNA–RNA interaction strength was predicted utilizing RNA-hybrid[45]. G-quadruplexes for both the plus and minus strand SARS-CoV-2 RNA genome were predicted using QGRS mapper[63] under default parameters. The following template sequences were utilized, position of TRS sequence is marked in red and YRRRY's is marked in yellow highlight.

Spike plus

GGGGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTC
TTCCTATTCTTTATTTGACATGAGTAAATTTCCCCTTAAATTAAGGGGTACTGCTGTTATGTC
TTTAAAAGAAGGTCAAATCAATGATATGATTTTATCTCTTCTTAGTAAAGGTAGACTTATAAT
TAGAGAAAACAACAGAGTTGTTATTTCTAGTGATGTTCTTGTTAACAACTAAACGAACAATG
TTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTACAACCAGAACT
CAATTACCCCCTGCATACACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAGTTTT
CAGATCCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTTTTCCAATGTTACTTG
GTTCCATGCTATACATGTCTCTGGGACCAATGGTACTAAGAGGTT

Envelope plus

GGGGTTGTATTACACAGTTACTTCACTTCAGACTATTACCAGCTGTACTCAACTCAATTGAG
TACAGACACTGGTGTTGAACATGTTACCTTCTTCATCTACAATAAAATTGTTGATGAGCCTG
AAGAACATGTCCAAATTCACACAATCGACGGTTCATCCGGAGTTGTTAATCCAGTAATGGA
ACCAATTTATGATGAACCGACGACGACTACTAGCGTGCCTTTGTAAGCACAAGCTGATGAG
TACGAACTTATGTACTCATTCGTTTCGGAAGAGACAGGTACGTTAATAGTTAATAGCGTACT
TCTTTTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCCATCCTTACTGCGCTTCGATT
GTGTGCGTACTGCTGCAATATTGTTAACGTGAGTCTTGTAAAACCTTCTTTTTACGTTTACT
CTCGTGTTAAAAATCTGAATTCTTCTAGAGTTCCTGATCTTCTGGTCTAA

Membrane plus

GGGTTATGTACTCATTCGTTTCGGAAGAGACAGGTACGTTAATAGTTAATAGCGTACTTCTT
TTTCTTGCTTTCGTGGTATTCTTGCTAGTTACACTAGCCATCCTTACTGCGCTTCGATTGTG
TGCGTACTGCTGCAATATTGTTAACGTGAGTCTTGTAAAACCTTCTTTTTACGTTTACTCTCG
TGTTAAAAATCTGAATTCTTCTAGAGTTCCTGATCTTCTGGTCTAAACGAACTAAATATTATA
TTAGTTTTTCTGTTTGGAACTTTAATTTTAGCCATGGCAGATTCCAACGGTACTATTACCGTT
GAAGAGCTTAAAAAGCTCCTTGAACAATGGAACCTAGTAATAGGTTTCCTATTCCTTACATG

1245 GATTTGTCTTCTACAATTTGCCTATGCCAACAGGAATAGGTTTTTGTATATAATTAAGTTAAT
1246 TTTCCTCTGGCTGTTATGGCCAGTAACTTTAGCTTGTTTTGTGCT
1247 **ORF7 plus**
1248 GGGATTCCAGTAGCAGTGACAATATTGCTTTGCTTGTACAGTAAGTGACAACAGATGTTTCA
1249 TCTCGTTGACTTTCAGGTTACTATAGCAGAGATATTACTAATTATTATGAGGACTTTTAAAGT
1250 TTCCATTTGGAATCTTGATTACATCATAAACCTCATAATTAAAAATTTATCTAAGTCACTAACT
1251 GAGAATAAATATTCTCAATTAGATGAAGAGCAACCAATGGAGATTGATTAAACGAACATGAA
1252 AATTATTCTTTTCTTGGCACTGATAACACTCGCTACTTGTGAGCTTTATCACTACCAAGAGT
1253 GTGTTAGAGGTACAACAGTACTTTTAAAAGAACCTTGCTCTTCTGGAACATACGAGGGCAA
1254 TTCACCATTTCATCCTCTAGCTGATAACAAATTTGCACTGACTTGCTTTAGCACTCAATTTGC
1255 TTTTGCTTGTCCTGACGGCGTAAAACACGTCTATCAGTTACGTGCC
1256 **ORF8 plus**
1257 GGGGTTCATCAGACAAGAGGAAGTTCAAGAACTTTACTCTCCAATTTTTCTTATTGTTGCGG
1258 CAATAGTGTTTATAACACTTTGCTTCACACTCAAAAGAAAGACAGAATGATTGAACTTTCATT
1259 AATTGACTTCTATTTGTGCTTTTTAGCCTTTCTGCTATTCCTTGTTTTAATTATGCTTATTATC
1260 TTTTGGTTCTCACTTGAACTGCAAGATCATAATGAAACTTGTCACGCCTAAACGAACATGAA
1261 ATTTCTTGTTTTCTTAGGAATCATCACAACTGTAGCTGCATTTCACCAAGAATGTAGTTTACA
1262 GTCATGTACTCAACATCAACCATATGTAGTTGATGACCCGTGTCCTATTCACTTCTATTCTA
1263 AATGGTATATTAGAGTAGGAGCTAGAAAATCAGCACCTTTAATTGAATTGTGCGTGGATGA
1264 GGCTGGTTCTAAATCACCCATTCAGTACATCGATATCGGTAATTAT
1265 **Nucleocapsid plus**
1266 GGGAAATGGTATATTAGAGTAGGAGCTAGAAAATCAGCACCTTTAATTGAATTGTGCGTGG
1267 ATGAGGCTGGTTCTAAATCACCCATTCAGTACATCGATATCGGTAATTATACAGTTTCCTGT
1268 TTACCTTTTACAATTAATTGCCAGGAACCTAAATTGGGTAGTCTTGTAGTGCGTTGTTCGTT
1269 CTATGAAGACTTTTTAGAGTATCATGACGTTCGTGTTGTTTTAGATTTCATCTAAACGAACAA
1270 ACTAAAATGTCTGATAATGGACCCCAAAATCAGCGAAATGCACCCCGCATTACGTTTGGTG
1271 GACCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCAGTGGGGCGCGATCAAAAC
1272 AACGTCGGCCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACCGCTCTCACTCAACA
1273 TGGCAAGGAAGACCTTAAATTCCCTCGAGGACAAGGCGTTCCAATTAACACCA
1274 **Spike minus**
1275 GGGAACCTCTTAGTACCATTGGTCCCAGAGACATGTATAGCATGGAACCAAGTAACATTGG
1276 AAAAGAAAGGTAAGAACAAGTCCTGAGTTGAATGTAAAACTGAGGATCTGAAAACTTTGTCA
1277 GGGTAATAAACACCACGTGTGAAAGAATTAGTGTATGCAGGGGGTAATTGAGTTCTGGTTG
1278 TAAGATTAACACACTGACTAGAGACTAGTGGCAATAAAACAAGAAAAACAAACATTGTTCGT
1279 TTAGTTGTTAACAAGAACATCACTAGAAATAACAACTCTGTTGTTTTCTCTAATTATAAGTCT
1280 ACCTTTACTAAGAAGAGATAAAATCATATCATTGATTTGACCTTCTTTTAAAGACATAACAGC
1281 AGTACCCCTTAATTTAAGGGGAAATTTACTCATGTCAAATAAAGAATAGGAAGACAACTGAA
1282 TTGGATTTGTATTCCTCCAAAATATGTAATTTGCATGCATGACATAACC
1283 **Envelope minus**
1284 GGGTTAGACCAGAAGATCAGGAACTCTAGAAGAATTCAGATTTTTAACACGAGAGTAAACG
1285 TAAAAAGAAGGTTTTACAAGACTCACGTTAACAATATTGCAGCAGTACGCACACAATCGAAG
1286 CGCAGTAAGGATGGCTAGTGTAACTAGCAAGAATACCACGAAAGCAAGAAAAAGAAGTAC
1287 GCTATTAACTATTAACGTACCTGTCTCTTCCGAAACGAATGAGTACATAAGTTCGTACTCAT
1288 CAGCTTGTGCTTACAAAGGCACGCTAGTAGTCGTCGTCGGTTCATCATAAATTGGTTCCAT
1289 TACTGGATTAACAACTCCGGATGAACCGTCGATTGTGTGAATTTGGACATGTTCTTCAGGC
1290 TCATCAACAATTTTATTGTAGATGAAGAAGGTAACATGTTCAACACCAGTGTCTGTACTCAA
1291 TTGAGTTGAGTACAGCTGGTAATAGTCTGAAGTGAAGTAACTGTGTAATACAAC
1292 **Membrane minus**
1293 GGGAGCACAAAACAAGCTAAAGTTACTGGCCATAACAGCCAGAGGAAAATTAACTTAATTA
1294 TATACAAAAACCTATTCCTGTTGGCATAGGCAAATTGTAGAAGACAAATCCATGTAAGGAAT
1295 AGGAAACCTATTACTAGGTTCCATTGTTCAAGGAGCTTTTTAAGCTCTTCAACGGTAATAGT

1296 ACCGTTGGAATCTGCCATGGCTAAAATTAAAGTTCCAAACAGAAAAACTAATATAATATTTA
1297 GTTCGTTTAGACCAGAAGATCAGGAACTCTAGAAGAATTCAGATTTTTAACACGAGAGTAAA
1298 CGTAAAAAGAAGGTTTTACAAGACTCACGTTAACAATATTGCAGCAGTACGCACACAATCG
1299 AAGCGCAGTAAGGATGGCTAGTGTAACTAGCAAGAATACCACGAAAGCAAGAAAAAGAAGT
1300 ACGCTATTAACTATTAACGTACCTGTCTCTTCCGAAACGAATGAGTACATAA
1301 **ORF7 minus**
1302 GGGGGCACGTAACTGATAGACGTGTTTTACGCCGTCAGGACAAGCAAAAGCAAATTGAGT
1303 GCTAAAGCAAGTCAGTGCAAATTTGTTATCAGCTAGAGGATGAAATGGTGAATTGCCCTCG
1304 TATGTTCCAGAAGAGCAAGGTTCTTTTAAAAGTACTGTTGTACCTCTAACACACTCTTGGTA
1305 GTGATAAAGCTCACAAGTAGCGAGTGTTATCAGTGCCAAGAAAAGAATAATTTTCATGTTCG
1306 TTTAATCAATCTCCATTGGTTGCTCTTCATCTAATTGAGAATATTTATTCTCAGTTAGTGACT
1307 TAGATAAATTTTTAATTATGAGGTTTATGATGTAATCAAGATTCCAAATGGAAACTTTAAAAG
1308 TCCTCATAATAATTAGTAATATCTCTGCTATAGTAACCTGAAAGTCAACGAGATGAAACATCT
1309 GTTGTCACTTACTGTACAAGCAAAGCAATATTGTCACTGCTACTGGAAT
1310 **ORF8 minus**
1311 GGGATAATTACCGATATCGATGTACTGAATGGGTGATTTAGAACCAGCCTCATCCACGCAC
1312 AATTCAATTAAAGGTGCTGATTTTCTAGCTCCTACTCTAATATACCATTTAGAATAGAAGTGA
1313 ATAGGACACGGGTCATCAACTACATATGGTTGATGTTGAGTACATGACTGTAAACTACATTC
1314 TTGGTGAAATGCAGCTACAGTTGTGATGATTCCTAAGAAAACAAGAAATTTCATGTTCGTTT
1315 AGGCGTGACAAGTTTCATTATGATCTTGCAGTTCAAGTGAGAACCAAAAGATAATAAGCATA
1316 ATTAAAACAAGGAATAGCAGAAAGGCTAAAAAGCACAAATAGAAGTCAATTAATGAAAGTTC
1317 AATCATTCTGTCTTTCTTTTGAGTGTGAAGCAAAGTGTTATAAACACTATTGCCGCAACAATA
1318 AGAAAAATTGGAGAGTAAAGTTCTTGAACTTCCTCTTGTCTGATGAAC
1319 **Nucleocapsid minus**
1320 GGGATAATTACCGATATCGATGTACTGAATGGGTGATTTAGAACCAGCCTCATCCACGCAC
1321 AATTCAATTAAAGGTGCTGATTTTCTAGCTCCTACTCTAATATACCATTTAGAATAGAAGTGA
1322 ATAGGACACGGGTCATCAACTACATATGGTTGATGTTGAGTACATGACTGTAAACTACATTC
1323 TTGGTGAAATGCAGCTACAGTTGTGATGATTCCTAAGAAAACAAGAAATTTCATGTTCGTTT
1324 AGGCGTGACAAGTTTCATTATGATCTTGCAGTTCAAGTGAGAACCAAAAGATAATAAGCATA
1325 ATTAAAACAAGGAATAGCAGAAAGGCTAAAAAGCACAAATAGAAGTCAATTAATGAAAGTTC
1326 AATCATTCTGTCTTTCTTTTGAGTGTGAAGCAAAGTGTTATAAACACTATTGCCGCAACAATA
1327 AGAAAAATTGGAGAGTAAAGTTCTTGAACTTCCTCTTGTCTGATGAAC
1328 **5′end plus**
1329 GGGTTAAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCGATCTCTTGTAGATCT
1330 GTTCTCTAAACGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCA
1331 CGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCT
1332 GCAGGCTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCATCAGCACATCTAGGTTTCGTCC
1333 GGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACG
1334 TCCAACTCAGTTTGCCTGTTTTACAGGTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTC
1335 CGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGGCTTAGTA
1336 GAAGTTGAAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTGTTCATCAAA
1337 **Spike minus 10% Overlap**
1338 GGGTGGATTTGTATTCCTCCAAAATATGTAATTTGCATGCATGACATAACCATCTATTTGTTC
1339 GCGTGGTTTGCCAAGATAATTACATCCAATTAAAAATGCTTCAGATGATGACGCATTCACAT
1340 TAGTAACAAAGGCTGTCCACCATGCGAAGTGTCCCATGAGCTTATAAAGATCAGCATTCCA
1341 AGAATGTTCTGTTATCTTTATAGCCACGGAACCTCCAAGAGCTAGCTTTTGTTGTATAAACC
1342 CACAAATGTAAGTGAAAAAACCCTCTTTAGAGTCATTTTCTTTTGTAACATTTTTAGTCTTAG
1343 GGTCGTACATATCACTAATAATGAGATCCCATTTATTAGCTGTATGTACAGTTGCACAATCA
1344 CCAATCAAAGTTGAATCTGCATCAGAGACAAAGTCATTAAGATCTGAATCGACAAGCAGCG
1345 TACCCGTAGGCAACCACTGTCTTAAAACAGCTGTACCTGGTGCAACTCC
1346 **Nucleocapsid minus 10% Overlap**

GGGTTCAATTAAAGGTGCTGATTTTCTAGCTCCTACTCTAATATACCATTTAGAATAGAAGT
GAATAGGACACGGGTCATCAACTACATATGGTTGATGTTGAGTACATGACTGTAAACTACAT
TCTTGGTGAAATGCAGCTACAGTTGTGATGATTCCTAAGAAAACAAGAAATTTCATGTTCGT
TTAGGCGTGACAAGTTTCATTATGATCTTGCAGTTCAAGTGAGAACCAAAAGATAATAAGCA
TAATTAAAACAAGGAATAGCAGAAAGGCTAAAAAGCACAAATAGAAGTCAATTAATGAAAGT
TCAATCATTCTGTCTTTCTTTTGAGTGTGAAGCAAAGTGTTATAAACACTATTGCCGCAACA
ATAAGAAAAATTGGAGAGTAAAGTTCTTGAACTTCCTCTTGTCTGATGAACAGTTTAGGTGA
AACTGATCTGGCACGTAACTGATAGACGTGTTTTACGCCGTCAGGACAA

RNA PRODUCTION

RNA was produced according to our established protocols with T7 in vitro transcription kit (NEB) with labeled cy3, cy5, and atto488 UTPs and purified with LiCl precipitation. Produced RNAs were quality controlled for concentration (nanodrop) as well as size and purity on a denaturing 1% agarose gel (stained with sybr gold). Data for the manuscript are representative of at least 2 independent batches of RNA for each tested sequence (1 labeled and 1 unlabeled). Label fraction for RNA was quantified via nanodrop and RNAs. Prior to use, RNAs were diluted to 1μM stock solutions and stored in labeled aliquots in the -80°C.

DEPC RNA structure probing

To validate RNA-RNA interaction arising from RNA complementarity, we performed a series of chemical probing experiments coupled with Nanopore direct RNA sequencing. First, all the RNAs are prepared using in-vitro transcription as described above. Then, 250nM of each RNA is independently probed and probed with its counterpart (e.g. sense with antisense, anti-sense with 5′end) to discern how it samples its thermodynamic landscape independently and with a binding partner, respectively.

For the RNAs that are independently probed, 10μL of 250nM RNA was denatured at 90°C for 3 min, cooled on ice, and renatured in 150mM NaCl, 20mM HEPES ph 7.4 for 20 min at 20°C. To test RNA-RNA interaction, a 10μL stock containing 250nM of RNA1 and 250uM of its counterpart RNA will be denatured using the same process as described above. For (+) DEPC, RNA was treated with 10% DEPC for 45 min at 20°C. Diethyl pyrocarbonate (DEPC) (Sigma) carbethoxylates unpaired adenosine at N-6 or N-7 by opening the imidazole ring. The reactions were immediately purified using RNA Clean & Concentrator (Zymo Research) or RNA XP bead (Beckman Coulter).

Sequencing libraries were prepared following the direct RNA sequencing protocol RNA-004 (Oxford Nanopore Technologies, ONT) using sequence specific 3′ adapters (which is made by annealing). After renaturation and treatment with DEPC, 3′ adapter-ligated RNAs were reverse-transcribed using superscript III. The reverse transcription step serves to prop up RNA using its complementary cDNA to obtain longer reads and minimize the impact of RNA structure on nanopore readthrough. (Note that the direct-RNA sequencing method is sequencing the RNA through Nanopore, not the cDNA) The reverse transcribed RNA was pooled for each experimental condition and a motor protein ligated, which were subjected to Oxford Nanopore sequencing using MinION Flow Cell (RNA) on GridION Mk1 for 72 hour duration. Reads were

basecalled using Super accurate (SUP) model (Oxford Nanopore Technologies, ONT). Reads were aligned to target RNAs allowing for mismatches, insertions and deletions to account for nanopore sequencing errors as well as probing adducts.

For each read, we identify the mismatched nucleotides. Then, we calculated the reactivity for each adenosine by using the equation below:

$$reactivity = \frac{\#mismatch}{\#coverage} \quad (1)$$

The uncertainty of reactivity is calculated using the standard error of binomial distribution, as the higher the read coverage associate with lower uncertainty:

$$\sigma_{reactivity} = \sqrt{\frac{1}{\#coverage}} \times reactivity \quad (2)$$

By comparing the DEPC (+) and DEPC (-) of the independently probed RNA, we concluded that the mismatch rate is a reliable way to gauge RNA structuredness, since the reactivity of the nucleotides on average is higher than the Nanopore sequencing error as reflected in DEPC (-). In addition, the ViennaRNA predicted centroid structure single stranded nucleotides has a higher reactivity on average than the double stranded nucleotides for all RNAs in question.

Next, we compared RNA when probed independently versus with its corresponding binding partner to investigate the nucleotides that are interacting with a binding partner, thereby protected from probing reagent.

For each adenosine position, the change in reactivity is calculated using:

$$\Delta reactivity = reactivity_{probed\ with\ couterpart} - reactivity_{probed\ independently} \quad (3)$$

The uncertainty for the change in reactivity is calculated by propagating the uncertainty of the independent reactivity and uncertainty of reactivity when probed with partners, as described by the equation below

$$\sigma_{\Delta reactivity} = \sqrt{{\sigma_{reactivity1}}^2 + {\sigma_{reactivity2}}^2} \quad (4)$$

DEPC RNA structure models

RNA structure modeling was carried out with DEPC reactivity data and Viennafold software using the following parameters: Fold algorithms and basic options (MFE and partition function); Energy parameters (Turner model 2004); Energy parameters (37C and .150 M salt); After conversion of SHAPE reactivities (All loops Zarringhalam et al 2012); Convert SHAPE reactivities into pairing (Linear mapping).

INOSINE CONTAINING RNA PRODUCTION

Inosine containing RNA was produced by modifying the in vitro transcription reaction replacing increasing percentages of ATPs with Inosine (cayman chemical).

ERROR PRONE RNA PRODUCTION

Error prone PCR templates were produced (Jena) by utilizing multiple 30 cycle reactions with 1 µl of PCR product used as a template for the subsequent PCR reaction thereby increasing error rates. Only 4 to 5 cycles of error prone PCR resulted in difference in arrested network formation.

CELL FREE CONDENSATION REACTION (N PROTEIN)

Condensates were formed in 384 well glass bottom dishes. 15 µl condensate buffer was added to the well (150mM NaCl, 20mM TRIS ph 7.4) followed by 5 µl of RNA and water (control water only). RNA, water, and condensate buffer were then mixed by pipetting a volume of 5µl three times. Following this, 5 µl of 16 µM (0.05% labeled) N protein in N protein buffer or N protein buffer only control (250 mM NaCl 20 mM phosphate buffer ph 7.4) was then added to the well and reactions were mixed by pipetting 5 µl 3 times. Reactions were incubated at 37°C for 2-21 hours prior to imaging. For RNA combinations 1 µM stocks of labeled RNA were premixed on ice prior to addition unless staggering experiments were undertaken as in **Figure 4**.

For staggered RNA addition experiments, plus RNA was added followed by protein followed by minus after incubation. Cell free condensation reaction (FUS): FUS reactions were carried out in the same wells by precoating the plate with blocking solution 15 µl reaction buffer was added followed by 1 µl of TEV protease, followed by RNA followed by protein. A280 absorbance: diffuse phase measurements were acquired for each technical replicate by nanodrop with 1 µl of solution taken from the top of the well after at least 20 hours of incubation.

CELL FREE CONDENSATION REACTION (N PROTEIN) WITH MG2+

Condensation was undertaking using staggered RNA addition with the addition of 10mM final concentration of MgCl2 to the reaction.

IMAGE ACQUISITION

Images were acquired utilizing a Nikon spinning disc confocal with 100× silicone objective. Images are representative of at least 2 frames for at least 3 technical replicates. All images in each panel were acquired during the same imaging session with the same acquisition/laser settings. Displayed images in each panel are contrasted the same and the same area. Scale bars are indicated with a white line. Unless otherwise noted, images were acquired at the glass to be more comparable between round condensates which are found extend only a few microns in z to arrested networks which can extend 30+microns in z.

IMAGE QUANTIFICATION: condensate signal was thresholded on protein channel signal and 0.2µm pixel was used as a cutoff to detect particles. Region of Interest (ROI) maps from detected particles were used to extract size, shape (circularity), RNA and protein maximum intensity signal. Colocalization of RNA signals was quantified using colocalization threshold parameter for ImageJ.

SIMULATION METHODS

For all simulations in this manuscript, we used LAMMPS[96] molecular simulation package (version 23 Jun 2022). For all visualizations of simulation results, OVITO[97] is used.

Coarse-grained Molecular Dynamics Simulations of Nucleocapsid protein and RNA (Model 1)

The nucleocapsid protein is modeled using the Mpipi model[40]. Since this model is mainly developed for studying the flexible disordered proteins, we used rigid bodies to describe the folded domains of N protein: RNA binding domain 1 (PDB:6YI3)[98] and RNA binding domain 2 (PDB: 6YUN)[99]. Similarly, we have used full FUS protein (Uniprot ID: P35637) and treated RRM (281–369) and Zinc finger domain (452–443) using rigid bodies (with representative structures obtained from AlphaFold: AF-P35637-F1-v4). The protein–protein interactions are described by Wang–Frenkel potential (short-ranged non-bonded contacts) and Debye–Hückel theory (long-ranged electrostatics). While the original Mpipi force field contains parameters for RNA at the nucleotide resolution, the RNA is treated as fully flexible and are not able to form base pairs. On the other hand, the SIS RNA model[42], which also represents RNA in a nucleotide level, supports description of RNA base-pairing[41,42]. The original SIS RNA model[42] assumes that the electrostatic interactions are mainly screened out by divalent ions, leaving effective base-pair potentials to describe the interactions between nucleotides. More specifically, the base pairing is captured by a many-body potential that describes canonical A–U, G–C, and noncanonical G–U wobble pairing between nucleotides. The detailed form of the potential and the parameters are explicitly described in [41]. Therefore, we have implemented a hybrid model that combines protein–protein and protein–RNA interactions from Mpipi with RNA–RNA interactions via a modified SIS RNA model. Specifically, we updated the RNA–RNA interactions with the SIS model parameters, and slightly adjusted the molecular diameters of nucleotide beads, such that they are consistent with the mapping used in the SIS RNA model. Moreover, since we are explicitly considering the electrostatics with Debye–Hückel screening, we updated the effective base pairing strengths under such consideration[100].

In each simulation, N protein and the RNA chains are placed in cubic box with enough separation initially. Then each simulation is performed in the *NVT* ensemble using Langevin thermostat at 300 K for up to 2×10^8 timesteps until the dimer formation is observed. The Debye screening length is set to 0.795 nm, which corresponds to a monovalent salt concentration of 0.15 M. The integration step size is set to 10 fs, and the friction is set to γ= 0.01ps^{-1}. Periodic boundary conditions are used.

For N protein–RNA contact maps, we monitor the inter-residue distance compared to the molecular diameters of nucleotides and amino acids, as set by the Wang–Frenkel σ parameter. Here, an amino acid i is in contact with a nucleotide j, if distance $r_{i,j} \leq 1.2\sigma$, which was empirically determined to be suitable for capturing key features in the contact matrices. For the N protein and RNA contact histogram, each distribution is normalized by the maximum number of contacts. For the NCPR (net charge per residue), a scanning window size of 5 amino acids are used.

To investigate the effect of excessive N protein or FUS protein on trans RNA contacts, we conduct two types of simulations for each protein. First, we perform simulations of 64 protein chains in the absence of RNA. Second, we simulate 64 protein chains with two complementary Envelope RNA strands. For both types of simulations, proteins and RNAs are premixed and placed closely in a cubic box of (600 Å)^3. Then, the systems are simulated for up to 10^8 timesteps using the exact same simulation parameters as described above for the N protein–RNA dimer simulations.

Coarse-grained Molecular Dynamics Simulations of RNA clusters (Model 2)

To investigate the relationship between RNA sequence and propensity to form trans RNA contacts, we simulate RNA strands in the absence of proteins. Here, we used the original SIS RNA model[41,42], which was designed to capture the effect of RNA base pairing on RNA cluster formation at nucleotide resolution. Using this model, we study RNA-only systems, where the computational expense remains high even at nucleotide resolution due to the length of the RNA strands. Notably, probing the systems in the absence of proteins helps us to interrogate whether RNA–RNA interactions are sufficient to lead to the arrested phenotypes observed in our experiments. We simulated three scenarios for each sequence, a system with only 54 plus strands, a system with only 54 minus strands and a system with half plus (27) and half minus (27) strands. In all simulations, overall concentration, annealing protocol, thermodynamic ensemble (*NVT*), and the damping coefficients are identical[41]. For simulating the Envelope RNA sequence with 162 chains, we use the same overall concentration for 1:1 and 5:1 plus/minus systems. We also characterized finite size effects for our simulations and concluded that 54-chain simulations are sufficient for capturing statistically robust results (**Supplementary Figure S3**).

Each simulation is performed for up to 1.2×10^8 timesteps, where the integration step size is set to 10 fs, and the friction is set to $\gamma = 0.01\text{ps}^{-1}$ for the Langevin thermostat at 293 K (following [42]). This yields a total simulation time of over 1us. Periodic boundary conditions are used, and a short annealing process is performed at the beginning of the simulations to relax the system before equilibration. The data from the last 100 ns is used to calculate the near/at equilibrium RNA–RNA interaction strength, number of cis and trans RNA contacts, and the mean-square-displacements of RNA strands in each system.

For obtaining the average interaction strength, we have obtained the potential energies U from the trajectories, which are then normalized by the total number of nucleotides in the simulations. To determine the number of trans RNA contacts, we evaluated the base-pair energies between nucleotides on different strands and define a trans RNA contact if the base-pairing energy is lower than $-1.5K_{\mathrm{B}}T$, which indicates the stability of the established contact against the thermal fluctuations. Finally, for the contact map analysis, we define contacts using the method described above (but for both intra- and inter-strand cases) and normalized the contact matrices by the maximum contacts of the corresponding simulations.

Limitations of wet-lab experiments:

This study addresses the regulation of biomolecular condensation using purified minimal components of the SARS-CoV-2 virus. Many additional cellular and viral factors likely play important roles in this process.

Limitations of simulations

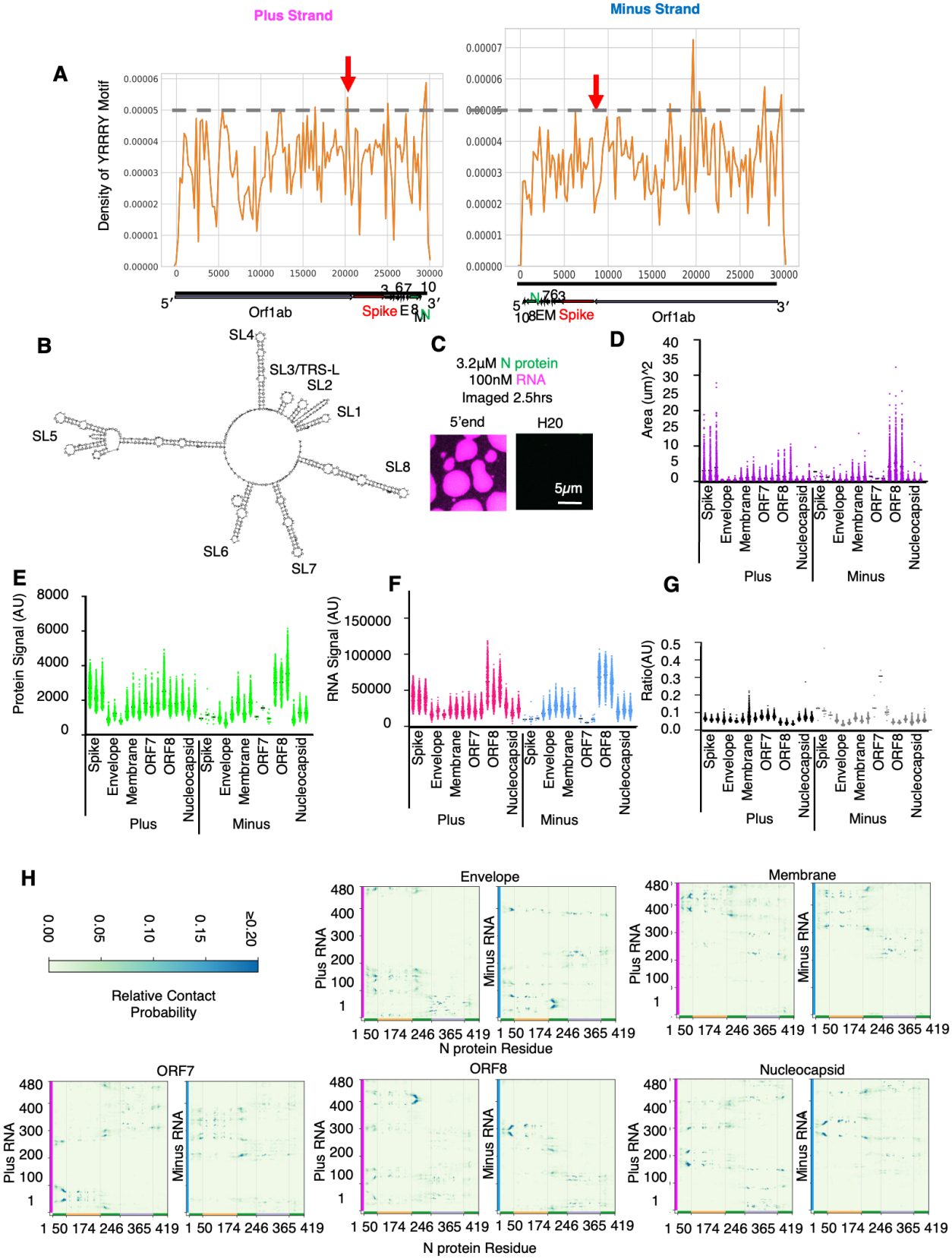
Coarse-grained simulation of protein and RNA is powerful as it can provide molecular insights into the studied system, while enabling us to probe longer timescales and larger system sizes, compared to all-atom simulations. However, these approximations may lead to certain limitations.

1. In all simulations, each amino acid or nucleotide is represented using a single bead or a single interaction site, which means that the structural and dynamic changes below this resolution are inaccessible from our simulations.

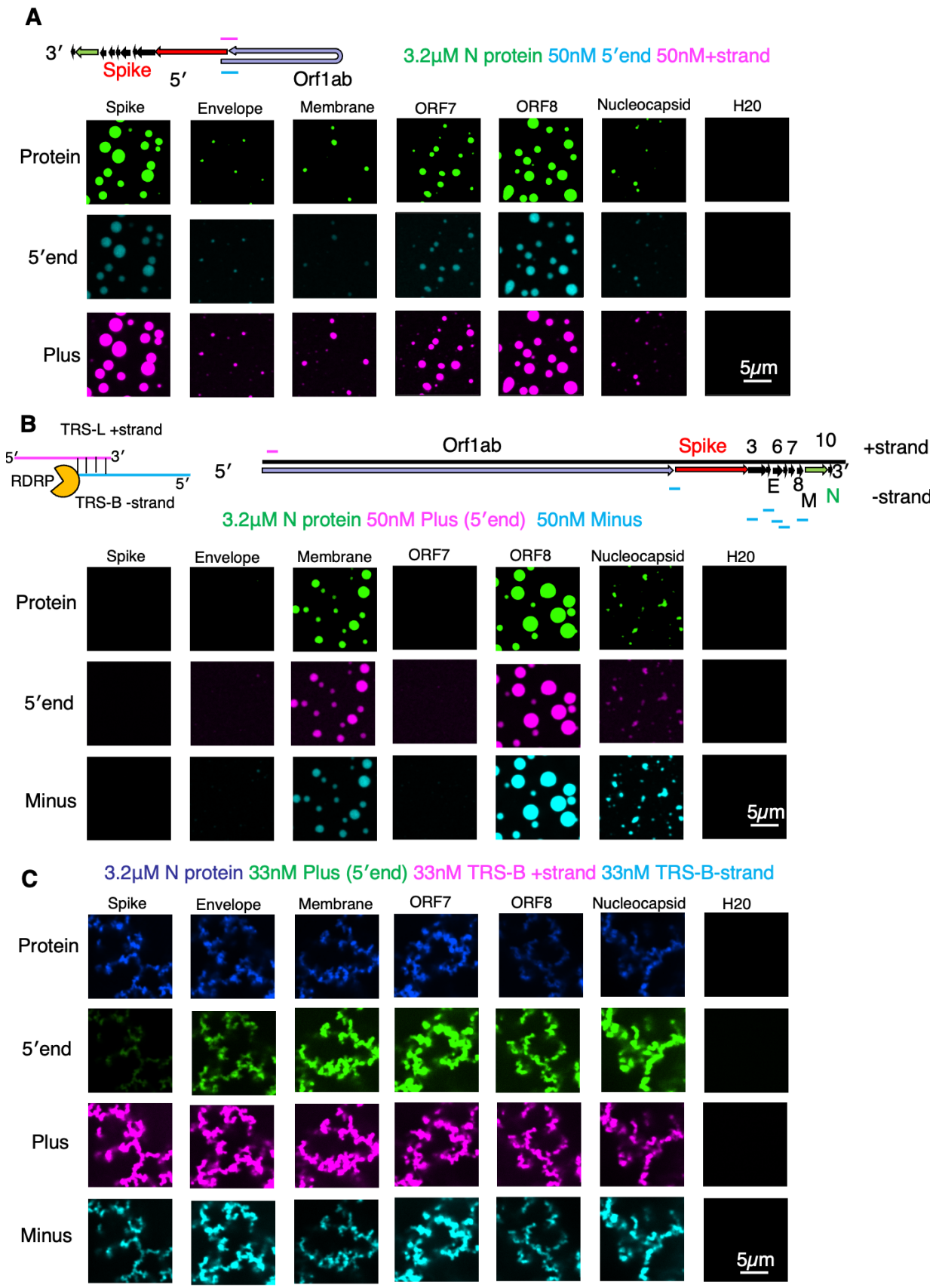
2. In representing the full N protein structure, we used the rigid-body representation for the RNA binding domains, under the assumption that such structured regions may experience minimal conformational changes under the simulation conditions. As a result, the simulation is not able to capture the effects that are related to structural changes in these domains.
3. In the study of N protein and RNA dimeric systems, we have found a nonspecific binding pattern that originates from the structure and the charge distribution of N protein. However, this model is not able to capture more specific binding interactions between N protein and RBD2, which may be important. Here, we capture mainly nonspecific binding effects.
4. Counterion, which is important to RNA condensation, is considered differently when proteins are involved (Model 1) and when RNA clustering is studied (Model 2). In the simulations when proteins are considered, the effective charges are explicitly modeled using Debye–Hückel theory as Mpipi force field is mainly parameterized under 0.15M NaCl concentrations. Therefore, the interaction sites in the RNA representation are also charged. However, when the RNA clustering is studied, it is assumed that the added monovalent and divalent ions effectively screen the electrostatic repulsions between the phosphate groups, and thus the remaining main interaction is the base-pairing between nucleotides at high salt concentrations.
5. In all simulations, only the canonical cis Watson-Crick (cWW) base pairing (GC and AU) and GU wobble pairs are considered. These interactions are favored in the model since cWW base pairing represents >70% of different pairing types in PDB structures and GC, AU, and GU pairing are the most dominant base pairs in cWW base pairing (>90%). However, effects of other base pairs could also serve to augment the degree of trans RNA contacts in our systems.

All relevant supporting data and codes are available in the Figshare data repository at: https://doi.org/10.6084/m9.figshare.28087196.

SUPPLEMENTARY MATERIAL

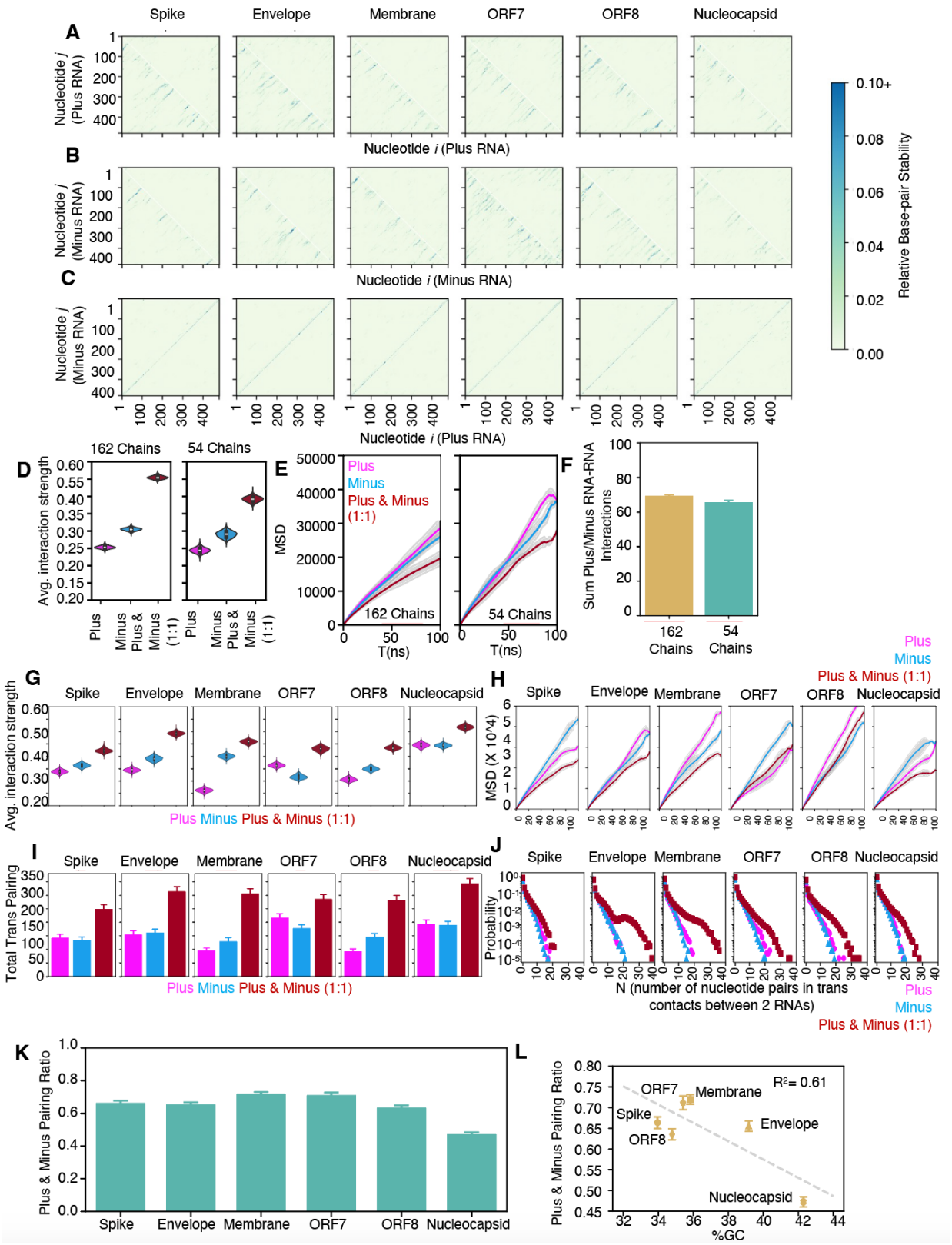


Supplemental Figure S1. TRS-N protein in reconstituted and simulated condensates. Related to **Figure 1**. **(A)** The local density of YRRRY motifs on plus and minus strands. (**B**) Structure model for 5′end RNA (**C**) 5′end RNA condensates under tested conditions for TRS-B fragments from **Figure 1F**. **(D)** Condensate area (μm2) quantification for **Figure 1F** taken from the protein signal. Data show the quantification of one representative image from 3 technical replicates. Plus Spike and minus ORF8 produce consistently larger condensates. Black line is the mean. **(E)** Measurement of max intensity of the protein signal for **Figure 1F**. **(F)** Measurement of max intensity of the RNA signal for panel **Figure 1F**. **(G)** Measurement of ratio max intensity of the RNA and protein signals for **Figure 1F**. **(H)** N protein and protein–RNA interactions are modeled via model 1 (see Methods). For each RNA, plus and minus strands are simulated with N protein separately. The corresponding domain boundaries of N protein are shown by vertical gray lines. The edges of the contact map are colored by the type of biomolecule/domain: yellow indicates RBD 1 of N protein, light purple for RBD 2 of N protein, green for disordered regions of N protein, purple for plus RNA strands, and blue for minus RNA strands. On the contact maps, the high probability region is indicated by bright blue colors. In terms of RNA sequences, the interaction hot spots appear to be random. However, as shown in main text **Figure 1**, the binding spots are relatively conserved with respect to the N protein, which is highly correlated with the charge distributions of N protein.


A
3′
Spike
5′
Orf1ab
3.2μM N protein 50nM 5′end 50nM+strand
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
H20
Protein
5′end
Plus
5μm
B
TRS-L +strand
5′
3′
RDRP
5′
TRS-B -strand
5′
Orf1ab
Spike
3
6 7
10
+strand
E
8
M
N
-strand
3.2μM N protein 50nM Plus (5′end) 50nM Minus
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
H20
Protein
5′end
Minus
5μm
C
3.2μM N protein 33nM Plus (5′end) 33nM TRS-B +strand 33nM TRS-B-strand
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
H20
Protein
5′end
Plus
Minus
5μm

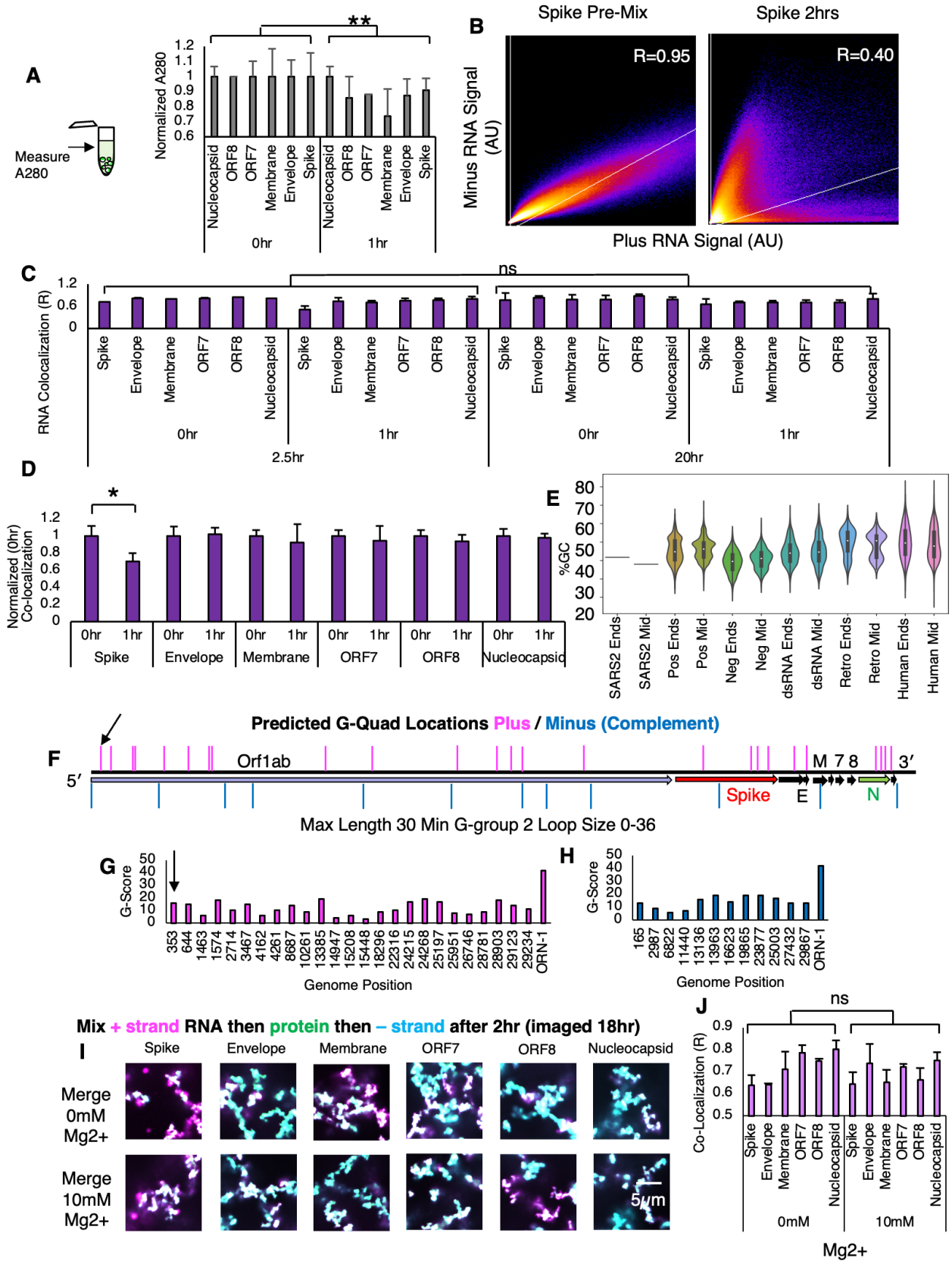
1656

1657 **Supplemental Figure S2. Separated RNA and protein channels from Figure 2.**
1658 For **A-C**, All images depict the unmerged images related to **Figure 2** to indicate that all
1659 reconstituted RNAs and proteins are well mixed and present in the same condensate. Of note,
1660 **panel C**, N protein is labeled in blue rather than green and 5′end is instead labeled in green.
1661
1662
1663


A
B
C
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
Nucleotide i (Plus RNA)
Nucleotide j (Plus RNA)
Nucleotide i (Minus RNA)
Nucleotide j (Minus RNA)
Relative Base-pair Stability
0.10+
0.08
0.06
0.04
0.02
0.00
D
162 Chains
54 Chains
Avg. interaction strength
Plus
Minus
Plus & Minus (1:1)
E
MSD
T(ns)
F
Sum Plus/Minus RNA-RNA Interactions
G
H
MSD (X 10^4)
I
Total Trans Pairing
J
Probability
N (number of nucleotide pairs in trans contacts between 2 RNAs)
K
Plus & Minus Pairing Ratio
L
R²= 0.61
%GC

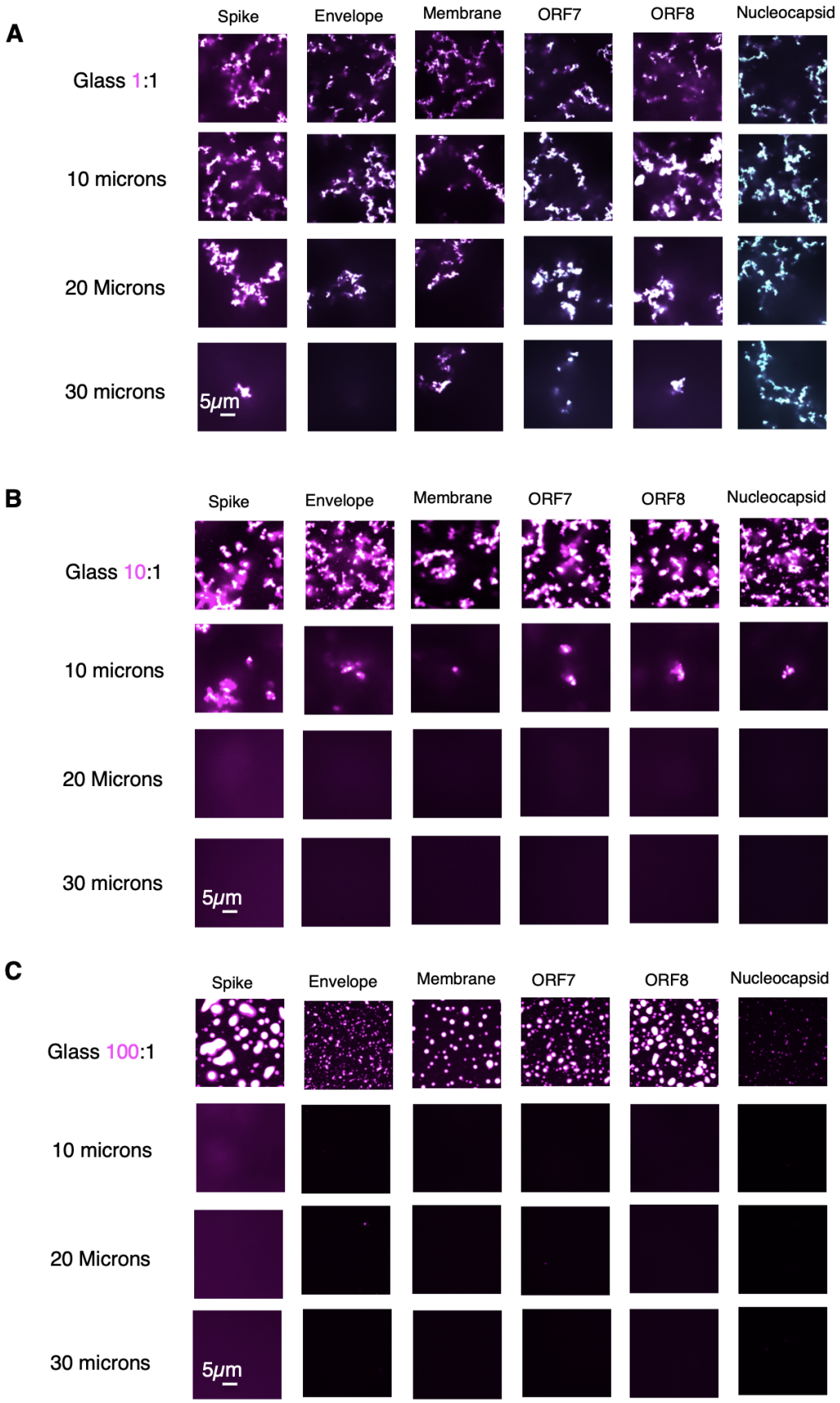
Supplemental Figure S3. Simulations of RNA–RNA contacts, interaction strengths, and dynamics of RNA chains.

Related to Figure 3. (**A**) Relative base pairing probability of intra-chain contacts (lower triangle) vs the inter-chain contacts (upper triangle) in plus RNA simulations. (**B**) Relative base pairing probability of intra-chain contacts (lower triangle) vs the inter-chain contacts (upper triangle) in minus RNA simulations. (**C**) Relative base-pair probability of inter-chain contacts in plus and minus 1:1 simulations. As expected, the contacts are along the diagonal due perfect complementary nature. (**D**) Comparison of the normalized interaction strength for different simulation sizes. The simulations are performed at the same concentration with periodic boundary conditions. The equilibrium potential energy absolute value |U| is normalized by calculating the average potential energy per nucleotide, |U|/N, where N is the number of nucleotides in the system (162x480 or 54x480). There is slight deviation of the values but the overall ordering of plus, minus, and 1:1 plus/minus systems are consistent between different system sizes. (**E**) Comparison of mean-squared displacement (MSD) of RNA chains for different systems sizes. The absolute values vary due to the number chains with 54-chain system moving faster for the same strands. However, the overall ordering is consistent. The pure plus and minus simulations result in an almost indistinguishable diffusion profile, and in both system sizes the movement of chains in the 1:1 plus/minus mixture system are noticeably reduced. (**F**) The percentage of plus–minus interactions in simulations of different sizes. This is obtained by dividing the plus–minus inter-chain base-pairing energies by overall inter-chain base-pairing energies. In both cases, the plus–minus pairing accounts for more than 60% inter-chain base-pairing interactions. (**G**) Average RNA–RNA interaction strength. This is obtained by normalizing the potential energy of each system by the total number of nucleotides. (**H**) Comparison of the MSD of the center of mass of chains. Here, the 1:1 plus/minus mixture exhibits the slowest dynamics for all strands. (**I**) Average number of nucleotides pairs in trans RNA contacts. (**J**) Probability distribution of number of nucleotide pairs in trans RNA contacts. (**K**) Comparison of base pairing ratio between simulations of 1:1 plus/minus systems. These values are obtained by dividing the plus–minus inter-chain base-pairing energies by overall inter-chain base-pairing energies. This ratio is over 50% for all the cases, except for nucleocapsid RNAs. (**L**) The percentage of GC in RNA sequences and base pairing ratio between simulations of 1:1 plus/minus mixtures. These two quantities are overall negatively correlated, indicating that GC pairing plays an important role in 1:1 plus/minus mixture inter-chain pairing.


A
Measure A280
Normalized A280
**
Nucleocapsid
ORF8
ORF7
Membrane
Envelope
Spike
0hr
1hr
B
Spike Pre-Mix
Spike 2hrs
R=0.95
R=0.40
Minus RNA Signal (AU)
Plus RNA Signal (AU)
C
RNA Colocalization (R)
ns
2.5hr
20hr
D
Normalized (0hr) Co-localization
*
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
E
%GC
SARS2 Ends
SARS2 Mid
Pos Ends
Pos Mid
Neg Ends
Neg Mid
dsRNA Ends
dsRNA Mid
Retro Ends
Retro Mid
Human Ends
Human Mid
Predicted G-Quad Locations Plus / Minus (Complement)
F
5′
Orf1ab
M 7 8
3′
Spike
E
N
Max Length 30 Min G-group 2 Loop Size 0-36
G
G-Score
Genome Position
H
G-Score
Genome Position
Mix + strand RNA then protein then – strand after 2hr (imaged 18hr)
I
Merge 0mM Mg2+
Merge 10mM Mg2+
5μm
J
Co-Localization (R)
ns
0mM
10mM
Mg2+

1701 **Supplemental Figure S4. Mixing of time-delayed RNA additions is controlled by GC content**
1702 Related to **Figure 4**.**(A)** A280, diffuse phase measurements of RNA and protein signal taken
1703 from **Figure 4C**. Consistent with plus TRS-B containing fragments being better able to drive N
1704 protein condensation (**Figure 1**, top panel), Reactions with longer incubation with plus RNA only
1705 (1 hour) have lower diffuse phase signal than 0-hour conditions ($p<0.01$ **, $P<0.05$ *). **(B)**
1706 Representative plots (Spike) taken from the colocalization calculation. For the plus (x-axis) and
1707 minus (y-axis) RNA signals for RNA premixed on ice (left panel) and minus RNA mixed 2 hours
1708 post plus and N protein mixing. Premixed RNAs show a higher degree of co-localization with R
1709 values approaching 1. Lack of co-localization is observed by staggering addition and low R
1710 values. **(C)** Related to **Figure 4B** and **4C**. Minor changes for most RNA combinations following
1711 18 hours of further incubation with minus RNA (images taken at 2.5 and 20 hours). No
1712 significant change in recorded values. **(D)** Related to **Figure 4C**. Normalized (0 hour) co-
1713 localization signal at 2.5 hours. The mixing of the highest GC RNA combinations is not
1714 enhanced by additional preincubation of N protein with plus RNA apart from spike. **(E)** Percent
1715 GC content in ends (200nt not including polyA) versus middle of the virus. SARS-CoV-2 shows
1716 higher GC content at ends on average with most RNA viruses and human transcripts showing
1717 the opposite enrichment. **(F)** G-quadruplexes were predicted using QGRS mapper using default
1718 conditions (Max length 30, Min G group 2, Loop size 0-36). Of note, these settings represent the
1719 least stringent possible criteria for g-quadruplex prediction. Pink lines indicate the approximate
1720 location of predicted G-quadruplexes on the plus stand and turquoise indicates the minus strand
1721 (in the complement configuration for easy comparison of duplex position sequences). Black
1722 arrow indicates the only predicted g-quadruplex which overlaps with a tested RNA fragment.
1723 0/12 TRS-B containing fragments in both the plus and minus strand sequence had predicted g-
1724 quadruplexes. **(G)** Quality score for plus and minus strand RNAs of the SARS-CoV-2 genome.
1725 Each bar is labeled with the start position of the indicated g-quadruplex in the plus or minus **(H)**
1726 strand genome. Both plus and minus g-quadruplexes have much lower quality scores than bona
1727 fide g-quadruplex from Orn-1. **(I)** Addition of non-physiologically high levels of magnesium,
1728 which should support g-quadruplex formation, has no obvious impact on plus (magenta) and
1729 minus (cyan) strand RNA co-localization following staggered addition pre-incubation experiment
1730 with N protein (green). Plus strand RNA mixed with N protein followed by 2 hours of incubation
1731 followed by minus strand addition. **(J)** Quantification of the co-localization of the plus and minus
1732 strand RNA co-localization with or without additional Mg^{2+} indicating that there is no significant
1733 difference (ns). Error bars represent technical replicates.

1734
1735
1736
1737
1738


A
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
Glass 1:1
10 microns
20 Microns
30 microns
5μm
B
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
Glass 10:1
10 microns
20 Microns
30 microns
5μm
C
Spike
Envelope
Membrane
ORF7
ORF8
Nucleocapsid
Glass 100:1
10 microns
20 Microns
30 microns
5μm

1739

Figure S5. Arrested networks extend farther into Z than non-arrested networks. Related to Figure 5 **(A)** Representative merged images from 1:1 plus/minus RNA condition at glass (row 1), 10 (row 2), 20 (row 3), and 30 (row 4) microns in Z. All but envelope extends 30 microns in Z in these conditions. **(B)** Representative merged images from 10:1 plus/minus RNA condition at glass (row 1), 10 (row 2), 20 (row 3), and 30 (row 4) microns in Z. All extend 10 microns in Z in these conditions. **(C)** Representative merged images from 100:1 plus/minus RNA condition at glass (row 1), 10 (row 2), 20 (row 3), and 30 (row 4) microns in Z. No tested conditions extend 10 microns in Z in these conditions.

References:

1. Higgs, P.G., and Lehman, N. (2015). The RNA World: Molecular cooperation at the origins of life. Preprint at Nature Publishing Group, https://doi.org/10.1038/nrg3841 https://doi.org/10.1038/nrg3841.
2. Robertson, M.P., and Joyce, G.F. (2012). The origins of the RNA World. Cold Spring Harb Perspect Biol *4*, 1. https://doi.org/10.1101/cshperspect.a003608.
3. Spruijt, E. (2023). Open questions on liquid–liquid phase separation. Preprint at Nature Research, https://doi.org/10.1038/s42004-023-00823-7 https://doi.org/10.1038/s42004-023-00823-7.
4. Rampersad, S., and Tennant, P. (2018). Replication and Expression Strategies of Viruses. In Viruses: Molecular Biology, Host Interactions, and Applications to Biotechnology (Elsevier), pp. 55–82. https://doi.org/10.1016/B978-0-12-811257-1.00003-6.
5. Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., and Snijder, E.J. (2006). Nidovirales: Evolving the largest RNA virus genome. Virus Res *117*, 17–37. https://doi.org/10.1016/j.virusres.2006.01.017.
6. Gulyaeva, A.A., and Gorbalenya, A.E. (2021). A nidovirus perspective on SARS-CoV-2. Biochem Biophys Res Commun *538*, 24–34. https://doi.org/10.1016/j.bbrc.2020.11.015.
7. Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. Nature *584*, 154–156. https://doi.org/10.1038/s41586-020-2368-8.
8. Wang, J., Shi, C., Xu, Q., and Yin, H. (2021). SARS-CoV-2 nucleocapsid protein undergoes liquid–liquid phase separation into stress granules through its N-terminal intrinsically disordered region. Cell Discov *7*, 3–7. https://doi.org/10.1038/s41421-020-00240-3.
9. Wu, Y., Ma, L., Cai, S., Zhuang, Z., Zhao, Z., Jin, S., Xie, W., Zhou, L., Zhang, L., Zhao, J., et al. (2021). RNA-induced liquid phase separation of SARS-CoV-2 nucleocapsid protein facilitates NF-κB hyper-activation and inflammation. Signal Transduct Target Ther *6*. https://doi.org/10.1038/s41392-021-00575-7.
10. Luo, L., Li, Z., Zhao, T., Ju, X., Ma, P., Jin, B., Zhou, Y., He, S., Huang, J., Xu, X., et al. (2021). SARS-CoV-2 nucleocapsid protein phase separates with G3BPs to disassemble stress granules and facilitate viral production. Sci Bull (Beijing) *66*, 1194–1204. https://doi.org/10.1016/j.scib.2021.01.013.

1785 11. Roden, C.A., Dai, Y., Giannetti, C.A., Seim, I., Lee, M., Sealfon, R., McLaughlin, G.A.,
1786 Boerneke, M.A., Iserman, C., Wey, S.A., et al. (2022). Double-stranded RNA drives
1787 SARS-CoV-2 nucleocapsid protein to undergo phase separation at specific temperatures.
1788 Nucleic Acids Res *50*, 8168–8192. https://doi.org/10.1093/nar/gkac596.
1789 12. Savastano, A., Ibáñez de Opakua, A., Rankovic, M., and Zweckstetter, M. (2020).
1790 Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-
1791 containing condensates. Nat Commun *11*. https://doi.org/10.1038/s41467-020-19843-1.
1792 13. Zhao, M., Yu, Y., Sun, L.M., Xing, J.Q., Li, T., Zhu, Y., Wang, M., Yu, Y., Xue, W., Xia,
1793 T., et al. (2021). GCG inhibits SARS-CoV-2 replication by disrupting the liquid phase
1794 condensation of its nucleocapsid protein. Nat Commun *12*, 1–14.
1795 https://doi.org/10.1038/s41467-021-22297-8.
1796 14. Savastano, A., Ibáñez de Opakua, A., Rankovic, M., and Zweckstetter, M. (2020).
1797 Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-
1798 containing condensates. Nat Commun *11*. https://doi.org/10.1038/s41467-020-19843-1.
1799 15. Cubuk, J., Alston, J.J., Incicco, J.J., Singh, S., Stuchell-Brereton, M.D., Ward, M.D.,
1800 Zimmerman, M.I., Vithani, N., Griffith, D., Wagoner, J.A., et al. (2021). The SARS-CoV-
1801 2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Nat
1802 Commun *12*, 1–17. https://doi.org/10.1038/s41467-021-21953-3.
1803 16. Iserman, C., Roden, C.A., Boerneke, M.A., Sealfon, R.S.G., McLaughlin, G.A., Jungreis,
1804 I., Fritch, E.J., Hou, Y.J., Ekena, J., Weidmann, C.A., et al. (2020). Genomic RNA
1805 Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. Mol Cell *80*, 1078–
1806 1091.e6. https://doi.org/10.1016/j.molcel.2020.11.041.
1807 17. Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., Wang, P.H., Song, G., Chen, W.,
1808 and Lou, J. (2020). Liquid–liquid phase separation by SARS-CoV-2 nucleocapsid protein
1809 and RNA. Cell Res *30*, 1143–1145. https://doi.org/10.1038/s41422-020-00408-2.
1810 18. Cai, S., Zhang, C., Zhuang, Z., Zhang, S., Ma, L., Yang, S., Zhou, T., Wang, Z., Xie, W.,
1811 Jin, S., et al. (2023). Phase-separated nucleocapsid protein of SARS-CoV-2 suppresses
1812 cGAS-DNA recognition by disrupting cGAS-G3BP1 complex. Signal Transduct Target
1813 Ther *8*, 170. https://doi.org/10.1038/s41392-023-01420-9.
1814 19. Perdikari, T.M., Murthy, A.C., Ryan, V.H., Watters, S., Naik, M.T., and Fawzi, N.L.
1815 (2020). SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human
1816 hnRNPs. EMBO J *39*, 1–15. https://doi.org/10.15252/embj.2020106478.
1817 20. Seim, I., Roden, C.A., and Gladfelter, A.S. (2021). Role of spatial patterning of N-protein
1818 interactions in SARS-CoV-2 genome packaging. Biophys J *120*, 2771–2784.
1819 https://doi.org/https://doi.org/10.1016/j.bpj.2021.06.018.
1820 21. Iserman, C., Roden, C.A., Boerneke, M.A., Sealfon, R.S.G., McLaughlin, G.A., Jungreis,
1821 I., Fritch, E.J., Hou, Y.J., Ekena, J., Weidmann, C.A., et al. (2020). Genomic RNA
1822 Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. Mol Cell *80*, 1078–
1823 1091.e6. https://doi.org/10.1016/j.molcel.2020.11.041.
1824 22. Jack, A., Ferro, L.S., Trnka, M.J., Wehri, E., Nadgir, A., Nguyenla, X., Fox, D., Costa, K.,
1825 Stanley, S., Schaletzky, J., et al. (2021). SARS-CoV-2 nucleocapsid protein forms
1826 condensates with viral genomic RNA. PLoS Biol *19*, e3001425.
1827 23. Van Treeck, B., and Parker, R. (2018). Emerging Roles for Intermolecular RNA-RNA
1828 Interactions in RNP Assemblies. Cell *174*, 791–802.
1829 https://doi.org/10.1016/j.cell.2018.07.023.

1830 24. Roden, C., and Gladfelter, A.S. (2021). RNA contributions to the form and function of
1831 biomolecular condensates. Nat Rev Mol Cell Biol *22*, 183–195.
1832 https://doi.org/10.1038/s41580-020-0264-6.
1833 25. Roden, C.A., Dai, Y., Giannetti, C.A., Seim, I., Lee, M., Sealfon, R., McLaughlin, G.A.,
1834 Boerneke, M.A., Iserman, C., Wey, S.A., et al. (2022). Double-stranded RNA drives
1835 SARS-CoV-2 nucleocapsid protein to undergo phase separation at specific temperatures.
1836 Nucleic Acids Res *50*, 8168–8192. https://doi.org/10.1093/nar/gkac596.
1837 26. Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., and Snijder, E.J. (2006). Nidovirales:
1838 Evolving the largest RNA virus genome. Virus Res *117*, 17–37.
1839 https://doi.org/10.1016/j.virusres.2006.01.017.
1840 27. Grossoehme, N.E., Li, L., Keane, S.C., Liu, P., Dann, C.E., Leibowitz, J.L., and Giedroc,
1841 D.P. (2009). Coronavirus N Protein N-Terminal Domain (NTD) Specifically Binds the
1842 Transcriptional Regulatory Sequence (TRS) and Melts TRS-cTRS RNA Duplexes. J Mol
1843 Biol *394*, 544–557. https://doi.org/10.1016/j.jmb.2009.09.040.
1844 28. Zúñiga, S., Sola, I., Alonso, S., and Enjuanes, L. (2004). Sequence Motifs Involved in the
1845 Regulation of Discontinuous Coronavirus Subgenomic RNA Synthesis. J Virol *78*, 980–
1846 994. https://doi.org/10.1128/jvi.78.2.980-994.2004.
1847 29. Yang, D., Liu, P., Giedroc, D.P., and Leibowitz, J. (2011). Mouse Hepatitis Virus Stem-
1848 Loop 4 Functions as a Spacer Element Required To Drive Subgenomic RNA Synthesis. J
1849 Virol *85*, 9199–9209. https://doi.org/10.1128/jvi.05092-11.
1850 30. Brian, D.A., and Baric, R.S. (2005). Coronavirus genome structure and replication. Curr
1851 Top Microbiol Immunol *287*, 1–30. https://doi.org/10.1007/3-540-26765-4_1.
1852 31. Stohlman, S.A., Baric, R.S., Nelson, G.N., Soe, L.H., Welter, L.M., and Deans, R.J.
1853 (1988). Specific interaction between coronavirus leader RNA and nucleocapsid protein. J
1854 Virol *62*, 4288–4295. https://doi.org/10.1128/jvi.62.11.4288-4295.1988.
1855 32. Baric, R.S., Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N., and
1856 Stohlman, S.A. (1988). Interactions between coronavirus nucleocapsid protein and viral
1857 RNAs: implications for viral transcription. J Virol *62*, 4280–4287.
1858 https://doi.org/10.1128/jvi.62.11.4280-4287.1988.
1859 33. Sawicki, S.G., and Sawicki, D.L. 79 CORONA VIRUSES USE DISCONTINUOUS
1860 EXTENSION FOR SYNTHESIS OF SUBGENOME-LENGTH NEGATIVE STRANDS.
1861 34. Sola, I., Almazán, F., Zúñiga, S., and Enjuanes, L. (2015). Continuous and Discontinuous
1862 RNA Synthesis in Coronaviruses. Preprint at Annual Reviews Inc.,
1863 https://doi.org/10.1146/annurev-virology-100114-055218 https://doi.org/10.1146/annurev-
1864 virology-100114-055218.
1865 35. Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F., and Miska,
1866 E.A. (2020). The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. Mol
1867 Cell *80*, 1067–1077.e5. https://doi.org/10.1016/j.molcel.2020.11.004.
1868 36. Nicholson, B.L., and White, K.A. (2014). Functional long-range RNA-RNA interactions
1869 in positive-strand RNA viruses. Nat Rev Microbiol *12*, 493–504.
1870 https://doi.org/10.1038/nrmicro3288.
1871 37. Seim, I., Zhang, V., Jalihal, A.P., Stormo, B.M., Cole, S.J., Ekena, J., Nguyen, H.T.,
1872 Thirumalai, D., and Gladfelter, A.S. (2024). RNA encodes physical information. Preprint,
1873 https://doi.org/10.1101/2024.12.11.627970 https://doi.org/10.1101/2024.12.11.627970.

38. Ma, W., Zheng, G., Xie, W., and Mayr, C. (2021). In vivo reconstitution finds multivalent RNA–RNA interactions as drivers of mesh-like condensates. Elife *10*, 1–32. https://doi.org/10.7554/eLife.64252.
39. Bizuayehu, T.T., Labun, K., Jakubec, M., Jefimov, K., Niazi, A.M., and Valen, E. (2022). Long-read single-molecule RNA structure sequencing using nanopore. Nucleic Acids Res *50*, e120–e120. https://doi.org/10.1093/nar/gkac775.
40. Joseph, J.A., Reinhardt, A., Aguirre, A., Chew, P.Y., Russell, K.O., Espinosa, J.R., Garaizar, A., and Collepardo-Guevara, R. (2021). Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. Nat Comput Sci *1*, 732–743. https://doi.org/10.1038/s43588-021-00155-3.
41. Aierken, D., and Joseph, J.A. (2024). Accelerated Simulations Reveal Physicochemical Factors Governing Stability and Composition of RNA Clusters. J Chem Theory Comput. https://doi.org/10.1021/acs.jctc.4c00803.
42. Nguyen, H.T., Hori, N., and Thirumalai, D. (2022). Condensates in RNA repeat sequences are heterogeneously organized and exhibit reptation dynamics. Nat Chem *14*, 775–785. https://doi.org/10.1038/s41557-022-00934-z.
43. Cubuk, J., Alston, J.J., Incicco, J.J., Singh, S., Stuchell-Brereton, M.D., Ward, M.D., Zimmerman, M.I., Vithani, N., Griffith, D., Wagoner, J.A., et al. (2021). The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. Nat Commun *12*, 1–17. https://doi.org/10.1038/s41467-021-21953-3.
44. Alston, J.J., Soranno, A., and Holehouse, A.S. (2024). Conserved molecular recognition by an intrinsically disordered region in the absence of sequence conservation. Biophys J *123*, 26a. https://doi.org/10.1016/j.bpj.2023.11.260.
45. Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: MicroRNA target prediction easy, fast and flexible. Nucleic Acids Res *34*. https://doi.org/10.1093/nar/gkl243.
46. Lan, T.C.T., Allan, M.F., Malsick, L.E., Woo, J.Z., Zhu, C., Zhang, F., Khandwala, S., Nyeo, S.S.Y., Sun, Y., Guo, J.U., et al. (2022). Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. Nat Commun *13*, 1128. https://doi.org/10.1038/s41467-022-28603-2.
47. Tian, S., Nguyen, H., Ye, Z., Rouskin, S., Thirumalai, D., and Trcek, T. (2024). The Folding of Germ Granule mRNAs Controls Intermolecular Base Pairing in Germ Granules and Maintains Normal Fly Development. Preprint, https://doi.org/10.1101/2024.05.31.596852 https://doi.org/10.1101/2024.05.31.596852.
48. Bar-On, Y.M., Flamholz, A., Phillips, R., and Milo, R. (2020). SARS-CoV-2 (COVID-19) by the numbers. Elife *9*. https://doi.org/10.7554/eLife.57309.
49. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: An open-source platform for biological-image analysis. Preprint, https://doi.org/10.1038/nmeth.2019 https://doi.org/10.1038/nmeth.2019.
50. Hyman, A.A., Weber, C.A., and Jülicher, F. (2014). Liquid-liquid phase separation in biology. Annu Rev Cell Dev Biol *30*, 39–58. https://doi.org/10.1146/annurev-cellbio-100913-013325.
51. Zhang, X., Zheng, R., Li, Z., and Ma, J. (2023). Liquid-liquid Phase Separation in Viral Function. J Mol Biol *435*, 167955. https://doi.org/10.1016/j.jmb.2023.167955.

52. Langdon, E.M., and Gladfelter, A.S. (2018). A New Lens for RNA Localization: Liquid-Liquid Phase Separation. Annu Rev Microbiol *72*, 255–271. https://doi.org/10.1146/annurev-micro-090817-062814.
53. Dai, Z., and Yang, X. (2023). The regulation of liquid-liquid phase separated condensates containing nucleic acids. 1–12. https://doi.org/10.1111/febs.16959.
54. Alberti, S., Gladfelter, A., and Mittag, T. (2019). Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. Cell *176*, 419–434. https://doi.org/10.1016/j.cell.2018.12.035.
55. Patel, A., Lee, H.O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M.Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T.M., et al. (2015). A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. Cell *162*, 1066–1077. https://doi.org/https://doi.org/10.1016/j.cell.2015.07.047.
56. Wang, B., Zhang, L., Dai, T., Qin, Z., Lu, H., Zhang, L., and Zhou, F. (2021). Liquid–liquid phase separation in human health and diseases. Signal Transduct Target Ther *6*, 290. https://doi.org/10.1038/s41392-021-00678-1.
57. Adame-Arana, O., Weber, C.A., Zaburdaev, V., Prost, J., and Jülicher, F. (2020). Liquid Phase Separation Controlled by pH. Biophys J *119*, 1590–1605. https://doi.org/https://doi.org/10.1016/j.bpj.2020.07.044.
58. Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science (1979) *357*. https://doi.org/10.1126/science.aaf4382.
59. Zhang, H., Elbaum-Garfinkle, S., Langdon, E.M., Taylor, N., Occhipinti, P., Bridges, A.A., Brangwynne, C.P., and Gladfelter, A.S. (2015). RNA Controls PolyQ Protein Phase Transitions. Mol Cell *60*, 220–230. https://doi.org/10.1016/j.molcel.2015.09.017.
60. Holehouse, A., Alston, J., and Soranno, A. (2024). Conserved molecular recognition by an intrinsically disordered region in the absence of sequence conservation. Preprint, https://doi.org/10.21203/rs.3.rs-4477977/v1 https://doi.org/10.21203/rs.3.rs-4477977/v1.
61. Auewarakul, P. (2005). Composition bias and genome polarity of RNA viruses. Virus Res *109*, 33–37. https://doi.org/10.1016/j.virusres.2004.10.004.
62. Ouizougun-Oubari, M., and Fearns, R. (2024). Structures and Mechanisms of Nonsegmented, Negative-Strand RNA Virus Polymerases. *251*, 47. https://doi.org/10.1146/annurev-virology-111821.
63. Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res *34*. https://doi.org/10.1093/nar/gkl253.
64. Bohálová, N., Cantara, A., Bartas, M., Kaura, P., Šťastný, J., Pečinka, P., Fojta, M., Mergny, J.L., and Brázda, V. (2021). Analyses of viral genomes for G-quadruplex forming sequences reveal their correlation with the type of infection. Biochimie *186*, 13–27. https://doi.org/10.1016/j.biochi.2021.03.017.
65. Qin, G., Zhao, C., Liu, Y., Zhang, C., Yang, G., Yang, J., Wang, Z., Wang, C., Tu, C., Guo, Z., et al. (2022). RNA G-quadruplex formed in SARS-CoV-2 used for COVID-19 treatment in animal models. Cell Discov *8*. https://doi.org/10.1038/s41421-022-00450-x.
66. Romani', A., and Scarpa, A. (1992). Regulation of Cell Magnesium.
67. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus biology and replication: implications for SARS-CoV-2. Preprint at Nature Research, https://doi.org/10.1038/s41579-020-00468-6 https://doi.org/10.1038/s41579-020-00468-6.

68. Bradley, C.C., Wang, C., Gordon, A.J.E., Wen, A.X., Luna, P.N., Cooke, M.B., Kohrn, B.F., Kennedy, S.R., Avadhanula, V., Piedra, P.A., et al. (2024). Targeted accurate RNA consensus sequencing (tARC-seq) reveals mechanisms of replication error affecting SARS-CoV-2 divergence. Nat Microbiol *9*, 1382–1392. https://doi.org/10.1038/s41564-024-01655-4.
69. Liu, J., Xu, Y.P., Li, K., Ye, Q., Zhou, H.Y., Sun, H., Li, X., Yu, L., Deng, Y.Q., Li, R.T., et al. (2021). The m6A methylome of SARS-CoV-2 in host cells. Cell Res *31*, 404–414. https://doi.org/10.1038/s41422-020-00465-7.
70. Yang, S.L., DeFalco, L., Anderson, D.E., Zhang, Y., Aw, J.G.A., Lim, S.Y., Lim, X.N., Tan, K.Y., Zhang, T., Chawla, T., et al. (2021). Comprehensive mapping of SARS-CoV-2 interactions in vivo reveals functional virus-host interactions. Nat Commun *12*. https://doi.org/10.1038/s41467-021-25357-1.
71. Krafcikova, P., Silhan, J., Nencka, R., and Boura, E. (2020). Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. Nat Commun *11*. https://doi.org/10.1038/s41467-020-17495-9.
72. Carrasco-Hernandez, R., Jácome, R., Vidal, Y.L., and de León, S.P. (2017). Are RNA viruses candidate agents for the next global pandemic? A review. ILAR J *58*, 343–358. https://doi.org/10.1093/ilar/ilx026.
73. Rimoin, A.W., Mulembakani, P.M., Johnston, S.C., Smith, J.O.L., Kisalu, N.K., Kinkela, T.L., Blumberg, S., Thomassen, H.A., Pike, B.L., Fair, J.N., et al. Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. https://doi.org/10.1073/pnas.1005769107/-/DCSupplemental.
74. Ishiguro, A., Lu, J., Ozawa, D., Nagai, Y., and Ishihama, A. (2021). ALS-linked FUS mutations dysregulate G-quadruplex-dependent liquid–liquid phase separation and liquid-to-solid transition. Journal of Biological Chemistry *297*. https://doi.org/10.1016/j.jbc.2021.101284.
75. Mcneil, P.L., and Warder, E. (1987). Glass beads load macromolecules into living cells.
76. Hur, S. (2019). Double-Stranded RNA Sensors and Modulators in Innate Immunity. Preprint at Annual Reviews Inc., https://doi.org/10.1146/annurev-immunol-042718-041356 https://doi.org/10.1146/annurev-immunol-042718-041356.
77. Decker, C.J., Burke, J.M., Mulvaney, P.K., and Parker, R. (2022). RNA is required for the integrity of multiple nuclear and cytoplasmic membrane-less RNP granules. EMBO J *41*. https://doi.org/10.15252/embj.2021110137.
78. Wang, X., and Zhu, B. (2024). SARS-CoV-2 nsp15 preferentially degrades AU-rich dsRNA via its dsRNA nickase activity. Nucleic Acids Res *52*, 5257–5272. https://doi.org/10.1093/nar/gkae290.
79. Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. Preprint at Nature Publishing Group, https://doi.org/10.1038/nrm.2016.50 https://doi.org/10.1038/nrm.2016.50.
80. Sommers, J.A., Loftus, L.N., Jones, M.P., Lee, R.A., Haren, C.E., Dumm, A.J., and Brosh, R.M. (2023). Biochemical analysis of SARS-CoV-2 Nsp13 helicase implicated in COVID-19 and factors that regulate its catalytic functions. Journal of Biological Chemistry *299*. https://doi.org/10.1016/j.jbc.2023.102980.
81. Patel, A., Malinovska, L., Saha, S., Wang, J., Alberti, S., Krishnan, Y., and Hyman, A.A. ATP as a biological hydrotrope.

2009 82. Yazdi, A.K., Pakarian, P., Perveen, S., Hajian, T., Santhakumar, V., Bolotokova, A., Li,
2010 F., and Vedadi, M. (2022). Kinetic Characterization of SARS-CoV-2 nsp13 ATPase
2011 Activity and Discovery of Small-Molecule Inhibitors. ACS Infect Dis *8*, 1533–1542.
2012 https://doi.org/10.1021/acsinfecdis.2c00165.
2013 83. Van Treeck, B., Protter, D.S.W., Matheny, T., Khong, A., Link, C.D., and Parker, R.
2014 (2018). RNA self-assembly contributes to stress granule formation and defining the stress
2015 granule transcriptome. Proc Natl Acad Sci U S A *115*, 2734–2739.
2016 https://doi.org/10.1073/pnas.1800038115.
2017 84. Parker, D.M., Tauber, D., and Parker, R. (2024). G3BP1 promotes intermolecular RNA-
2018 RNA interactions during RNA condensation. Mol Cell.
2019 https://doi.org/10.1016/j.molcel.2024.11.012.
2020 85. Hillman, B.I., and Cai, G. (2013). The Family Narnaviridae. Simplest of RNA Viruses. In
2021 Advances in Virus Research (Academic Press Inc.), pp. 149–176.
2022 https://doi.org/10.1016/B978-0-12-394315-6.00006-4.
2023 86. Simmonds, P. (2024). C→U transition biases in SARS-CoV-2: still rampant 4 years from
2024 the start of the COVID-19 pandemic. mBio *15*. https://doi.org/10.1128/mbio.02493-24.
2025 87. Pecori, R., Di Giorgio, S., Paulo Lorenzo, J., and Nina Papavasiliou, F. (2022). Functions
2026 and consequences of AID/APOBEC-mediated DNA and RNA deamination. Nat Rev
2027 Genet *23*, 505–518. https://doi.org/10.1038/s41576-022-00459-8.
2028 88. Srinivasan, S., Torres, A.G., and Ribas de Pouplana, L. (2021). Inosine in biology and
2029 disease. Preprint at MDPI AG, https://doi.org/10.3390/genes12040600
2030 https://doi.org/10.3390/genes12040600.
2031 89. Sun, L., Li, P., Ju, X., Rao, J., Huang, W., Ren, L., Zhang, S., Xiong, T., Xu, K., Zhou, X.,
2032 et al. (2021). In vivo structural characterization of the SARS-CoV-2 RNA genome
2033 identifies host proteins vulnerable to repurposed drugs. Cell *184*, 1865–1883.e20.
2034 https://doi.org/10.1016/j.cell.2021.02.008.
2035 90. Monette, A., Niu, M., Chen, L., Rao, S., Gorelick, R.J., and Mouland, A.J. (2020). Pan-
2036 retroviral Nucleocapsid-Mediated Phase Separation Regulates Genomic RNA Positioning
2037 and Trafficking. Cell Rep *31*, 107520. https://doi.org/10.1016/j.celrep.2020.03.084.
2038 91. Roden, C.A., and Gladfelter, A.S. (2024). Experimental Considerations for the Evaluation
2039 of Viral Biomolecular Condensates. Annual Review of Virology Downloaded from
2040 www.annualreviews.org. Guest *43*, 22. https://doi.org/10.1146/annurev-virology-093022.
2041 92. Cui, J., Schlub, T.E., and Holmes, E.C. (2014). An Allometric Relationship between the
2042 Genome Length and Virion Volume of Viruses. J Virol *88*, 6403–6410.
2043 https://doi.org/10.1128/jvi.00362-14.
2044 93. Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.
2045 V., and Koonin, E. V. (2018). Origins and evolution of the global RNA virome. mBio *9*.
2046 https://doi.org/10.1128/mBio.02329-18.
2047 94. Poranen, M.M., and Bamford, D.H. (2012). Assembly of Large Icosahedral Double-
2048 Stranded RNA Viruses. In Viral Molecular Machines, M. G. Rossmann and V. B. Rao,
2049 eds. (Springer US), pp. 379–402. https://doi.org/10.1007/978-1-4614-0980-9_17.
2050 95. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F.,
2051 and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms for Molecular Biology *6*,
2052 1–14. https://doi.org/10.1186/1748-7188-6-26.
2053 96. Thompson, A.P., Aktulga, H.M., Berger, R., Bolintineanu, D.S., Brown, W.M., Crozier,
2054 P.S., in 't Veld, P.J., Kohlmeyer, A., Moore, S.G., Nguyen, T.D., et al. (2022). LAMMPS

- a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. Comput Phys Commun *271*. https://doi.org/10.1016/j.cpc.2021.108171.

97. Stukowski, A. (2010). Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. Model Simul Mat Sci Eng *18*. https://doi.org/10.1088/0965-0393/18/1/015012.

98. Dinesh, D.C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., and Boura, E. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. PLoS Pathog *16*, 1–16. https://doi.org/10.1371/journal.ppat.1009100.

99. Zinzula, L., Basquin, J., Bohn, S., Beck, F., Klumpe, S., Pfeifer, G., Nagy, I., Bracher, A., Hartl, F.U., and Baumeister, W. (2021). High-resolution structure and biophysical characterization of the nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute respiratory syndrome coronavirus 2. Biochem Biophys Res Commun *538*, 54–62. https://doi.org/10.1016/j.bbrc.2020.09.131.

100. Maity, H.I., Nguyen, H.T., Hori, N.I., Thirumalai, D., and Edited by Gerhard Hummer, I. (2023). Odd-even disparity in the population of slipped hairpins in RNA repeat sequences with implications for phase separation. *120*. https://doi.org/10.1073/pnas.