# The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA

Jasmine Cubuk[1,2], Jhullian J. Alston[1,2], J. Jeremías Incicco[1,2] , Sukrit Singh[1,2], Melissa D. Stuchell-Brereton[1,2], Michael D. Ward[1,2], Maxwell I. Zimmerman[1,2], Neha Vithani[1,2], Daniel Griffith[1,2], Jason A. Wagoner[3], Gregory R. Bowman[1,2], Kathleen B. Hall[1], Andrea Soranno[1,2]*, Alex S. Holehouse[1,2]*

[1]**Department of Biochemistry and Molecular Biophysics**, Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO, 63110, USA
[2]**Center for Science and Engineering of Living Systems (CSELS)**, Washington University in St. Louis, St. Louis, MO 63130, USA
[3]**Laufer Center for Physical and Quantitative Biology**, Stony Brook University, Stony Brook, NY 11794

**Correspondence**: soranno@wustl.edu, alex.holehouse@wustl.edu

**Abstract:**
The SARS-CoV-2 nucleocapsid (N) protein is an abundant RNA binding protein that plays a variety of roles in the viral life cycle including replication, transcription, and genome packaging. Despite its critical and multifunctional nature, the molecular details that underlie how N protein mediates these functions are poorly understood. Here we combine single-molecule spectroscopy with all-atom simulations to uncover the molecular details that contribute to the function of SARS-CoV-2 N protein. N protein contains three intrinsically disordered regions and two folded domains. All three disordered regions are highly dynamic and contain regions of transient helicity that appear to act as local binding interfaces for protein-protein or protein-RNA interactions. The two folded domains do not significantly interact with one another, such that full-length N protein is a flexible and multivalent RNA binding protein. As observed for other proteins with similar molecular features, we found that N protein undergoes liquid-liquid phase separation when mixed with RNA. Polymer models predict that the same multivalent interactions that drive phase separation also engender RNA compaction. We propose a simple model in which symmetry breaking through specific binding sites promotes the formation of metastable single-RNA condensate, as opposed to large multi-RNA phase separated droplets. We speculate that RNA compaction to form dynamic single-genome condensates may underlie the early stages of genome packaging. As such, assays that measure how compounds modulate phase separation could provide a convenient tool for identifying drugs that disrupt viral packaging.

## Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped, positive-strand RNA virus that causes the disease COVID-19 (Coronavirus Disease-2019) (Zhu et al., 2020). While coronaviruses typically cause relatively mild respiratory diseases, COVID-19 is on course to kill half a million people in the first six months since its emergence in late 2019 (Corman et al., 2018; Roser et al., 2020; Zhu et al., 2020). While efforts are ongoing to develop

a vaccine, the timeframe from development to distribution is anticipated to be on the order of months to years (Lurie et al., 2020). An alternative clinical approach is through small-molecule drugs that could attenuate viral function, but we currently lack pharmacological agents that are effective against SARS-CoV-2 (Sanders et al., 2020b). As a result, a major effort is currently underway to identify compounds which could ameliorate viral morbidity and mortality.

A challenge in identifying candidate drugs is our relatively sparse understanding of the molecular details that underlie the function of SARS-CoV-2 proteins. As a result, there is a surge of biochemical and biophysical exploration of these proteins, with the ultimate goal of identifying proteins that are suitable targets for disruption, ideally with insight into the molecular details of how disruption could be achieved (Gordon et al., 2020; Sanders et al., 2020b).

While much attention has been focused on the Spike (S) protein, many other SARS-CoV-2 proteins play equally critical roles in viral physiology, yet we know relatively little about their structural or biophysical properties (Hoffmann et al., 2020; Lan et al., 2020; Shang et al., 2020; Walls et al., 2020). To address this limitation we performed a high-resolution structural and biophysical characterization of the SARS-CoV-2 nucleocapsid (N) protein, the protein responsible for genome packaging (Laude and Masters, 1995; Masters, 2019). A large fraction of N protein is predicted to be intrinsically disordered, which constitutes a major barrier to conventional structural characterization (van der Lee et al., 2014). Here we combined single-molecule spectroscopy with all-atom simulations to build a residue-by-residue description of all three disordered regions in the context of their folded domains. The combination of single-molecule spectroscopy and simulations to reconstruct structural ensembles has been applied extensively to uncover key molecular details underlying disordered protein regions (Borgia et al., 2018; Chung et al., 2015; Dimura et al., 2016; Fuertes et al., 2017; Holmstrom et al., 2019; Warner et al., 2017). Our goal here is to provide biophysical and structural insights into the physical basis of N protein function.

In exploring the molecular properties of N protein, we discovered it undergoes phase separation with RNA, as was also reported recently (Iserman et al., 2020; Perdikari et al., 2020). Given N protein underlies viral packaging, we reasoned phase separation may in fact be an unavoidable epiphenomenon that reflects the physical properties necessary in a protein to drive compaction of long RNA molecules. To explore this principle further, we developed a simple physical model, which suggested symmetry breaking through a small number of high-affinity binding sites can organize anisotropic multivalent interactions to drive single-polymer compaction, as opposed to multi-polymer phase separation. Irrespective of its physiological role, our results suggest that phase separation provides a macroscopic readout (visible droplets) of a nanoscopic process (protein:RNA and protein:protein interaction). In the context of SARS-CoV-2, those interactions are expected to be key for viral packaging, such that an assay that monitors phase separation of N protein may offer a convenient route to identify compounds that will also attenuate viral assembly.

## Results

Coronavirus nucleocapsid proteins are multi-domain RNA binding proteins that play a critical role in many aspects of the viral life cycle (Laude and Masters, 1995; McBride et al., 2014). The SARS-CoV-2 N protein shares a number of sequence features with other nucleocapsid proteins from coronaviruses (**Fig. S1-5**). Work on N protein from a range of model coronaviruses has shown that N protein undergoes both self-association, interaction with other proteins, and interaction with RNA, all in a highly multivalent manner.

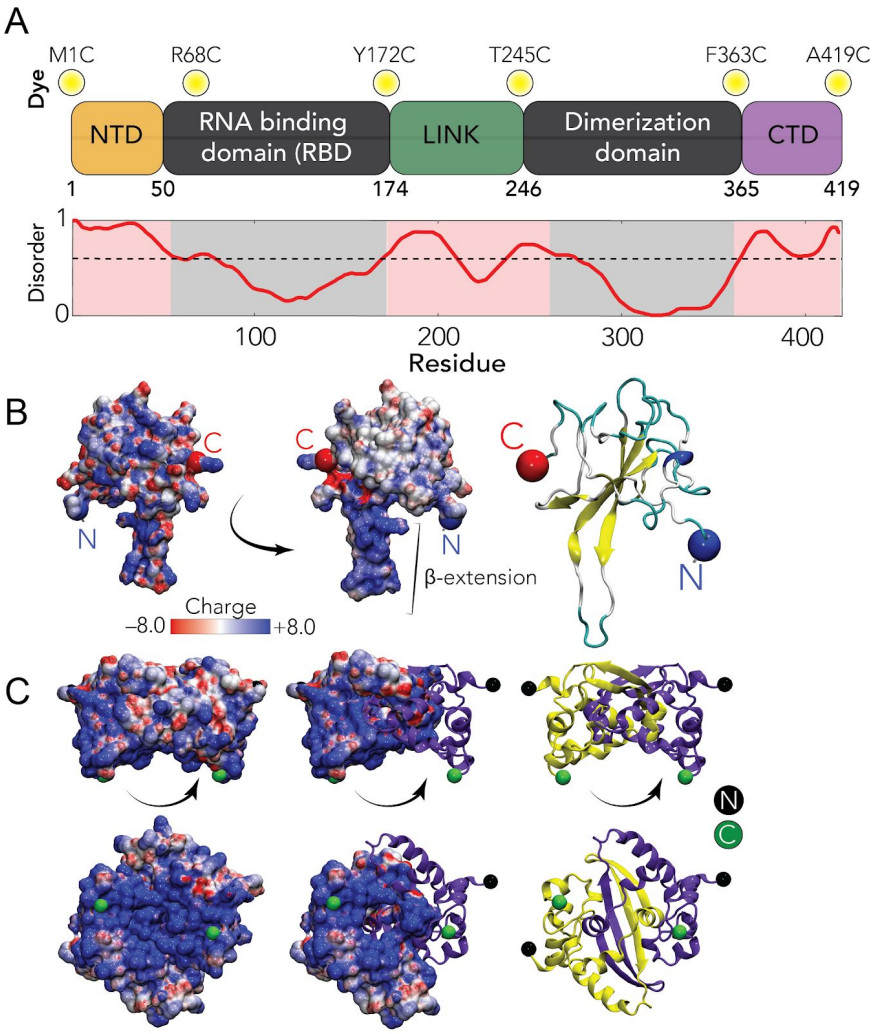


**Figure 1: Sequence and structural summary of N protein**

**A** Domain architecture of the SARS-CoV-2 N protein. Dye positions used in this study are annotated across the top, disorder prediction calculated across the bottom.

**B** Structure of the SARS-CoV-2 RNA binding domain (RBD) (PBD: 6yi3). Center and left: coloured based on surface potential, revealing the highly basic surface of the RBD. Right: ribbon structure with N- and C-termini highlighted.

**C** Dimer structure of the SARS-CoV-2 dimerization domain (PDB: 6yun). Center and left: coloured based on surface potential, revealing the highly basic surface. Right: ribbon structure with N- and C-termini highlighted.

The SARS-CoV-2 N protein can be divided into five domains; a predicted intrinsically disordered N-terminal domain (NTD), an RNA binding domain (RBD), a predicted disordered central linker (LINK), a dimerization domain, and a predicted disordered C-terminal domain (CTD) (**Fig. 1**). While SARS-CoV-2 is a novel coronavirus, decades of work on model coronaviruses (including SARS coronavirus) have revealed a number of features expected to hold true in the SARS-CoV-2 N protein. Notably, all five domains are predicted to bind RNA (Chang et al., 2009; Cui et al., 2015; Grossoehme et al., 2009; Jayaram et al., 2006; Luo et al., 2006; Takeda et al., 2008; Yu et al., 2005), and while the dimerization domain facilitates the formation of well-defined stoichiometric dimers, RNA-independent higher-order oligomerization is also expected to occur (Chang et al., 2013; He et al., 2004a; Robbins et al., 1986; Yu et al., 2005). Importantly, protein-protein and protein-RNA interaction sites have been mapped to all three disordered regions.

Despite recent structures of the RBD and dimerization domains from SARS-CoV-2, the solution-state conformational behaviour of the full-length protein remains elusive (Kang et al., 2020; Zinzula et al., 2020). Understanding N protein function necessitates a mechanistic understanding of the flexible predicted disordered regions and their interplay with the folded domains. A recent small-angle X-ray study shows good agreement with previous work on SARS, suggesting the LINK is relatively extended, but neither the structural basis for this extension nor the underlying dynamics are known (Chang et al., 2009; Zeng et al., 2020a).

Here, we address this question by combining single-molecule fluorescence spectroscopy and all-atom simulations, with the goal of reconstructing an atomistic description of the conformational ensemble of the N protein and a map of its intrachain interactions. To this end, we have created three full-length constructs of the N protein with fluorescent labels (Alexa 488 and 594) flanking the NTD, the LINK, and the CTD (see **Fig. 1A**). The specific positions were selected such that fluorophores are sufficiently close to be in the dynamic range of FRET measurements and sufficiently far apart (in sequence and structure) from tryptophan and tyrosine residues to avoid static and dynamic quenching (Doose et al., 2005; Soranno et al., 2017; Zosel et al., 2017). Labeling was achieved using cysteine mutations in the following positions: M1C and R68C for the NTD, Y172C and T245C for the LINK, and F363C A419C for the CTD (see **Fig. 1A**). These constructs allow us to probe conformations and dynamics of the disordered regions in the context of the full-length protein using single-molecule Förster Resonance Energy Transfer (FRET) and Fluorescence Correlation Spectroscopy (FCS) (see SI for details). In parallel to the experiments, we performed all-atom Monte Carlo simulations of each of the three IDRs in isolation and in context with their adjacent folded domains.

**The NTD is disordered, flexible, and transiently interacts with the RBD**

We started our analysis by investigating the NTD conformations. Under native conditions, single-molecule FRET measurements revealed the occurrence of a single population

with a mean transfer efficiency of 0.61 ± 0.03 (**Fig. 2A** and **Fig. S6**). A quantitative interpretation of this transfer efficiency requires the investigation of protein dynamics to assess whether the transfer efficiency reports about a rigid distance (e.g. structure formation or persistent interaction with the RBD) or is a dynamic average across multiple conformations sampled by the protein during its diffusion across the confocal volume. To this end, we first compare the lifetime of the fluorophores with transfer efficiency. The interdependence of these two factors is expected to be linear if the protein conformations are identical on both timescales (nanoseconds as detected by the fluorescence lifetime, milliseconds as computed from the number of photons in each burst). Alternatively, protein dynamics give rise to a departure from the linear relation and an analytical limit can be computed for configurations rearranging much faster than the burst duration (see SI). Under native conditions, the donor and acceptor lifetimes for the NTD construct lie on the dynamic curved line, indicating that the single population observed in the FRET efficiency histogram is the result of an ensemble in fast conformational exchange (**Fig. S7A**).

To properly quantify the timescale associated with these fast structural rearrangements, we leveraged nanoseconds FCS. As expected for a dynamic population (Nettels et al., 2009; Soranno et al., 2012), the cross-correlation of acceptor-donor photons for the NTD is anticorrelated (**Fig. 2B and S11**). A global fit of the donor-donor, acceptor-acceptor, and acceptor-donor correlations yields a reconfiguration time $t_r$ = 170 ± 30 ns. This is a longer time than the one observed for other proteins with similar persistence length and charge content (Borgia et al., 2012; Schuler et al., 2016; Soranno et al., 2012, 2018), hinting at a large contribution from internal friction due to rapid intrachain contacts (formed either within the NTD or with the RBD) or transient formation of short structural motifs (Soranno et al., 2017).

A conversion from transfer efficiency to chain dimensions can be obtained by assuming the distribution of distances computed from polymer models. Assuming a Gaussian chain distribution yields a root mean square distance between the fluorophores $r_{1\text{-}68}$ of 50 ± 2 Å; when using a SAW model recently proposed by (Zheng et al., 2018) (see SI), we compute a value of $r_{1\text{-}68}$ 48 ± 2 Å. This corresponds to values of persistence length (see SI) $l_p$ equal to 4.9 ± 0.4 Å and 4.5 ± 0.4 Å for Gaussian and SAW distribution, respectively, which are similar to values reported for another unfolded protein under native conditions (Borgia et al., 2012; Schuler et al., 2016; Soranno et al., 2012, 2018). Overall these results confirm the NTD is disordered, as predicted by sequence analysis.

As a next step, we assessed the stability of the folded RBD and its influence on the conformations of the NTD by studying the effect of a chemical denaturant on the protein. The titration with guanidinium chloride (GdmCl) reveals a decrease of transfer efficiencies when moving from native buffer conditions to 1 M GdmCl, followed by a plateau of the transfer efficiencies at concentrations between 1 M and 2 M and a subsequent further decrease at higher concentrations (**Fig. S6** and **S8**). This behavior can be understood assuming that the plateau between 1 M and 2 M GdmCl represents the average of transfer efficiencies between two populations in equilibrium that have very close transfer efficiency and are not resolved because of shot noise. Indeed, this interpretation is supported by a broadening in the transfer

efficiency peak between 1 M and 2 M GdmCl, which is expected if two overlapping populations react differently to denaturant. Besides the effect of the unfolding of the RBD, the dimensions of the NTD are also modulated by change in the solvent quality when adding denaturant **(Fig. 2C ,S6, S8)** and this contribution to the expansion of the chain can be described using an empirical weak-binding model (Aznauryan et al., 2016; Borgia et al., 2016; Hofmann et al., 2012; Schellman, 1987; Zheng et al., 2016). A fit of the interdye root-mean-square distances to this model and the extracted stability of RBD (midpoint: 1.25 ± 0.2 M; $\Delta G_0$ = (3 ± 0.6) RT) is presented in **Fig. 2C** A comparative fit of the histograms with two populations yields an identical result in terms of RBD stability and protein conformations (**Fig. S9**).

These observations provide two important insights. Firstly, the RBD is completely folded under native conditions (**Fig. 2C**). Secondly, the RBD contributes significantly to the conformations of the measured NTD construct, mainly by reducing the accessible space of the disordered tail and favoring expanded configurations, as shown by the shift in transfer efficiency when the RBD is unfolded.

Finally, we tested the effects of electrostatic interactions on the NTD conformational ensemble. Starting from buffer conditions and increasing concentration of KCl, we observed a small but noticeable shift toward lower transfer efficiencies, which represents an expansion of the NTD due to screening of electrostatic interactions. This can be rationalized in terms of the polyampholyte theory of Higgs and Joanny (Higgs and Joanny, 1991; Müller-Späth et al., 2010) (see SI and Table S2), where the increasing concentration of ions screens the interaction between oppositely charged residues (see **Fig. S10**).

To better understand the sequence-dependent conformational behaviour of the NTD we turned to all-atom simulations. We used a sequential sampling approach in which long timescale MD simulations of the RBD in isolation performed on the Folding@home distributed computing platform were first used to generate hundreds of starting conformations (Shirts and Pande, 2000). Those RBD conformations were then used as starting structures for independent all-atom Monte Carlo simulations. Monte Carlo simulations were performed with the ABSINTH forcefield in which the RBD backbone dihedral angles are held fixed but the NTD is fully sampled, as are RBD sidechains (see methods and SI for details). This approach allowed us to construct an ensemble of almost 400 K conformations.

We calculated the equivalent inter-residue distance as measured by FRET, finding excellent agreement (**Fig. 2D**). The peaks on the left side of the histogram reflect specific simulations where the NTD engages more extensively with the RBD through a fuzzy interaction, leading to local kinetic traps (Tompa and Fuxreiter, 2008/1). We also identified several regions of transient helicity that showed perfect agreement between simulations performed for the NTD in isolation vs. simulations performed with the RBD (**Fig. 2E**). We computed normalized distances between the NTD and the folded domain, identifying regions of transient attractive and repulsive interaction, notably the basic beta-strand extension from the RBD (**Fig. 1B**) repels the arginine-rich C-terminal region of the NTD, while a phenylalanine residue (F17) in the NTD

engages with a hydrophobic face on the RBD (**Fig. 2G**). Finally, we noticed the arginine-rich C-terminal residues (residues 31 - 41) form a transient alpha helix projecting three of the four arginines in the same direction (**Fig. 2H**). These features provide molecular insight into previously reported functional observations (see Discussion).

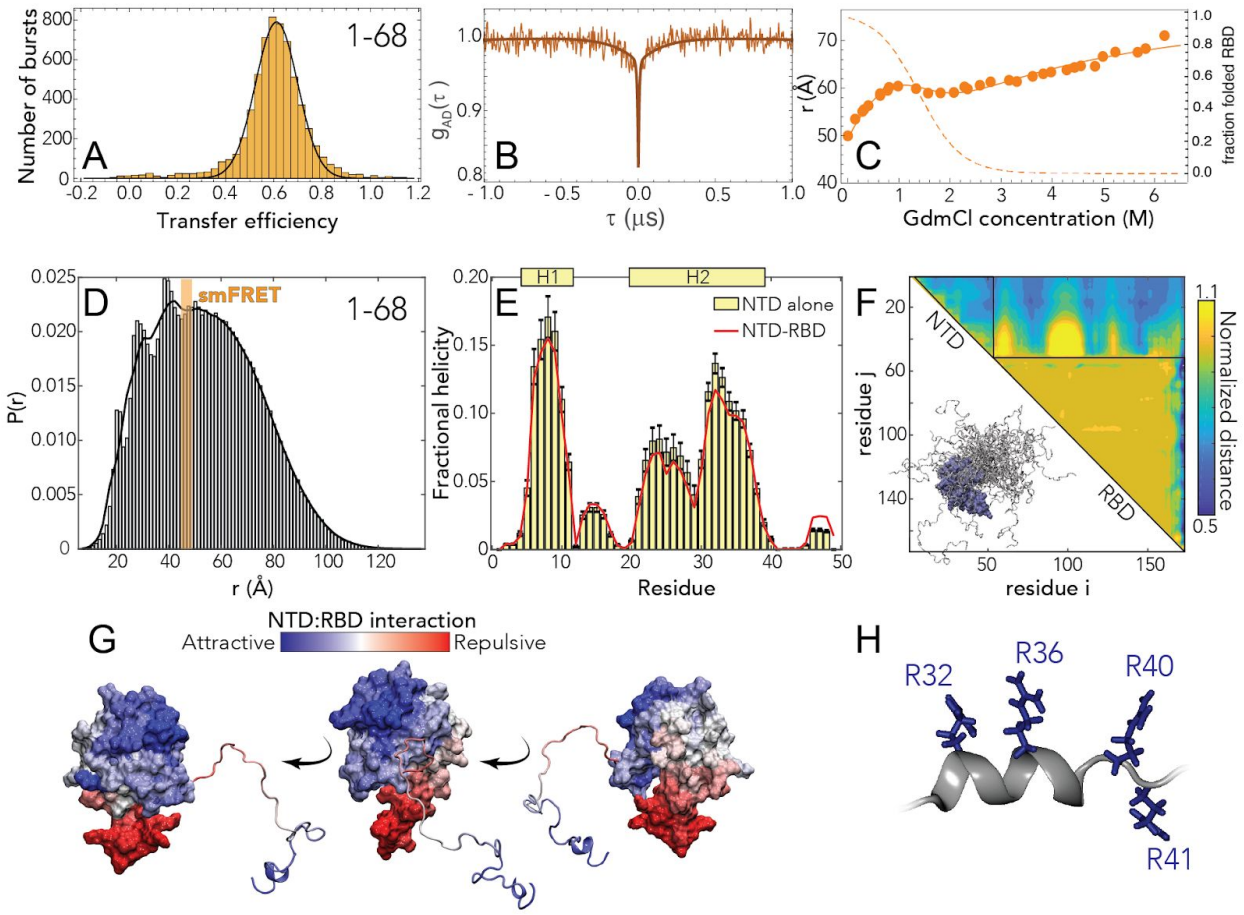


**Figure 2 : The N-terminal domain (NTD) is disordered with residual helical motifs**

**A** Histogram of the transfer efficiency distribution measured across the labeling positions 1 and 68 in the context of the full-length protein, under native conditions (50 mM Tris buffer).

**B** Donor-acceptor crosscorrelation measured by ns-FCS (see SI). The observed anticorrelated rise is the characteristic signature of FRET dynamics and the timescale associated is directly related to the reconfiguration time of the probed segment.

**C** Interdye distance as extracted from single-molecule FRET experiments across different concentrations using a Gaussian chain distribution, examining residues 1-68 in the context of the full length protein. The full line represents a fit to the model in **Eq. S7**, which accounts for denaturant binding (see Table S1) and unfolding of the folded RBD. The dashed line represents the estimate of folded RBD across different denaturant concentrations based on **Eq. S8** .

**D** All-atom simulations of the NTD in the context of RBD. Overall good agreement with smFRET average distance. The peaks on the left shoulder of the histogram are due to persistent NTD-RBD interactions in a small subset of simulations.

**E** Transient helicity in the NTD in isolation or in the context of the RBD. Perfect profile overlap suggests interaction between the NTD and the RBD does not lead to a loss of helicity.
**F** Scaling maps quantify heterogeneous interaction between each pair of residues in terms of average distance normalized by distance expected for the same system if the IDR had no attractive interactions (the "excluded volume" limit (Holehouse et al., 2015)). Both repulsive (yellow) and attractive (blue) regions are observed for NTD-RBD interactions.
**G** Projection of normalized distances onto the folded domain reveals repulsion is through electrostatic interaction (positively charged NTD is repelled by the positive face of the RBD, which is proposed to engage in RNA binding) while attractive interactions are between positive, aromatic, and polar residues in the NTD and a slightly negative and hydrophobic surface on the RBD (see **Fig. 1B**, center).
**H** The C-terminal half of transient helicity in H2 encodes an arginine-rich surface.

**The linker is highly dynamic and there is minimal interaction between the RBD and the dimerization domain**

We next turned to the linker (LINK) construct to investigate how the disordered region modulates the interaction and dynamics between the two folded domains. Under native conditions (50 mM Tris buffer), single-molecule FRET reveals a narrow population with mean transfer efficiency of 0.52 ± 0.03. Comparison of the fluorescence lifetime and transfer efficiency indicates that, like the NTD, the transfer efficiency represents a dynamic conformational ensemble sampled by the LINK (Fig. **S7B**). ns-FCS confirms fast dynamics with a characteristic reconfiguration time $t_r$ of 120 ± 20 ns (**Fig. 3B** and **S11**). This reconfiguration time is compatible with high internal friction effects, as observed for other unstructured proteins (Borgia et al., 2012; Soranno et al., 2012), though usually less charged, but may also account for the drag of the surrounding domains. Overall the lifetime vs transfer efficiency dependence and the reconfiguration time of the chain supports evidence for dynamics occurring on a timescale much faster then residence time of the molecule in the confocal volume. Operating under this assumption, we compute the root-mean-square distance adopted by the LINK. $r_{172\text{-}245}$ is equal to 57 ± 2 Å ($l_p$ = 5.8 ± 0.4 Å) when assuming a Gaussian Chain model as a distribution for the conformational ensemble, and 53 ± 2 Å ($l_p$ = 5.1 ± 0.4 Å) when using a SAW model (see SI).

Next, we addressed whether the LINK segment populates elements of persistent secondary structure or forms stable interaction with the RBD or dimerization domains. Addition of the denaturant shows a continuous shift of the transfer efficiency toward lower values (**Fig. S6,S8**), that corresponds to an almost linear expansion of the chain (see **Fig. 3C**). These observations support a model in which LINK is unstructured and flexible and do not reveal a significant fraction of folding or persistent interactions with or between folded domains.

Finally, we investigate how salt affects the conformations of the LINK. Interestingly, we find a negligible effect of salt screening on the root mean square distance $r_{172\text{-}245}$ as measured by FRET (see **Fig. S10**). Predictions of the Higgs & Joanny theory (see SI) for the content of negative and positive charges within the LINK construct indicates a variation of interdye distance dimension that is comparable with the measurement error. It has to be noted that in this case the excluded volume term in the Higgs and Joanny theory will empirically account not

only for the excluded volume of the amino acids in the chain, but also for the excluded volume occupied by the two folded domains.

Overall, our single-molecule observations report a relatively extended average inter-domain distance, suggesting a low number of interactions between folded domains. To further explore this conclusion, we turned again to Monte Carlo simulations of the same region.

As with the NTD, all-atom Monte Carlo simulations provide atomistic insight that can be compared with our spectroscopic results. Given the size of the system (318 residues for the RBD-LINK-DIM construct) and the inherent combinatorial challenges of sampling two different folded domains via molecular dynamics to generate seeds, we opted to use a single starting seed structure for the folded domains based on the NMR and crystal-structure conformations for the RBD and dimerization domains, respectively. We again performed multiple independent all-atom Monte Carlo simulations in which the LINK is fully sampled while the backbone dihedrals for the folded domains were held fixed. We also performed simulations of the LINK in isolation.

We found good agreement between simulations and experiment (**Fig. 3D**). The root mean square inter-residue distance between simulated positions 172 and 245 is 59.1 Å, which is within the experimental error of the single-molecule observations. Normalized distance map shows a number of regions of repulsion, notably that the RBD repels the N-terminal part of the LINK and the dimerization domain repels the C-terminal part of the LINK (**Fig. 3E**). We tentatively suggest this may reflect sequence properties chosen to prevent aberrant interactions between the LINK and the two folded domains. In the LINK-only simulations we identified two regions that form transient helices at low populations (20-25%), although these are much less prominent in the context of the full length protein (**Fig. 3F)**. Those helices encompass a serine-arginine (SR) rich region known to mediate both protein-protein and protein-RNA interaction, and leads to the alignment of three arginine residues along one face of a helix. The second helix (H4) is a leucine/alanine-rich hydrophobic helix which may contribute to oligomerization, or act as a helical recognition motif for other protein interactions (notably as a nuclear export signal for Crm1, see Discussion).

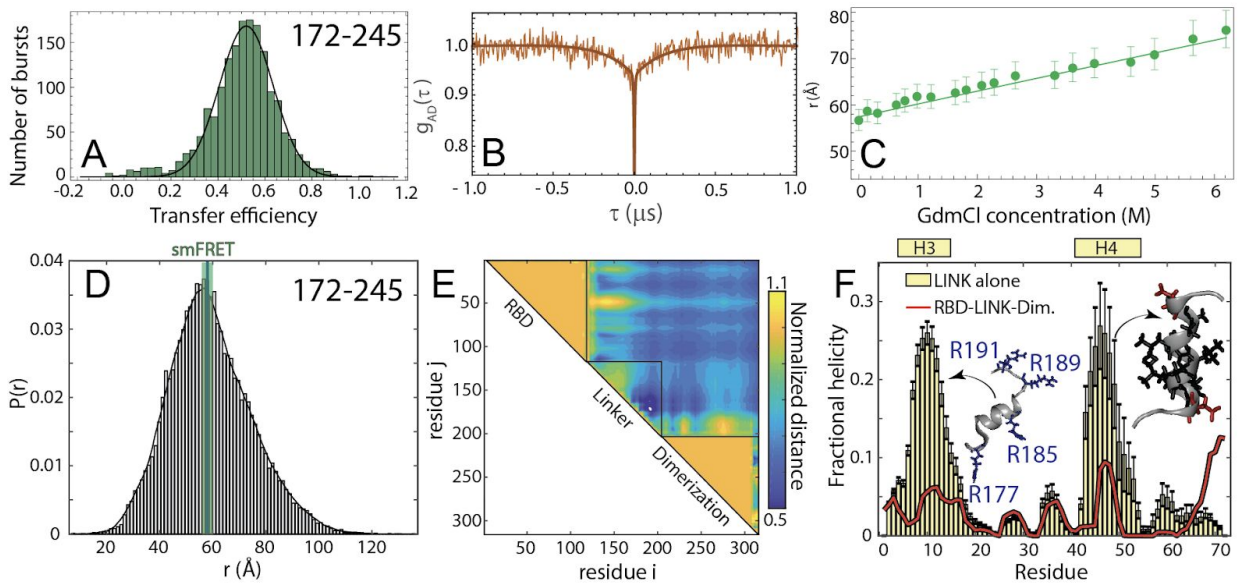


**Figure 3: The RNA binding domain (RBD) and dimerization domains do not significantly interact and are connected by a disordered linker (LINK)**

**A** Histogram of the transfer efficiency distribution measured across the labeling positions 172 and 245 in the context of the full-length protein, under native conditions (50 mM Tris buffer).

**B** Donor-acceptor crosscorrelation measured by ns-FCS (see SI). The observed anticorrelated rise is the characteristic signature of FRET dynamics and the timescale associated is directly related to the reconfiguration time of the probed segment.

**C** Interdye distance as extracted from single-molecule FRET experiments across different denaturant concentrations. The full line represents a fit to the model in **Eq. S6**, which accounts for denaturant binding.

**D** Inter-residue distance distributions calculated from simulations (histogram) show good agreement with distances inferred from single-molecule FRET measurements (green bar).

**E** Scaling maps reveal repulsive interactions between the N- and C-terminal regions of the LINK with the adjacent folded domains. We also observe relatively extensive intra-LINK interactions around helix H4 (see **Fig. 3F**).

**F** Two transient helices are observed in the linker. The N-terminal helix H3 overlaps with part of the SR-region and orientates three arginine residues in the same direction, analogous to behaviour observed for H2 in the NTD. The C-terminal helix H4 overlaps with a Leu/Ala rich motif which we believe is a conserved nuclear export signal (see Discussion).

## The CTD engages in transient but non-negligible interactions with the dimerization domain

Finally, we turned to the CTD. Single-molecule FRET experiments again reveal a single population with a mean transfer efficiency of 0.59 ± 0.03 (**Fig. 4A**) and the denaturant dependence follows the expected trend for a disordered region, with a shift of the transfer efficiency toward lower values (**Fig. S6** and **S8**), from 0.59 to 0.35. Interestingly, when studying the denaturant dependence of the protein, we noticed that the width of the distribution increases while moving toward native conditions. This suggests that the protein may form transient contacts or adopt local structure. To investigate this aspect, we turned first to the investigation

of the dynamics. Though the comparison of the fluorophore lifetimes against transfer efficiency (**Fig. S7c**) appears to support a dynamic nature underlying this population, nanosecond FCS reveals a flat acceptor-donor cross-correlation on the μs timescale (**Fig. 4B**). However, inspection of the donor-donor and acceptor-acceport autocorrelations reveal a correlated decay with a characteristic time of 240 ± 50 ns. This is different from what expected for a completely static system such as polyprolines (Nettels et al., 2007), where the donor-donor and acceptor-acceptor autocorrelation are also flat. An increase in the autocorrelation can be observed for static quenching due to the interaction of the dye with aromatic residues, and correlations of the donor only population reveals a decay time of 230 ± 50 ns, supporting the hypothesis that the donor-acceptor crosscorrelation of the CTD FRET population contains a contribution due to donor quenching. Interestingly, donor dye quenching can also contribute to a positive amplitude in the acceptor-acceptor and donor-acceptor correlation (Haenni et al., 2013; Sauer and Neuweiler, 2014). Therefore, a plausible interpretation of the flat cross-correlation data is that we are observing two populations in equilibrium whose correlations (one anticorrelated, reflecting conformational dynamics, and one correlated, reflecting quenching due contact formation) compensate each other.

To further investigate the possible coexistence of these different species, we performed ns-FCS at 0.2 M GdmCl, where the width of the FRET population starts decreasing and the mean transfer efficiency is slightly shifted to larger values, under the assumption that the decreased width of the population reflects reduced interactions. Indeed, the cross-correlation of ns-FCS reveals a dynamic behavior with a reconfiguration time $t_r$ = 70 ± 15 ns (**Fig. S11**). Based on these observations, we suggest that a very similar disordered population as the one observed at 0.2 M is also present under native conditions, but in equilibrium with a quenched species that forms long-lived contacts. Under the assumption that the mean transfer efficiency still originates (at least partially) from a dynamic distribution, the estimate of the inter-residue root-mean-square distance is $r_{363\text{-}419}$ = 51 ± 2 Å ($l_p$ = 6.1 ± 0.4 Å) for a Gaussian chain distribution and $r_{363\text{-}419}$ = 48 ± 2 Å ($l_p$ = 5.4 ± 0.4 Å) for the SAW model (see SI). However, some caution should be used when interpreting these numbers since we know there is some contribution from fluorophore quenching, which may in turn contribute to an underestimate of the effective transfer efficiency (Zosel et al., 2017).

Finally, we test if the addition of salt can provide similar effects than those obtained by GdmC: interestingly, we do not observe any significant variation either in transfer efficiency or distribution width (Fig S.10), suggesting that the broadening of the population does not originate from electrostatic interactions.

Having performed single-molecule FRET experiments, we again turned to Monte Carlo simulations to examine the interplay between the CTD and dimerization domain. We found good agreement between simulations and experiment (**Fig. 4D**). We identified two transient helices, one (H5) is minimally populated but the second (H6) is more highly populated in the IDR-only simulation and still present at ~20% in the folded state simulations (**Fig. 4E**). The difference

reflects the fact that several of the helix-forming residues interacted with the dimerization domain, leading to a competition between helix formation and intramolecular interaction. Scaling maps revealed extensive intramolecular interaction by the residues that make up H6, both in terms of local intra-IDR interactions and interaction with the dimerization domain (**Fig. 4F**). Mapping normalized distances onto the folded structure revealed that interactions occurred primarily with the N-terminal portion of the dimerization domain (**Fig. 4G**). As with the LINK and the NTD, a positively charged set of residues immediately adjacent to the folded domain in the CTD drive repulsion between this region and the dimerization domain. H6 is the most robust helix observed across all three IDRs, and is a perfect amphipathic helix with a hydrophobic surface on one side and charged/polar residues on the other (**Fig. 4H**). The cluster of hydrophobic residues in H6 engage in intramolecular contacts and offer a likely physical explanation for the complex lifetime data.

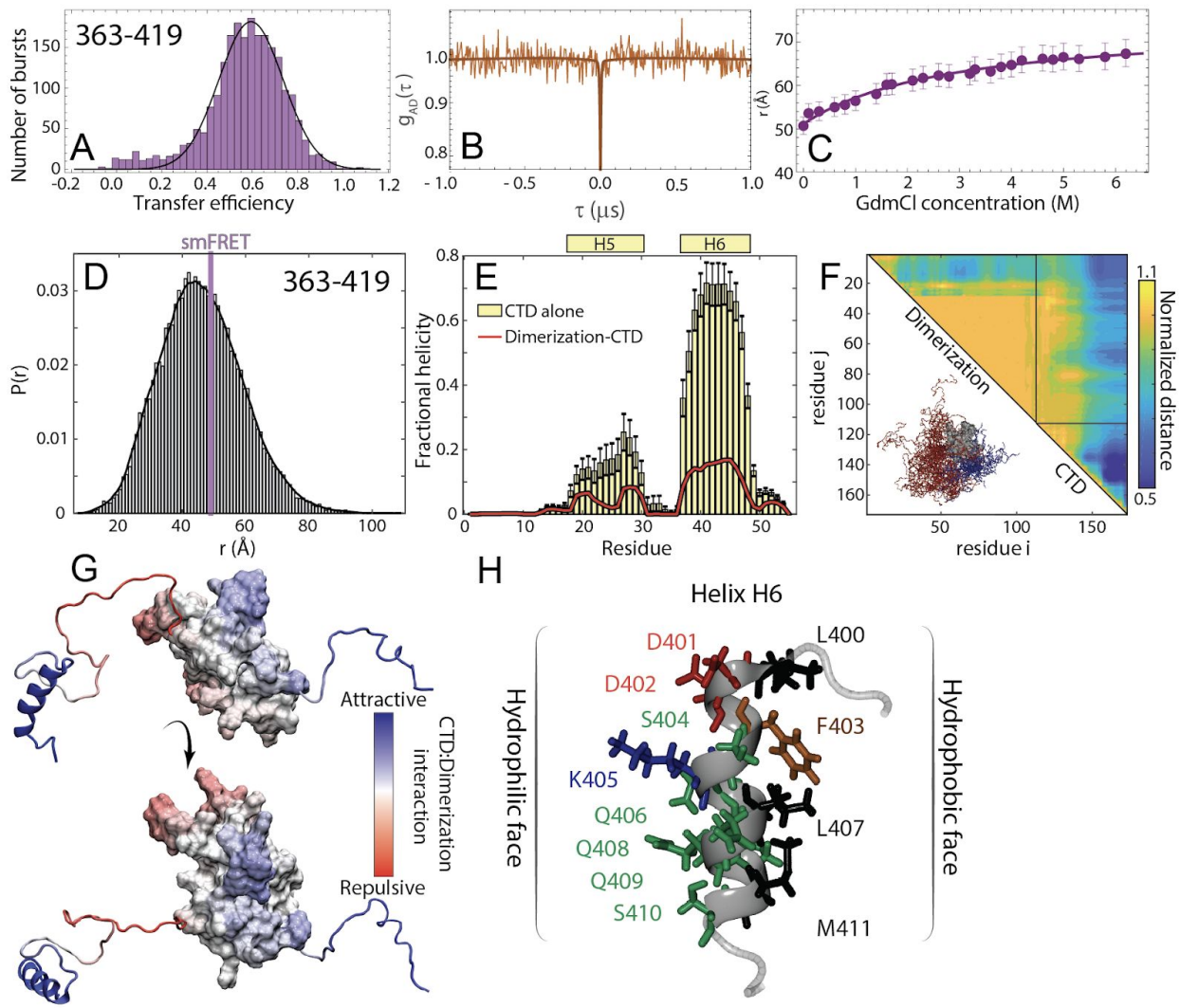


**Figure 4: The C-terminal domain (CTD) is disordered, engages in transient interaction with the dimerization domain, and contains a putative helical binding motif**

**A** Histogram of the transfer efficiency distribution measured across the labeling positions 363 and 419 in the context of the full-length protein, under native conditions (50 mM Tris buffer).
**B** Donor-acceptor crosscorrelation measured by ns-FCS (see SI). The flat correlation indicates a lack of dynamics in the studied timescale or the coexistence of two populations in equilibrium whose correlations (one correlated and the other anticorrelated) compensate each other.
**C** Interdye distance as extracted from single-molecule FRET experiments across different denaturant concentrations. The full line represents a fit to the model in **Eq. S6**, which accounts for denaturant binding.
**D** Inter-residue distance distributions calculated from simulations (histogram) show good agreement with distances inferred from single-molecule FRET measurements (purple bar)
**E** Two transient helices (H5 and H6) are observed in the CTD. Both show a reduction in population in the presence of the dimerization domain at least in part because the same sets of residues engage in transient interactions with the dimerization domain.
**F** Normalized contacts maps describe the average inter-residue distance between each pair of residues, normalized by the distance expected if the CTD behaved as a self-avoiding random coil. H6 engages in extensive intra-CTD interactions and also interacts with the dimerization domain. We observe repulsion between the dimerization domain and the N-terminal region of the CTD.
**G** The normalized distances are projected onto the surface to map CTD-dimerization interaction. The helical region drives intra-molecular interaction, predominantly with the N-terminal side of the dimerization domain.
**H** Helix H6 is an amphipathic helix with a polar/charged surface (left) and a hydrophobic surface (right).

**N protein undergoes phase separation with RNA**

Over the last decade, biomolecular condensates formed through phase separation have emerged as a new mode of cellular organization (Banani et al., 2017; Brangwynne et al., 2009; Li et al., 2012; Shin and Brangwynne, 2017). Given the high interaction valency and the presence of molecular features similar to other proteins we had previously studied, we anticipated that N protein would undergo phase separation with RNA (Boeynaems et al., 2019; Guillén-Boixet et al., 2020; Martin et al., 2020; Pak et al., 2016; Wang et al., 2018).

In line with this expectation, we observed robust droplet formation with homopolymeric RNA (**Fig. 5A-B**) under native buffer conditions (50 mM Tris) and at higher salt concentration (50 mM NaCl). Turbidity assays at different concentrations of protein and poly(rU) demonstrate the classical reentrant phase behaviour expected for a system undergoing heterotypic interaction (**Fig. 5C-D**). It is to be noted that turbidity experiments do not exhaustively cover all the conditions for phase-separation and are only indicative of the regime explored in the current experiments. In particular, turbidity experiments do not provide a measurement of tie-lines, though they are inherently a reflection of the free energy and chemical potential of the solution mixture (Stockmayer, 1950). Interestingly, phase-separation occurs at relatively low concentrations, in the low μM range, which are compatible with physiological concentration of the protein and nucleic acids. Though increasing salt concentration results in an upshift of the phase boundaries, one has to consider that in a cellular environment this effect might be counteracted by cellular crowding.

One peculiar characteristic of our measured phase-diagram is the narrow regime of conditions in which we observe phase-separation of non-specific RNA at a fixed concentration of protein. This leads us to hypothesize that the protein may have evolved to maintain a tight control of concentrations at which phase-separation can (or cannot) occur. Interestingly, when rescaling the turbidity curves as a ratio between protein and RNA, we find all the curve maxima aligning at a similar stoichiometry, approximately 20 nucleotides per protein in absence of added salt and 30 nucleotides when adding 50 mM NaCl (**Fig. S12**). These ratios are in line with the charge neutralization criterion proposed by Banerjee *et al.*, since the estimated net charge of the protein at pH 7.4 is +24 (Banerjee et al., 2017). Finally, given we observed phase separation with poly(rU), the behaviour we are observing is likely driven by relatively non-specific protein:RNA interactions. In agreement, work from the Gladfelter (Iserman et al., 2020), Fawzi (Perdikari et al., 2020) and Yildiz (unpublished) labs has also established this phenomenon across a range of solution conditions and RNA types.

As mentioned, the phase boundaries identified in our turbidity experiments are not completely representative of physiological conditions, since there is no account of the contribution of cellular crowding and other interaction partners. Having established phase separation through a number of assays, we wondered what -if any- physiological relevance this may have for the normal biology of SARS-CoV-2.

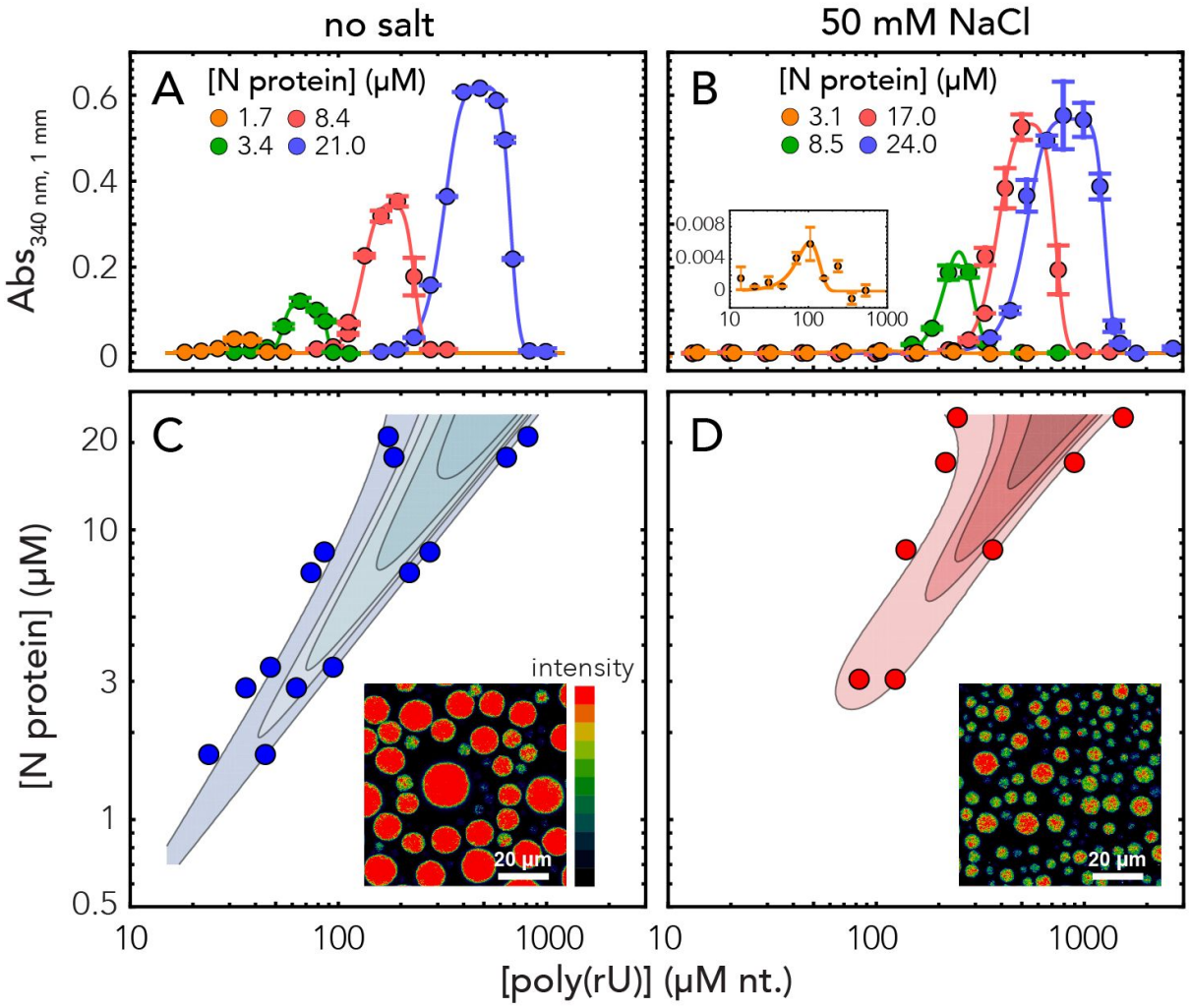


**Figure 5 - Nucleocapsid protein undergoes phase separation with RNA**

**A-B.** Appearance of solution turbidity upon mixing was monitored to determine the concentration regime in which N protein and poly(rU) undergo phase separation. Representative turbidity titrations with poly(rU) in 50 mM Tris, pH 7.5 (HCl) at room temperature, in absence of added salt (A) and in presence of 50 mM NaCl (B), at the indicated concentrations of N protein. Points and error bars represent the mean and standard deviation of 2-4 consecutive measurements from the same sample. Solid lines are simulations of an empirical equation fitted individually to each titration curve (see SI). An inset is provided for the titration at 3.1 μM N protein in 50 mM NaCl to show the small yet detectable change in turbidity on a different scale.

**C-D.** Projection of phase boundaries for poly(rU) and N protein mixtures highlights a re-entrant behavior, as expected for phase-separations induced by heterotypic interactions. Turbidity contour lines are computed from a global fit of all titrations curves (see SI).

Insets: confocal fluorescence images of droplets doped with fluorescently labeled N protein. Total concentrations are 22 μM N protein, 0.5 nM labeled N protein and 0.54 μM nt. poly(rU). At higher salt concentration, a lower concentration of protein in the droplet is detected.

**A simple polymer model shows symmetry-breaking can facilitate multiple metastable single-polymer condensates instead of a single multi-polymer condensate**

Why might phase separation of N protein with RNA be advantageous to SARS-CoV-2? One possible model is that large, micron-sized cytoplasmic condensates of N protein with RNA form through phase separation and play a role in genome packaging. These condensates may act as molecular factories that help concentrate the components for pre-capsid assembly (where we define a pre-capsid here simply as a species that contains a single copy of the genome with multiple copies of the associated N protein). Such a model has been proposed in elegant work on measles virus, in which the measles-virus nucleoprotein and phosphoprotein phase separate together to accelerate the formation of the helical capsid state inside condensates (Guseva et al., 2020). Moreover, phase separation has been invoked in the context of several other viruses, with various proposed functions (Heinrich et al., 2018; Metrick et al., 2020; Nikolic et al., 2017; Zhou et al., 2019).

However, given that phase separation is unavoidable when high concentrations of multivalent species are combined, we propose that an alternative interpretation of our data is that in this context, phase separation is simply an inevitable epiphenomenon that reflects the inherent multi-valency of the N protein for itself and for RNA. This poses questions about the origin of specificity for viral genomic RNA (gRNA), and, of focus in our study, how phase separation might relate to a single genome packaging through RNA compaction.

Given the expectation of a single genome per virion, we reasoned SARS-CoV-2 may have evolved a mechanism to limit phase separation with gRNA (i.e. to avoid multi-genome condensates), with a preference instead for single-genome packaging (single-genome condensates). This mechanism may exist in competition with the intrinsic phase separation of the N protein with other non-specific RNAs (nsRNA).

One possible way to limit phase separation between two components (e.g. gRNA/nsRNA and N protein) is to ensure the levels of these components are held at a sufficiently low total concentration such that the phase boundary is never crossed. While certainly possible, this behaviour must contend with a number of hard-to-control variables. These variables include the impact of temperature, the role of other RNA binding proteins, and other cellular RNAs, all of which could substantially shift the phase boundary in complex ways (Posey et al., 2018; Riback et al., 2020; Sanders et al., 2020a). Most importantly, we need to consider conditions in which not only is phase-separation prevented but gRNA compaction is promoted through binding of N protein. In this scenario, the affinity between gRNA and N protein plays a central role in determining the required concentration for condensation of the macromolecule (gRNA) by the ligand (N protein).

Given a defined valence of the system components, phase boundaries are encoded by the strength of interaction between the interacting domains in the components. Considering a long polymer (e.g. gRNA) with proteins adsorbed onto that polymer as adhesive points (“stickers”), the physics of associative polymers predicts that the same interactions that cause

phase separation will also control the condensation of individual long polymers (Choi et al., 2019, 2020; Martin et al., 2020; Post and Zimm, 1979; Rubinstein and Colby, 2003; Semenov and Rubinstein, 1998). With this in mind, we hypothesized that phase separation is reporting on the physical interactions that underlie genome compaction.

To explore this hypothesis, we developed a simple computational model where the interplay between compaction and phase separation could be explored. Our setup consists of two types of species: long multivalent polymers and short multivalent binders (**Fig. 6A**). All interactions are isotropic and each bead is inherently multivalent as a result. In the simplest instantiation of this model, favourable polymer:binder and binder:binder interactions are encoded, mimicking the scenario in which a binder (e.g. a protein) can engage in non-specific polymer (RNA) interaction as well as binder-binder (protein-protein) interaction.

Simulations of binder and polymer undergo phase separation in a concentration-dependent manner, as expected (**Fig. 6B,C**). Phase separation gives rise to a single large spherical cluster with multiple polymers and binders (**Fig. 6D, 6H**). The formation of a single large droplet is the thermodynamic minimum and the expected end state. Importantly, no compaction of the long polymers is observed in the one-phase regime with the model parameters used here. For a homopolymer the balance of chain-compaction and phase separation is determined in part due to chain length and binder $K_d$. In our system the polymer is largely unbound in the one-phase regime (suggesting the concentration of ligand in the one-phase space is below the $K_d$) but entirely coated in the two-phase regime, consistent with highly-cooperative binding behaviour. In the limit of long, multivalent polymers with multivalent binders, the sharpness of the coil-to-globule transition is such that an effective two-state description of the chain emerges, in which the chain is *either* expanded (non-phase separation-competent) OR compact (coated with binders, phase separation competent).

In light of these observations, we wondered if a break in the symmetry between intra- and inter-molecular interactions would be enough to promote single-polymer condensation in the same concentration regime over which we had previously observed phase-separation. This hypothesis was motivated by the fact that many viruses (including coronaviruses) contain viral packaging signals that confer high-affinity binding (Cologna and Hogue, 2000; Fosmire et al., 1992; Hsieh et al., 2005; Stockley et al., 2013; Woo et al., 1997). Therefore, in keeping with the spirit of simple models, we added a single high-affinity multivalent binding site to the center of our polymer (**Fig. 6A**). While our simple model cannot begin to approach the real complexity of the genome packaging problem, we felt providing the right ‘class’ of symmetry breaking feature would be sensible.

We performed identical simulations to those in **Fig. 6C-D** using the same system with polymers that now possess a single high affinity binding site (**Fig. 6E**). Under these conditions and using the same simulation parameters, we did not observe large phase separated droplets (**Fig. 6F**). Instead, each individual polymer undergoes collapse to form a single-polymer condensate. Collapse is driven by the recruitment of binders to the high-affinity site, where they

“coat” the chain, forming a local cluster of binders on the polymer. This cluster is then able to interact with the remaining regions of the polymer through weak “nonspecific” interactions, the same interactions that drove phase separation in **Fig. 6 B,C,D**. Symmetry breaking is achieved because the local concentration of binder around the site is high, such that intramolecular interactions are favoured over intermolecular interaction. This high local concentration also drives compaction at low binder concentrations. As a result, instead of a single multi-polymer condensate, we observe multiple single-polymers condensates, where the absolute number matches the number of polymers in the system (**Fig. 6G**).

Our results can also be cast in terms of two distinct concentration (phase) boundaries - one for binder:high affinity site interaction ($c_1$), and a second boundary for “nonspecific” binder:polymer interactions ($c_2$) at a higher concentration. $c_2$ reflects the boundary observed in **Fig. 6C** that delineated the one and two-phase regimes. At global concentrations below $c_2$, (but above $c_1$) the clustering of binders at a high affinity site raises the apparent local concentration of binders above $c_2$, from the perspective of other beads on the chain. In this way, a local high affinity binding site can drive “local” phase separation of a single polymer.

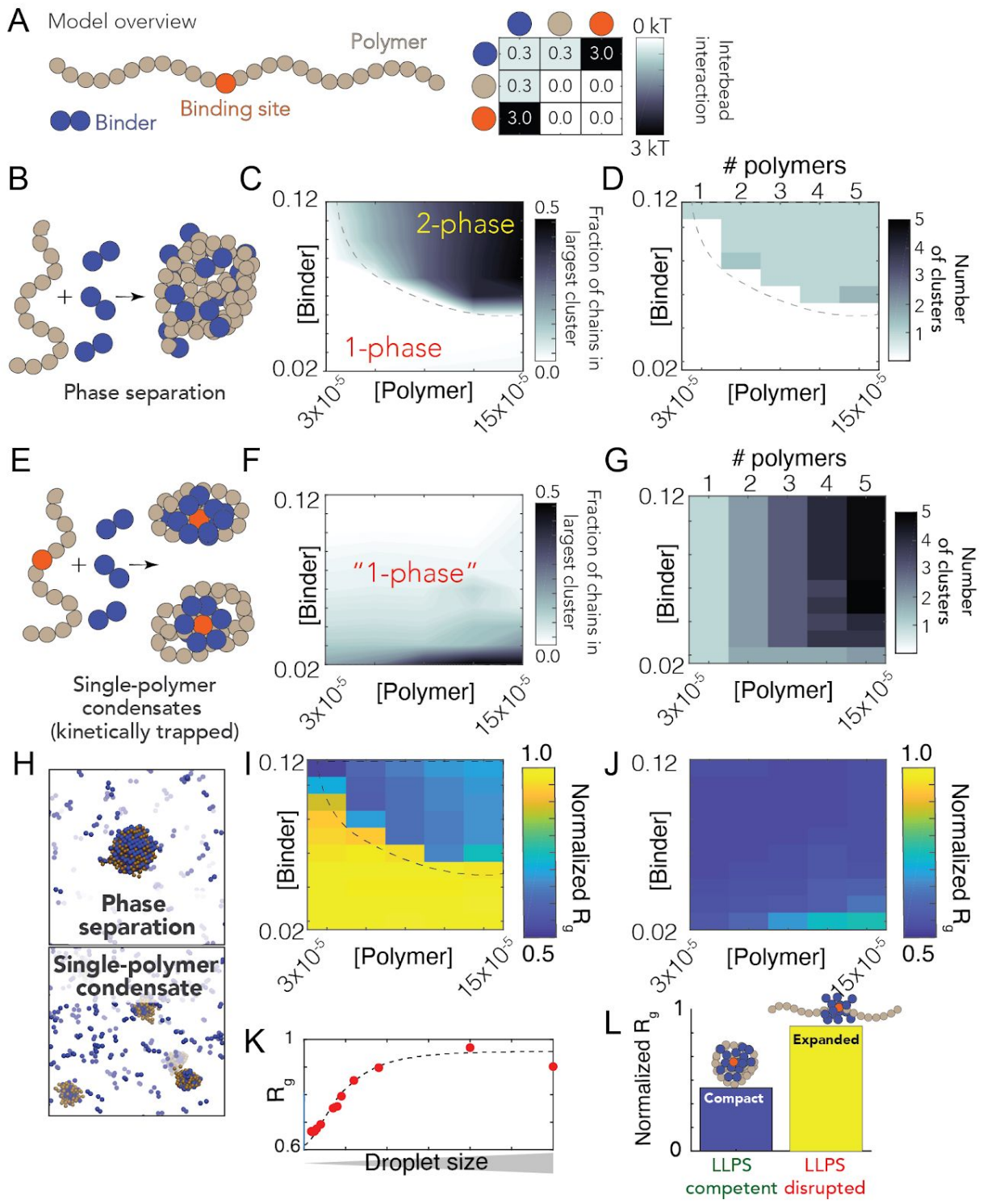


**Figure 6: A simple polymer suggests symmetry breaking can promote single-polymer condensates over multi-polymer assemblies**

**A** Summary of our model setup, which involves long 'polymers' (61 beads) or short 'binders' (2 beads). Each bead is multivalent and can interact with every adjacent lattice site. The interaction matrix to the right defines the pairwise interaction energies associated with each of the bead times.

**B** Concentration dependent assembly behaviour for polymers lacking a high-affinity binding site.
**C** Phase diagram showing the concentration-dependent phase regime - dashed line represents the binodal (phase boundary) and is provided to guide the eye.
**D** Analysis in the same 2D space as C assessing the number of droplets at a given concentration. When phase separation occurs a single droplet appears in almost all cases.
**E** Concentration dependent assembly behaviour for polymers with a high-affinity binding site.
**F** No large droplets are formed in any of the systems, although multiple polymer:binder complexes form.
**G** The number of clusters observed matches the number of polymers in the system - i.e. each polymer forms an individual cluster.
**H** Simulation snapshots from equivalent simulations for polymers with (top) or without (bottom) a single high-affinity binding site
**I** Polymer dimensions in the dense and dilute phase (for the parameters in our model) for polymers with no high-affinity binding site. Note that compaction in the dense phase reflects finite-size dependent effects, as addressed in panel K.
**J** Polymer dimensions across the same concentration space for polymers with a single high-affinity binding site. Across all concentrations, each individual polymer is highly compact.
**K** Compaction in the dense phase (panel I) is due to small droplets. When droplets are sufficiently large we observe chain expansion as expected.
**L** Simulations performed under conditions in which non-specific interactions between binder and polymer are reduced (interaction strength = 0 kT). Under these conditions phase separation is suppressed. Equivalent simulations for polymers with a high-affinity site reveal these chains are no longer compact. As such, phase separation offers a readout that - in our model - maps to single-polymer compaction.

The high affinity binding site polarizes the single-polymer condensate, such that they are organized, recalcitrant to fusion, and kinetically stable. A convenient physical analogy is that of a micelle, which are non-stoichiometric stable assemblies. Even for micelles that are far from their optimal size, fusion is slow because it requires substantial molecular reorganization and the breaking of highly stable interactions (Denkova et al., 2010; Pool and Bolhuis, 2007). If our simulations are run in a way deliberately designed to rapidly reach equilibrium using enhanced sampling approaches eventually all single-polymer condensates coalesce into one large multi-polymer condensate. The difference between the simulations with and without the high affinity binder is that in the limit of low-to-intermediate binder concentration, fusion of single-polymer condensates is extremely slow. This supports a general argument that a slow down in the fusion kinetics may play a role in favoring the packaging of a single gRNA.

Finally, we ran simulations under conditions in which binder:polymer interactions were reduced, mimicking the scenario in which protein:RNA interactions are inhibited (**Fig. 6L**). Under these conditions our polymers with a high-affinity binding site do not undergo chain collapse (in contrast to when binder:polymer interactions are present, see **Fig. 6J**), illustrating that phase separation offers a convenient readout for molecular interactions that might otherwise be challenging to measure.

We emphasize that our conclusions from simulations are subject to the parameters in our model. We present these results to demonstrate an example of "*how this single-genome*

*packaging could be achieved*", as opposed to the much stronger statement of proposing "*this is how it is"* achieved. Of note, recent elegant work by Ranganathan and Shakhnovich identified kinetically arrested microclusters, where slow kinetics result from the saturation of stickers within those clusters (Ranganathan and Shakhnovich, 2020). This is completely analogous to our results (albeit with homotypic interactions, rather than heterotypic interactions), giving us confidence that the physical principles uncovered are robust and, we tentatively suggest, quite general. Future simulations are required to systematically explore the details of the relevant parameter space in our system. However, regardless of those parameters, our model does establish that if weak multivalent interactions underlie the formation of large multi-polymer droplets, those same interactions cannot *also* drive polymer compaction inside the droplet.

An important technical point is that it would appear that we observe compaction inside the droplets (**Fig. 6I**). This is an artefact of the relatively small droplets formed in our systems (relative to the size of the polymer). The droplets act as a bounding cage for the polymer, driving their compaction indirectly. Simulations performed on a larger scale reveal that inside the droplet, polymers are highly expanded, simultaneously maximizing their interactions with binders while minimizing the loss of configurational entropy (**Fig. 6K**).

## Discussion

The nucleocapsid (N) protein from SARS-CoV-2 is a multivalent RNA binding protein critical for viral replication and genome packaging (Laude and Masters, 1995; Masters, 2019). To better understand how the various folded and disordered domains interact with one another, we applied single-molecule spectroscopy and all-atom simulations to perform a detailed biophysical dissection of the protein, uncovering several putative interaction motifs. Furthermore, based on both sequence analysis and our single-molecule experiments, we anticipated that N protein would undergo phase separation with RNA. In agreement with this prediction, and in line with work from the Gladfelter and Yildiz groups working independently from us, we find that N protein robustly undergoes phase separation in vitro with model RNA under a range of different salt conditions. Using simple polymer models, we propose that the same interactions that drive phase separation may also drive genome packaging into a dynamic, single-genome condensate. The formation of single-genome condensates (as opposed to multi-genome droplets) is influenced by the presence of one (or more) symmetry-breaking interaction sites, which we tentatively suggest could reflect packaging signals in viral genomes.

### All three IDRs are highly dynamic

Our single-molecule experiments and all-atom simulations are in good agreement with one another, and reveal that all three IDRs are extended and highly dynamic. Simulations suggest the NTD may interact transiently with the RBD, which offers an explanation for the slightly slowed reconfiguration time measured by nanosecond FCS. The LINK shows rapid rearrangement, demonstrating the RBD and dimerization domain are not interacting. Finally, we

see more pronounced interaction between the CTD and the dimerization domain, although these interactions are still highly transient.

Single molecule experiments and all-atom simulations were performed on monomeric versions of the protein, yet N protein has previously been shown to undergo dimerization and form higher-order oligomers in the absence of RNA (Chang et al., 2013). To assess the formation of oligomeric species, we use a combination of nativePAGE, crosslinking and FCS experiments (see **Fig. S13** and SI). These experiments also verified that under the conditions used for single molecule experiments the protein exists only as a monomer.

**Simulations identify multiple transient helices**

We identified a number of transient helical motifs which provide structural insight into previously characterized molecular interactions. Transient helices are ubiquitous in viral disordered regions and have been shown to underlie molecular interactions in a range of systems (Feuerstein et al., 2012; Guseva et al., 2020; Jensen et al., 2008; Leyrat et al., 2011).

Transient helix H2 (in the NTD) and H3 (in the LINK) flank the RBD and organize a set of arginine residues to face the same direction (**Fig. 2E**). Both the NTD and LINK have been shown to drive RNA binding, such that we propose these helical arginine-rich motifs (ARMs) may engage in both both nonspecific binding and may also contribute to RNA specificity, as has been proposed previously (Battiste et al., 1996; Bayer et al., 2005; Chang et al., 2009). The SR-rich region (which includes H3) has been previously identified as engaging in interaction with a structured acidic helix in Nsp3 in the model coronavirus MHV, consistent with an electrostatic helical interaction (Hurst et al., 2010, 2013). The RS-region is necessary for recruitment to replication-transcription centers (RTCs) in MHV, and also undergoes phosphorylation, setting the stage for a complex regulatory system awaiting exploration (Surjit et al., 2005; Verheije et al., 2010).

Transient helix H4 (**Fig. 3H**), was previously predicted bioinformatically and identified as a conserved feature across different coronaviruses (Chang et al., 2009). Furthermore, the equivalent region was identified in SARS coronavirus as a nuclear export signal (NES), such that we suspect this too is a classical Crm1-binding leucine-rich NES (Timani et al., 2005).

Transient helix H6 is an amphipathic helix with a highly hydrophobic face (**Fig. 4H**). Residues in this region have been identified as mediating M-protein binding in other coronaviruses, such that we propose H6 underlies that interaction (Hurst et al., 2005; Kuo and Masters, 2002; Verma et al., 2006). Recent work has also identified amphipathic transient helices in disordered proteins as interacting directly with membranes, such that an additional (albeit entirely speculative) role could involve direct membrane interaction, as has been observed in other viral phosphoproteins (Brass et al., 2002; Braun et al., 2017).

**The physiological relevance of nucleocapsid protein phase separation in SARS-CoV-2 physiology**

Our work has revealed that SARS-CoV-2 N protein undergoes phase separation with RNA when reconstituted *in vitro*. The solution environment and types of RNA used in our experiments are very different from the cytoplasm and viral RNA. However, similar results have been obtained in published and unpublished work by several other groups under a variety of conditions, including via *in cell* experiments (Yildiz group, unpublished) (Iserman et al., 2020; Perdikari et al., 2020). Taken together, these results demonstrate that N protein *can* undergo *bona fide* phase separation, and that N protein condensates *can* form in cells. Nevertheless, the complexity introduced by multidimensional linkage effects *in vivo* could substantially influence the phase behaviour and composition of condensates observed in the cell (Choi et al., 2019; Riback et al., 2020; Wyman and Gill, 1990). Of note, the regime we have identified in which phase separation occurs (**Fig. 5**) is remarkably relatively narrow, a prerequisite for the assembly of virion particles containing a single viral genome .

Does phase separation play a physiological role in SARS-CoV-2 biology? Under a phase-separation and assembly model as proposed for other viruses, the process of phase separation is decoupled from genome packaging, where packaging emerges from nucleocapsid assembly (Guseva et al., 2020). If applied to SARS-CoV-2, such a model would suggest that (1) Initially N protein and RNA phase separate in the cytosol, (2) that some discrete pre-capsid state forms within condensates and, (3) upon maturation, the pre-capsid is released from the condensate and undergoes subsequent virion assembly by interacting with the membrane-bound M, E, and S structural proteins. While this model is attractive it places a number of constraints on the physical properties of this pre-capsid, not least that the ability to escape the “parent” condensate dictates that the assembled pre-capsid must interact less strongly with the condensate components than in the unassembled state. This requirement introduces some thermodynamic complexities: how is a pre-capsid state driven to assemble if it is necessarily less stable than the unassembled pre-capsid, and how is incomplete pre-capsid formation avoided if – as assembly occurs – the pre-capsid becomes progressively less stable?
This model raises additional questions, such as the origins of specificity for recruitment of viral proteins and viral RNA, the kinetics of assembly in a droplet, and preferential packaging of gRNA over sub-genomic RNA. None of these questions are unanswerable, nor do they invalidate this model, but they should be addressed if the physiological relevance of large cytoplasmic condensates is to be further explored in the context of viral assembly.

Our preferred interpretation is that N protein has evolved to drive genome compaction for packaging (**Fig. 7**). In this model, a single-genome condensate forms through N protein gRNA interaction, driven by a small number of high-affinity sites. This (meta)-stable single-genome condensate undergoes subsequent maturation, leading to virion assembly. In this model, condensate-associated N proteins are in exchange with a bulk pool of soluble N protein, such that the interactions that drive compaction are heterogeneous and dynamic. The resulting condensate is then in effect a multivalent binder for M protein, which interacts with N directly (He et al., 2004b).

An open question pertains to specificity of packaging gRNA while excluding other RNAs. While we have not explored this question directly, based on our model for genome packaging the easiest conceptual route to ensure exclusive gRNA compaction would be the presence of two-high affinity sites at the extreme 5' and 3' ends of the gRNA. By ensuring both sites are required for compaction, only full-length gRNA will be packaged as only full-length gRNA has both sites. A recent map of N protein binding to gRNA maps to this architecture and suggests high-affinity binding regions at the extreme 5' and 3' ends of the gRNA, in good agreement with our qualitative prediction (Iserman et al., 2020).

Genome compaction through dynamic multivalent interactions would be especially relevant for coronaviruses, which have extremely large single-stranded RNA genomes. This is evolutionarily appealing, in that as the genome grows larger, compaction becomes increasingly efficient, as the effective valence of the genome is increased (Choi et al., 2020; Rubinstein and Colby, 2003). The ability of multivalent disordered proteins to drive RNA compaction has been observed previously in various contexts (Holmstrom et al., 2018, 2019). Furthermore, genome compaction by RNA binding protein has been proposed and observed in other viruses (Borodavka et al., 2013; Linger et al., 2004; Rodríguez et al., 2004), and the SARS coronavirus N protein has previously been shown to act as an RNA chaperone, an expected consequence of compaction to a dynamic single-RNA condensate that accommodates multiple N proteins with a single RNA (Holmstrom et al., 2019; Zúñiga et al., 2007). Finally, N protein has been shown to interact directly with a number of proteins we and others have studied in the context of biological phase separation (notably hnRNPA1 and G3BP1) which may influence assembly *in vivo* (Gordon et al., 2020; Luo et al., 2005; Martin et al., 2020; Perdikari et al., 2020; Sanders et al., 2020a; Yang et al., 2020).

Our model is also in good empirical agreement with recent observations made for other viruses (van Rosmalen et al., 2020). Taken together, we speculate that viral packaging may -in general- involve an initial genome compaction through multivalent protein:RNA and protein:protein interactions, followed by a liquid-to-solid transition in cases where well-defined crystalline capsid structures emerge. Liquid-to-solid transitions are well established in the context of neurodegeneration with respect to disease progression (Alberti and Dormann, 2019; Dao et al., 2018; Patel et al., 2015; Weber and Brangwynne, 2012). Here we suggest nature is leveraging those same principles as an evolved mechanism for monodisperse particle assembly.

Regardless of if phase separated condensates form inside cells, all available evidence suggests phase separation is reporting on a physiologically important interaction that underlies genome compaction (**Fig. 6L**). With this in mind, from a biotechnology standpoint, phase separation may be a convenient readout for *in vitro* assays to interrogate protein:RNA interaction. Regardless of which model is correct, N protein:RNA interaction is key for viral replication. As such, phase separation provides a macroscopic reporter on a nanoscopic phenomenon, in line with previous work (Dignon et al., 2018; Martin et al., 2020; Rubinstein and

Colby, 2003; Zeng et al., 2020b). In this sense, we believe the therapeutic implications of understanding and modulating phase separation here (and elsewhere in biology) are conveniently decoupled from the physiological relevance of actual, large phase separated “liquid droplets”, but instead offer a window into the underlying physical interactions that lead to condensate formation.

**The physics of single polymer condensates**

Depending on the molecular details, single-polymer condensates may be kinetically stable (but thermodynamically unstable, as in our model simulations) or thermodynamically stable. Delineation between these two scenarios will depend on the nature, strength, valency and anisotropy of the interactions. It is worth noting that from the perspective of functional biology, kinetic stability may be essentially indistinguishable from thermodynamic stability, depending on the lifetime of a metastable species.

It is also important to emphasize that at higher concentrations of N protein and/or after a sufficiently long time period we expect robust phase separation with viral RNA, regardless of the presence of a symmetry-breaking site. Symmetry breaking is achieved when the apparent local concentration of N protein (from the “perspective” of gRNA) is substantially higher than the actual global concentration. As effective local and global concentrations approach one another, the entropic cost of intra-molecular interaction is outweighed by the availability of inter-molecular partners. On a practical note, if the readout in question is the presence/absence of liquid droplets, a high affinity site may be observed as a shift in the saturation concentration which, confusingly, could either suppress or enhance phase separation. Further, if single-genome condensates are kinetically stable and driven through electrostatic interactions, we would expect a complex temperature dependence, in which larger droplets are observed at higher temperature (up to some threshold). Recent work is showing a strong temperature-dependence of phase separation is consistent with these predictions (Iserman et al., 2020).

Finally, we note no reason to assume single-RNA condensates should be exclusively the purview of viruses. RNAs in eukaryotic cells may also be processed in these types of assemblies, as opposed to in large multi-RNA RNPs. The role of RNA:RNA interactions both here and in other systems is also of particular interest and not an aspect explored in our current work, but we anticipate may play a key role in the relevant biology (Langdon et al., 2018; Ma et al., 2020; Zhang et al., 2015).

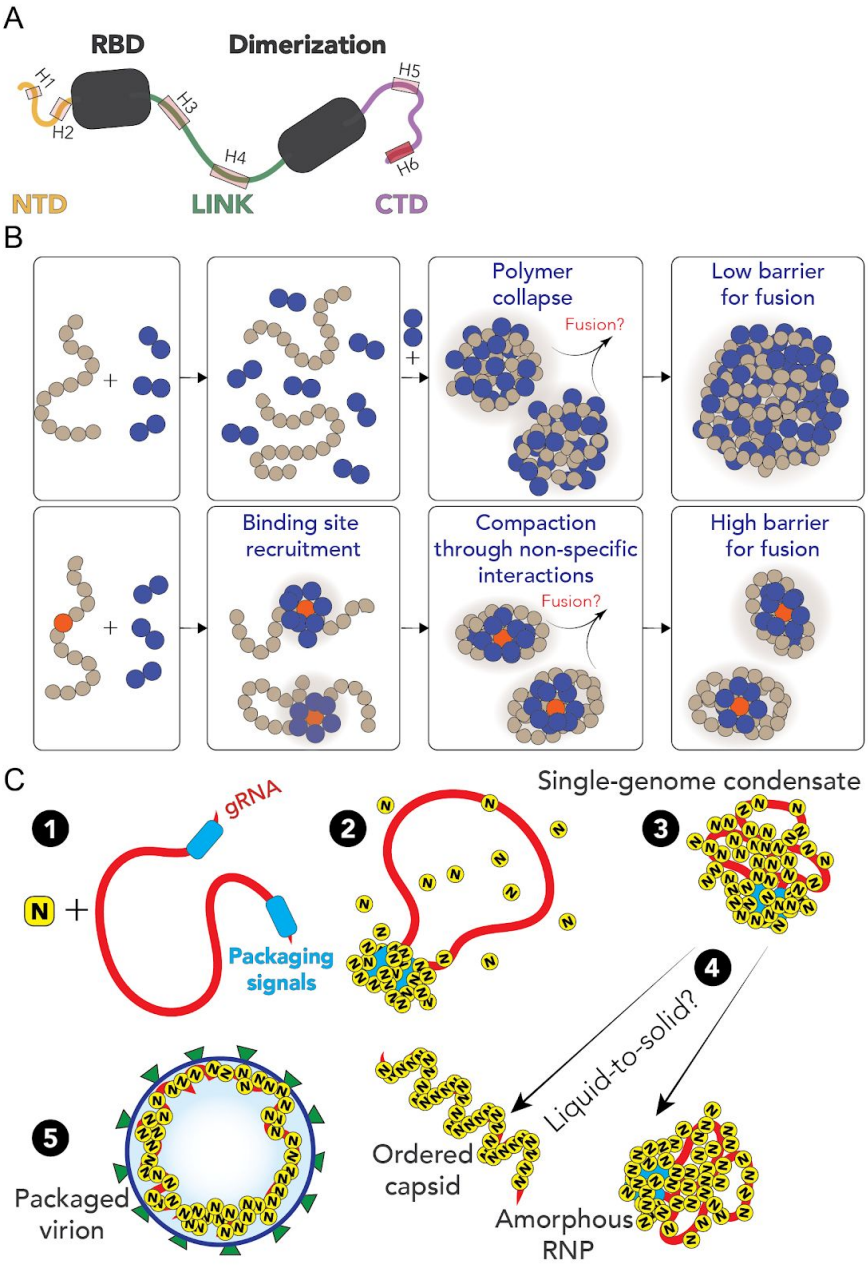


**Figure 7: Summary and proposed model**

**A.** Summary of results from single-molecule spectroscopy experiments and all-atom simulations. All three predicted IDRs are disordered, highly flexible, and house a number of putative helical binding regions which overlap with subregions identified previously to drive N protein function.

**B.** Overview of general symmetry breaking model. For homopolymeric polymers, local collapse leads to single-polymer condensates that undergo barrier-less fusion, rapidly assembling into large multi-polymer condensates. When one (or a small number of) high affinity sites are present, local clustering of binders at a lower concentration organize the polymer such that single-polymer condensates are kinetically stable.

**C.** Proposed model for SARS-CoV-2 genome packaging. **(1)** Simplified model of SARS-CoV-2 genome with a pair of packaging region at the 5' and 3' end of the genome **(2)** N protein preferentially binds to packaging signal regions in the genome, leadling to a local cluster of N protein at the packaging signal

RNA. **(3)** The high local concentration of N protein drives condensation of distal regions of the genome, forming a stable single-genome condensate. **(4)** Single-genome condensates may undergo subsequent maturation through a liquid-to-solid (crystallization) transition to form an ordered crystalline capsid, or solidify into an amorphous ribonuclear particle (RNP). While in some viruses an ordered capsid clearly forms, we favour a model in which the SARS-CoV-2 capsid is an amorphous RNP. Compact single-genome condensate ultimately recruits E, S and M proteins at the membrane, driving envelope formation and final virion packaging.

## Acknowledgements

We thank Amy Gladfelter, Christiane Iserman, Christine Roden, Ahmet Yildiz, Amanda Jack, Luke Ferro, Steve Michnick, Pascale Legault, and Jim Omichinski for sharing data and extensive discussion. We also thank Rohit Pappu for placing our groups in contact with one another.

We thank the labs of John Cooper, Carl Frieden, and Silvia Jansen for providing some of the reagents we have used in this work. We thank Ben Schuler and Daniel Nettels for developing, maintaining, and sharing with us the software package used to analyze the single-molecule data.

J.C. and J.J.A are supported by NIGMS R25 IMSD Training Grant GM103757. We are grateful to the citizen-scientists of Folding@home for donating their computing resources. G.R.B holds an NSF CAREER Award MCB-1552471, NIH R01GM12400701, a Career Award at the Scientific Interface from the Burroughs Wellcome Fund, and a Packard Fellowship for Science and Engineering from The David and Lucile Packard Foundation.

A.S.H. is a scientific consultant with Dewpoint Therapeutics.

## Methods

### All-atom simulations

All-atom Monte Carlo simulations were performed with the ABSINTH implicit solvent model and CAMPARI simulation engine (http://campari.sourceforge.net/) (Mittal et al., 2015; Vitalis and Pappu, 2009) with the solution ion parameters of Mao et al.(Mao and Pappu, 2012). Simulations were performed using movesets and Hamiltonina parameters as reported previously (Martin et al., 2020; Sherry et al., 2017). All simulations were performed in sufficiently large box sizes to prevent finite size effects (where box size varies from system to system).

All-atom molecular dynamics simulations were performed using GROMACS, using the FAST algorithm in conjunction with the Folding@home platform (Abraham et al., 2015/9; Shirts and Pande, 2000; Zimmerman and Bowman, 2015). Post-simulation analysis was performed with Enspara (Porter et al., 2019). For additional simulation details see the supplementary information.

**Coarse-grained Polymer Simulations**

Coarse-grained Monte Carlo simulations were performed using the PIMMS simulation engine (Holehouse and Pappu, 2019). All simulations were performed in a 70 x 70 x 70 lattice-site box. The results averaged over the final 20% of the simulation to give average values at equivalent states. The "polymer" is represented as a 61-residue polymer with either a central high-affinity binding site or not. The binder is a 2-bead species. Every simulation was run for 20 x $10^9$ Monte Carlo steps, with four independent replicas. Bead interaction strengths were defined as shown in **Fig. 6A**. For additional simulation details see the supplementary information.

**Protein Expression, purification, and labeling**.

SARS-CoV-2 Nucleocapsid protein (NCBI Reference Sequence: YP_009724397.2) including an N term extension containing $His_9$-HRV 3C protease site was cloned into the BamHI EcoRI sites in the MCS of pGEX-6P-1 vector (GE Healthcare). Site-directed mutagenesis was performed on the $His_9$-SARS-CoV-2 Nucleocapsid pGEX vector to create M1C R68C, Y172C T245C, and F363C A419C variant N protein constructs and sequences were verified using Sanger sequencing. All variants were expressed recombinantly in BL21 Codon-plus pRIL cells (Agilent) or Gold BL21(DE3) cells (Agilent) and purified using a FF HisTrap column. The GST-$His_9$-N tag was then cleaved using HRV 3C protease and further purified to remove the cleaved tag. Finally, purified N protein variants were analyzed using SDS-PAGE and verified by electrospray ionization mass spectrometry (LC-MS). Activity of the protein was assessed by testing whether the protein is able to bind and condense nucleic acids (see phase-separation experiments) as well as to form dimers (see oligomerization in SI).

All Nucleocapsid variants were labeled with Alexa Fluor 488 maleimide and Alexa Fluor 594 maleimide (Molecular Probes) under denaturing conditions following a two-step sequential labeling procedure (see SI).

**Single-molecule fluorescence spectroscopy.**

Single-molecule fluorescence measurements were performed with a Picoquant MT200 instrument (Picoquant, Germany). FRET experiments were performed by exciting the donor dye with a laser power of 100 µW (measured at the back aperture of the objective). For pulsed interleaved excitation of donor and acceptor, the power used for exciting the acceptor dye was adjusted to match the acceptor emission intensity to that of the donor (between 50 and 70 mW). Single-molecule FRET efficiency histograms were acquired from samples with protein concentrations between 50 pM and 100 pM and the population with stoichiometry corresponding to 1:1 donor:acceptor labeling was selected. Trigger times for excitation pulses (repetition rate 20 MHz) and photon detection events were stored with 16 ps resolution. For FRET-FCS, samples of double-labeled protein with a concentration of 100 pM were excited by either the diode laser or the supercontinuum laser at the powers indicated above.

All samples were prepared in 50 mM Tris pH 7.32, 143 mM β-mercaptoethanol (for photoprotection), 0.001% Tween 20 (for limiting surface adhesion) and GdmCl at the reported concentrations. All measurements were performed in uncoated polymer coverslip cuvettes (Ibidi, Wisconsin, USA), which significantly decrease the fraction of protein adhering to the surface (compared to normal glass cuvettes) under native conditions. For comparison, experiments have been performed also in glass cuvette coated with PEG, which provided analogous results to the polymeric cuvette. Each sample was measured for at least 30 min at room temperature (295 ± 0.5 K) (see SI).

## References


Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015/9). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25.

Alberti, S., and Dormann, D. (2019). Liquid-Liquid Phase Separation in Disease. Annu. Rev. Genet. *53*, 171–194.

Aznauryan, M., Delgado, L., Soranno, A., Nettels, D., Huang, J.-R., Labhardt, A.M., Grzesiek, S., and Schuler, B. (2016). Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. Proc. Natl. Acad. Sci. U. S. A. *113*, E5389–E5398.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Banerjee, P.R., Milin, A.N., Moosa, M.M., Onuchic, P.L., and Deniz, A.A. (2017). Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. Angew. Chem. Int. Ed Engl. *56*, 11354–11359.

Battiste, J.L., Mao, H., Rao, N.S., Tan, R., Muhandiram, D.R., Kay, L.E., Frankel, A.D., and Williamson, J.R. (1996). Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. Science *273*, 1547–1551.

Bayer, T.S., Booth, L.N., Knudsen, S.M., and Ellington, A.D. (2005). Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA *11*, 1848–1857.

Boeynaems, S., Holehouse, A.S., Weinhardt, V., Kovacs, D., Van Lindt, J., Larabell, C., Van Den Bosch, L., Das, R., Tompa, P.S., Pappu, R.V., et al. (2019). Spontaneous driving forces give rise to protein-RNA condensates with coexisting phases and complex material properties. Proc. Natl. Acad. Sci. U. S. A. *116*, 7889–7898.

Borgia, A., Wensley, B.G., Soranno, A., Nettels, D., Borgia, M.B., Hoffmann, A., Pfeil, S.H., Lipman, E.A., Clarke, J., and Schuler, B. (2012). Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. Nat. Commun. *3*, 1195.

Borgia, A., Zheng, W., Buholzer, K., Borgia, M.B., Schüler, A., Hofmann, H., Soranno, A., Nettels, D., Gast, K., Grishaev, A., et al. (2016). Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. J. Am. Chem. Soc. *138*, 11714–11726.

Borgia, A., Borgia, M.B., Bugge, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D., et al. (2018). Extreme disorder in an ultrahigh-affinity protein complex. Nature *555*, 61.

Borodavka, A., Tuma, R., and Stockley, P.G. (2013). A two-stage mechanism of viral RNA

compaction revealed by single molecule fluorescence. RNA Biol. *10*, 481–489.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Juelicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729–1732.

Brass, V., Bieck, E., Montserret, R., Wölk, B., Hellings, J.A., Blum, H.E., Penin, F., and Moradpour, D. (2002). An Amino-terminal Amphipathic α-Helix Mediates Membrane Association of the Hepatitis C Virus Nonstructural Protein 5A. J. Biol. Chem. *277*, 8130–8139.

Braun, A.R., Lacy, M.M., Ducas, V.C., Rhoades, E., and Sachs, J.N. (2017). α-Synuclein's Uniquely Long Amphipathic Helix Enhances its Membrane Binding and Remodeling Capacity. J. Membr. Biol. *250*, 183–193.

Chang, C.-K., Hsu, Y.-L., Chang, Y.-H., Chao, F.-A., Wu, M.-C., Huang, Y.-S., Hu, C.-K., and Huang, T.-H. (2009). Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging. J. Virol. *83*, 2255–2264.

Chang, C.-K., Chen, C.-M.M., Chiang, M.-H., Hsu, Y.-L., and Huang, T.-H. (2013). Transient oligomerization of the SARS-CoV N protein--implication for virus ribonucleoprotein packaging. PLoS One *8*, e65045.

Choi, J.-M., Dar, F., and Pappu, R.V. (2019). LASSI: A lattice model for simulating phase transitions of multivalent proteins. PLoS Comput. Biol. *15*, e1007028.

Choi, J.-M., Holehouse, A.S., and Pappu, R.V. (2020). Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. Annu. Rev. Biophys. *49*, 107–133.

Chung, H.S., Piana-Agostinetti, S., Shaw, D.E., and Eaton, W.A. (2015). Structural origin of slow diffusion in protein folding. Science *349*, 1504–1510.

Cologna, R., and Hogue, B.G. (2000). Identification of a bovine coronavirus packaging signal. J. Virol. *74*, 580–583.

Corman, V.M., Muth, D., Niemeyer, D., and Drosten, C. (2018). Chapter Eight - Hosts and Sources of Endemic Human Coronaviruses. In Advances in Virus Research, M. Kielian, T.C. Mettenleiter, and M.J. Roossinck, eds. (Academic Press), pp. 163–188.

Cui, L., Wang, H., Ji, Y., Yang, J., Xu, S., Huang, X., Wang, Z., Qin, L., Tien, P., Zhou, X., et al. (2015). The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells. J. Virol. *89*, 9029–9043.

Dao, T.P., Kolaitis, R.-M., Kim, H.J., O'Donovan, K., Martyniak, B., Colicino, E., Hehnly, H., Taylor, J.P., and Castañeda, C.A. (2018). Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. Mol. Cell *69*, 965–978.e6.

Denkova, A.G., Mendes, E., and Coppens, M.-O. (2010). Non-equilibrium dynamics of block copolymer micelles in solution: recent insights and open questions. Soft Matter *6*, 2351–2357.

Dignon, G.L., Zheng, W., Best, R.B., Kim, Y.C., and Mittal, J. (2018). Relation between

single-molecule properties and phase behavior of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U. S. A. *115*, 9929–9934.

Dimura, M., Peulen, T.O., Hanke, C.A., Prakash, A., Gohlke, H., and Seidel, C.A. (2016). Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. Curr. Opin. Struct. Biol. *40*, 163–185.

Doose, S., Neuweiler, H., and Sauer, M. (2005). A close look at fluorescence quenching of organic dyes by tryptophan. Chemphyschem *6*, 2277–2285.

Feuerstein, S., Solyom, Z., Aladag, A., Favier, A., Schwarten, M., Hoffmann, S., Willbold, D., and Brutscher, B. (2012). Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. J. Mol. Biol. *420*, 310–323.

Fosmire, J.A., Hwang, K., and Makino, S. (1992). Identification and characterization of a coronavirus packaging signal. J. Virol. *66*, 3522–3530.

Fuertes, G., Banterle, N., Ruff, K.M., Chowdhury, A., Mercadante, D., Koehler, C., Kachala, M., Estrada Girona, G., Milles, S., Mishra, A., et al. (2017). Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. Proc. Natl. Acad. Sci. U. S. A. *114*, E6342–E6351.

Gordon, D.E., Jang, G.M., Bouhaddou, M., Xu, J., Obernier, K., White, K.M., O'Meara, M.J., Rezelj, V.V., Guo, J.Z., Swaney, D.L., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature.

Grossoehme, N.E., Li, L., Keane, S.C., Liu, P., Dann, C.E., 3rd, Leibowitz, J.L., and Giedroc, D.P. (2009). Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. J. Mol. Biol. *394*, 544–557.

Guillén-Boixet, J., Kopach, A., Holehouse, A.S., Wittmann, S., Jahnel, M., Schlüßler, R., Kim, K., Irmela R E, Wang, J., Mateju, D., et al. (2020). RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. Cell *181*, 346–361.e17.

Guseva, S., Milles, S., Jensen, M.R., Salvi, N., Kleman, J.-P., Maurin, D., Ruigrok, R.W.H., and Blackledge, M. (2020). Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. Sci Adv *6*, eaaz7095.

Haenni, D., Zosel, F., Reymond, L., Nettels, D., and Schuler, B. (2013). Intramolecular distances and dynamics from the combined photon statistics of single-molecule FRET and photoinduced electron transfer. J. Phys. Chem. B *117*, 13015–13028.

He, R., Dobie, F., Ballantine, M., Leeson, A., Li, Y., Bastien, N., Cutts, T., Andonov, A., Cao, J., Booth, T.F., et al. (2004a). Analysis of multimerization of the SARS coronavirus nucleocapsid protein. Biochem. Biophys. Res. Commun. *316*, 476–483.

He, R., Leeson, A., Ballantine, M., Andonov, A., Baker, L., Dobie, F., Li, Y., Bastien, N., Feldmann, H., Strocher, U., et al. (2004b). Characterization of protein–protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res.

*105*, 121–125.

Heinrich, B.S., Maliga, Z., Stein, D.A., Hyman, A.A., and Whelan, S.P.J. (2018). Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments. MBio *9*.

Higgs, P.G., and Joanny, J. (1991). Theory of polyampholyte solutions. J. Chem. Phys. *94*, 1543–1554.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell *181*, 271–280.e8.

Hofmann, H., Soranno, A., Borgia, A., Gast, K., Nettels, D., and Schuler, B. (2012). Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc. Natl. Acad. Sci. U. S. A. *109*, 16155–16160.

Holehouse, A.S., and Pappu, R.V. (2019). PIMMS (0.24 pre-beta).

Holehouse, A.S., Garai, K., Lyle, N., Vitalis, A., and Pappu, R.V. (2015). Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. J. Am. Chem. Soc. *137*, 2984–2995.

Holmstrom, E.D., Nettels, D., and Schuler, B. (2018). Conformational Plasticity of Hepatitis C Virus Core Protein Enables RNA-Induced Formation of Nucleocapsid-like Particles. J. Mol. Biol. *430*, 2453–2467.

Holmstrom, E.D., Liu, Z., Nettels, D., Best, R.B., and Schuler, B. (2019). Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. Nat. Commun. *10*, 2453.

Hsieh, P.-K., Chang, S.C., Huang, C.-C., Lee, T.-T., Hsiao, C.-W., Kou, Y.-H., Chen, I.-Y., Chang, C.-K., Huang, T.-H., and Chang, M.-F. (2005). Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. J. Virol. *79*, 13848–13855.

Hurst, K.R., Kuo, L., Koetzner, C.A., Ye, R., Hsue, B., and Masters, P.S. (2005). A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. J. Virol. *79*, 13285–13297.

Hurst, K.R., Ye, R., Goebel, S.J., Jayaraman, P., and Masters, P.S. (2010). An Interaction between the Nucleocapsid Protein and a Component of the Replicase-Transcriptase Complex Is Crucial for the Infectivity of Coronavirus Genomic RNA. J. Virol. *84*, 10276–10288.

Hurst, K.R., Koetzner, C.A., and Masters, P.S. (2013). Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J. Virol. *87*, 9159–9172.

Iserman, C., Roden, C.A., Boerneke, M., Sealfon, R., McLaughlin, G., Jungreis, I., Park, C.Y.,

Boppana, A., Fritch, E., Hou, Y., et al. (2020). Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate.

Jayaram, H., Fan, H., Bowman, B.R., Ooi, A., Jayaram, J., Collisson, E.W., Lescar, J., and Prasad, B.V.V. (2006). X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. J. Virol. *80*, 6612–6620.

Jensen, M.R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R.W.H., and Blackledge, M. (2008). Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. J. Am. Chem. Soc. *130*, 8055–8061.

Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., He, S., Zhou, Z., Zhou, Z., Chen, Q., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharm Sin B.

Kuo, L., and Masters, P.S. (2002). Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. J. Virol. *76*, 4987–4999.

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., et al. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature *581*, 215–220.

Langdon, E.M., Qiu, Y., Ghanbari Niaki, A., McLaughlin, G.A., Weidmann, C.A., Gerbich, T.M., Smith, J.A., Crutchley, J.M., Termini, C.M., Weeks, K.M., et al. (2018). mRNA structure determines specificity of a polyQ-driven phase separation. Science *360*, 922–927.

Laude, H., and Masters, P.S. (1995). The Coronavirus Nucleocapsid Protein. In The Coronaviridae, S.G. Siddell, ed. (Boston, MA: Springer US), pp. 141–163.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of intrinsically disordered regions and proteins. Chem. Rev. *114*, 6589–6631.

Leyrat, C., Jensen, M.R., Ribeiro, E.A., Jr, Gérard, F.C.A., Ruigrok, R.W.H., Blackledge, M., and Jamin, M. (2011). The N0-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α-helices. Protein Sci. *20*, 542–556.

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Linger, B.R., Kunovska, L., Kuhn, R.J., and Golden, B.L. (2004). Sindbis virus nucleocapsid assembly: RNA folding promotes capsid protein dimerization. RNA *10*, 128–138.

Luo, H., Chen, Q., Chen, J., Chen, K., Shen, X., and Jiang, H. (2005). The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1. FEBS Lett. *579*, 2623–2628.

Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006). Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. Biochemistry *45*, 11827–11835.

Lurie, N., Saville, M., Hatchett, R., and Halton, J. (2020). Developing Covid-19 Vaccines at Pandemic Speed. N. Engl. J. Med. *382*, 1969–1973.

Ma, W., Zhen, G., Xie, W., and Mayr, C. (2020). Unstructured mRNAs form multivalent RNA-RNA interactions to generate TIS granule networks.

Mao, A.H., and Pappu, R.V. (2012). Crystal lattice properties fully determine short-range interaction parameters for alkali and halide ions. J. Chem. Phys. *137*, 064104.

Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R.V., and Mittag, T. (2020). Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science *367*, 694–699.

Masters, P.S. (2019). Coronavirus genomic RNA packaging. Virology *537*, 198–207.

McBride, R., van Zyl, M., and Fielding, B.C. (2014). The coronavirus nucleocapsid is a multifunctional protein. Viruses *6*, 2991–3018.

Metrick, C.M., Koenigsberg, A.L., and Heldwein, E.E. (2020). Conserved Outer Tegument Component UL11 from Herpes Simplex Virus 1 Is an Intrinsically Disordered, RNA-Binding Protein. MBio *11*.

Mittal, A., Das, R.K., Vitalis, A., and Pappu, R.V. (2015). The ABSINTH implicit solvation model and forcefield paradigm for use in simulations of intrinsically disordered proteins. In Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods, M. Fuxreiter, ed. (CRC Press), pp. 188–222.

Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rüegger, S., Reymond, L., Nettels, D., and Schuler, B. (2010). Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U. S. A. *107*, 14609–14614.

Nettels, D., Gopich, I.V., Hoffmann, A., and Schuler, B. (2007). Ultrafast dynamics of protein collapse from single-molecule photon statistics. Proc. Natl. Acad. Sci. U. S. A. *104*, 2655–2660.

Nettels, D., Müller-Späth, S., Küster, F., Hofmann, H., Haenni, D., Rüegger, S., Reymond, L., Hoffmann, A., Kubelka, J., Heinz, B., et al. (2009). Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proc. Natl. Acad. Sci. U. S. A. *106*, 20740–20745.

Nikolic, J., Le Bars, R., Lama, Z., Scrima, N., Lagaudrière-Gesbert, C., Gaudin, Y., and Blondel, D. (2017). Negri bodies are viral factories with properties of liquid organelles. Nat. Commun. *8*, 58.

Pak, C.W., Kosno, M., Holehouse, A.S., Padrick, S.B., Mittal, A., Ali, R., Yunus, A.A., Liu, D.R., Pappu, R.V., and Rosen, M.K. (2016). Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. Mol. Cell *63*, 72–85.

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.

Perdikari, T.M., Murthy, A.C., Ryan, V.H., Watters, S., Naik, M.T., and Fawzi, N.L. (2020). SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins.

Pool, R., and Bolhuis, P.G. (2007). Sampling the kinetic pathways of a micelle fusion and fission transition. J. Chem. Phys. *126*, 244703.

Porter, J.R., Zimmerman, M.I., and Bowman, G.R. (2019). Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. J. Chem. Phys. *150*, 044108.

Posey, A.E., Holehouse, A.S., and Pappu, R.V. (2018). Chapter One - Phase Separation of Intrinsically Disordered Proteins. In Methods in Enzymology, E. Rhoades, ed. (Academic Press), pp. 1–30.

Post, C.B., and Zimm, B.H. (1979). Internal condensation of a single DNA molecule. Biopolymers *18*, 1487–1501.

Ranganathan, S., and Shakhnovich, E.I. (2020). Dynamic metastable long-living droplets formed by sticker-spacer proteins. Elife *9*.

Riback, J.A., Zhu, L., Ferrolino, M.C., Tolbert, M., Mitrea, D.M., Sanders, D.W., Wei, M.-T., Kriwacki, R.W., and Brangwynne, C.P. (2020). Composition-dependent thermodynamics of intracellular phase separation. Nature *581*, 209–214.

Robbins, S.G., Frana, M.F., McGowan, J.J., Boyle, J.F., and Holmes, K.V. (1986). RNA-binding proteins of coronavirus MHV: detection of monomeric and multimeric N protein with an RNA overlay-protein blot assay. Virology *150*, 402–410.

Rodríguez, L., Cuesta, I., Asenjo, A., and Villanueva, N. (2004). Human respiratory syncytial virus matrix protein is an RNA-binding protein: binding properties, location and identity of the RNA contact residues. J. Gen. Virol. *85*, 709–719.

Roser, M., Ritchie, H., Ortiz-Ospina, E., and Hasell, J. (2020). Coronavirus Pandemic (COVID-19). Our World in Data.

van Rosmalen, M.G.M., Kamsma, D., Biebricher, A.S., Li, C., Zlotnick, A., Roos, W.H., and Wuite, G.J.L. (2020). Revealing in real-time a multistep assembly mechanism for SV40 virus-like particles. Science Advances *6*, eaaz1639.

Rubinstein, M., and Colby, R.H. (2003). Polymer Physics (New York: Oxford University Press).

Sanders, D.W., Kedersha, N., Lee, D.S.W., Strom, A.R., Drake, V., Riback, J.A., Bracha, D., Eeftens, J.M., Iwanicki, A., Wang, A., et al. (2020a). Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. Cell *181*, 306–324.e28.

Sanders, J.M., Monogue, M.L., Jodlowski, T.Z., and Cutrell, J.B. (2020b). Pharmacologic

Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA.

Sauer, M., and Neuweiler, H. (2014). PET-FCS: probing rapid structural fluctuations of proteins and nucleic acids by single-molecule fluorescence quenching. Methods Mol. Biol. *1076*, 597–615.

Schellman, J.A. (1987). Selective binding and solvent denaturation. Biopolymers *26*, 549–559.

Schuler, B., Soranno, A., Hofmann, H., and Nettels, D. (2016). Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. Annu. Rev. Biophys. *45*, 207–231.

Semenov, A.N., and Rubinstein, M. (1998). Thermoreversible Gelation in Solutions of Associative Polymers. 1. Statics. Macromolecules *31*, 1373–1385.

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. Nature *581*, 221–224.

Sherry, K.P., Das, R.K., Pappu, R.V., and Barrick, D. (2017). Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. Proc. Natl. Acad. Sci. U. S. A. *114*, E9243–E9252.

Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. Science *357*.

Shirts, M., and Pande, V.S. (2000). Screen Savers of the World Unite! Science *290*, 1903–1904.

Soranno, A., Buchli, B., Nettels, D., Cheng, R.R., Müller-Späth, S., Pfeil, S.H., Hoffmann, A., Lipman, E.A., Makarov, D.E., and Schuler, B. (2012). Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. Proc. Natl. Acad. Sci. U. S. A. *109*, 17800–17806.

Soranno, A., Holla, A., Dingfelder, F., Nettels, D., Makarov, D.E., and Schuler, B. (2017). Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. Proc. Natl. Acad. Sci. U. S. A. *114*, E1833–E1839.

Soranno, A., Cabassi, F., and Orselli, M.E. (2018). Dynamics of Structural Elements of GB1 β-Hairpin Revealed by Tryptophan–Cysteine Contact Formation Experiments. The Journal of.

Stockley, P.G., Twarock, R., Bakker, S.E., Barker, A.M., Borodavka, A., Dykeman, E., Ford, R.J., Pearson, A.R., Phillips, S.E.V., Ranson, N.A., et al. (2013). Packaging signals in single-stranded RNA viruses: nature's alternative to a purely electrostatic assembly mechanism. J. Biol. Phys. *39*, 277–287.

Stockmayer, W.H. (1950). Light Scattering in Multi-Component Systems. J. Chem. Phys. *18*, 58–61.

Surjit, M., Kumar, R., Mishra, R.N., Reddy, M.K., Chow, V.T.K., and Lal, S.K. (2005). The severe acute respiratory syndrome coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation. J. Virol. *79*, 11476–11486.

Takeda, M., Chang, C.-K., Ikeya, T., Güntert, P., Chang, Y.-H., Hsu, Y.-L., Huang, T.-H., and Kainosho, M. (2008). Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. J. Mol. Biol. *380*, 608–622.

Timani, K.A., Liao, Q., Ye, L., Zeng, Y., Liu, J., Zheng, Y., Ye, L., Yang, X., Lingbao, K., Gao, J., et al. (2005). Nuclear/nucleolar localization properties of C-terminal nucleocapsid protein of SARS coronavirus. Virus Res. *114*, 23–34.

Tompa, P., and Fuxreiter, M. (2008/1). Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends Biochem. Sci. *33*, 2–8.

Verheije, M.H., Hagemeijer, M.C., Ulasli, M., Reggiori, F., Rottier, P.J.M., Masters, P.S., and de Haan, C.A.M. (2010). The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes. J. Virol. *84*, 11575–11579.

Verma, S., Bednar, V., Blount, A., and Hogue, B.G. (2006). Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. J. Virol. *80*, 4344–4355.

Vitalis, A., and Pappu, R.V. (2009). ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. J. Comput. Chem. *30*, 673–699.

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell *181*, 281–292.e6.

Wang, J., Choi, J.-M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., et al. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. Cell *174*, 688–699.e16.

Warner, J.B., 4th, Ruff, K.M., Tan, P.S., Lemke, E.A., Pappu, R.V., and Lashuel, H.A. (2017). Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths. J. Am. Chem. Soc. *139*, 14456–14469.

Weber, S.C., and Brangwynne, C.P. (2012). Getting RNA and protein in phase. Cell *149*, 1188–1191.

Woo, K., Joo, M., Narayanan, K., Kim, K.H., and Makino, S. (1997). Murine coronavirus packaging signal confers packaging to nonviral RNA. J. Virol. *71*, 824–827.

Wyman, J., and Gill, S.J. (1990). Binding and Linkage: Functional Chemistry of Biological Macromolecules (University Science Books).

Yang, P., Mathieu, C., Kolaitis, R.-M., Zhang, P., Messing, J., Yurtsever, U., Yang, Z., Wu, J., Li, Y., Pan, Q., et al. (2020). G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. Cell *181*, 325–345.e28.

Yu, I.-M., Gustafson, C.L.T., Diao, J., Burgner, J.W., 2nd, Li, Z., Zhang, J., and Chen, J. (2005). Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain. J. Biol. Chem. *280*, 23280–23286.

Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., Mohammed, A., Zhao, C., Yang, Y., Xie, J., et al. (2020a). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem. Biophys. Res. Commun. *527*, 618–623.

Zeng, X., Holehouse, A.S., Mittag, T., Chilkoti, A., and Pappu, R.V. (2020b). Connecting coil-to-globule transitions to full phase diagrams for intrinsically disordered proteins.

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.

Zheng, W., Borgia, A., Buholzer, K., Grishaev, A., Schuler, B., and Best, R.B. (2016). Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment. J. Am. Chem. Soc. *138*, 11702–11713.

Zheng, W., Zerze, G.H., Borgia, A., Mittal, J., Schuler, B., and Best, R.B. (2018). Inferring properties of disordered chains from FRET transfer efficiencies. J. Chem. Phys. *148*, 123329.

Zhou, Y., Su, J.M., Samuel, C.E., and Ma, D. (2019). Measles Virus Forms Inclusion Bodies with Properties of Liquid Organelles. J. Virol. *93*.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. *382*, 727–733.

Zimmerman, M.I., and Bowman, G.R. (2015). FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. J. Chem. Theory Comput. *11*, 5747–5757.

Zinzula, L., Nagy, M.O., and Bracher, A. (2020). 1.45 Angstrom Resolution Crystal Structure of C-terminal Dimerization Domain of Nucleocapsid Phosphoprotein from SARS-CoV-2 (PDB: 6YUN). Protein Data Bank.

Zosel, F., Haenni, D., Soranno, A., Nettels, D., and Schuler, B. (2017). Combining short- and long-range fluorescence reporters with simulations to explore the intramolecular dynamics of an intrinsically disordered protein. J. Chem. Phys. *147*, 152708.

Zúñiga, S., Sola, I., Moreno, J.L., Sabella, P., Plana-Durán, J., and Enjuanes, L. (2007). Coronavirus nucleocapsid protein is an RNA chaperone. Virology *357*, 215–227.