MULTI-SCALE STUDIES OF MOLECULAR SELF-ASSEMBLY IN BIOLOGY

Ian Seim

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Bioinformatics and Computational Biology.

Chapel Hill
2022

Approved by:

Amy S. Gladfelter

Daphne Klotsa

M. Gregory Forest

Ronit Freeman

Kevin C. Slep

Alain Laederach

ABSTRACT

Ian Seim: Multi-scale studies of molecular self-assembly in biology
(Under the direction of Amy S. Gladfelter and Daphne Klotsa)

Self-assembly of molecules into higher-order structures of specific shapes and sizes is central to biological organization. My graduate work is focused on understanding how molecular-scale interactions encode the emergent mesoscale properties in three model systems involving biomolecular condensates, viral genomic RNA packaging, and the septin cytoskeleton.

Biomolecular condensates are formed when proteins and nucleic acids undergo phase separation coupled to percolation (PSCP) which allows for compartmentalization of biochemistry without the use of membranes. Studies to date have shown that oligomerization of phase separating proteins promotes PSCP. However, we have shown that the fungal protein Whi3 contains a transiently structured region which can promote oligomerization and opposes PSCP. Stabilizing structure in this region increases dilute phase oligomerization and reduces the concentration of Whi3 in condensates. Both trends are reversed when structure formation is abrogated. Further, condensates with low dense phase protein concentrations experience an accelerated dynamical arrest, while condensates with higher proteins concentrations coarsen over longer periods of time. The relationships between dilute phase oligomerization and PSCP have implications for myriad systems, including the packaging of viral RNA genomes.

Viruses must package their genomes into nascent virions while excluding unwanted cellular and viral molecules. Recent work has shown that viral packaging proteins can undergo PSCP with and on viral genomes. The Gladfelter lab previously demonstrated that the 5′ and 3′

ends of the SARS-CoV-2 RNA genome promote PSCP of nucleocapsid protein (N-protein), whereas the middle of the genome is predicted to consist mostly of solubilizing elements. Using a coarse-grained polymer model, we show that solubilization of N-protein can be achieved if RNA-protein binding interactions are sufficiently strong, sequestering proteins into inert complexes in the dilute phase. PSCP is instead promoted by weak, multivalent interactions. We outline how the spatial patterning of these distinct RNA-protein binding modalities along the genome affect genome cyclization, compaction, and packaging into complexes with single genomes.

Septins are membrane-binding cytoskeletal proteins found in all eukaryotes except plants. Combining experiments and agent-based modeling, we show that septin polymerization is non-cooperative and explore the effects of reaction rates on higher-order septin assemblies.

ACKNOWLEDGEMENTS

Many people have supported me throughout my scientific career without whom I would not be able to write this thesis. While I was a relatively aimless undergraduate math student, I emailed Dr. Greg Forest out of the blue, asking if he knew of any research opportunities I could get involved in. He promptly answered that I could work with him. I am grateful for his patience with me while I was finishing my undergraduate degree, since I did not really contribute much during that time. After graduating though, I committed myself to the work and ended up staying for three years until starting graduate school. Working with Greg is what clarified for me that I love interdisciplinary scientific research and want to pursue it as my career.

I'd also like to thank Dr. Amy Gladfelter for giving me the chance to join her lab, in spite of the fact that I had taken no biology classes in undergrad. Working with her and her lab have shown me that molecular and cell biology is the world that I am most excited to investigate. I have even become something of an experimentalist along the way. I am forever grateful for the freedom, collaboration, and fun I have had while being a part of the Gladfelter lab.

Rohit Pappu and several members of his group have been very important teachers throughout the last three years of my time in graduate school. They have helped shape and re-shape my perspectives many times and have been fun and inspiring collaborators.

My parents have been a lifelong support and have instilled in me the belief that I can do anything. My partner and fiancée Sol has given me boundless love and support for which I am so grateful. I only hope to be able to give back what I am so lucky to have received.

PREFACE

This work contains two previously published articles and references to several articles that are either published or currently under peer review. Chapter 2 is an article published in *PNAS* on March 25, 2022. The project originated while I was graciously hosted by the Pappu lab for a week in the summer of 2019, where I first used their atomistic simulation software, CAMPARI, to study a putative coiled-coil domain in Whi3 which had piqued my curiosity. The article was completed in collaboration with Ammon Posey and Rohit Pappu from the Pappu lab at Washington University in St. Louis, as well as Wilton Snead and Benjamin Stormo from the Gladfelter lab. Ammon Posey performed all of the circular dichroism experiments and analysis, Wilton Snead performed the protein-only phase separation assays and the TIRF microscopy experiments and analysis, and Benjamin Stormo cloned the Whi3 mutants and assisted with purification. I performed all other experiments and all of the computational and theoretical work. The article has the following citation:

> Ian Seim, Ammon E. Posey, Wilton T. Snead, Benjamin M. Stormo, Daphne Klotsa, Rohit V. Pappu, Amy S. Gladfelter. Dilute phase oligomerization can oppose phase separation and modulate material properties of a ribonucleoprotein condensate. *Proceedings of the National Academy of Sciences USA* **119** (13) e2120799119 (2022).

PNAS does not require authors of manuscripts to obtain permission for use in a dissertation.

Chapter 3 is an article published in *Biophysical Journal* on July 20, 2021. The project is a follow-up to an article[1] published and driven by Christiane Iserman and Christine Roden from

---

[1] Christiane Iserman, Christine A. Roden, *et al*. Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Molecular Cell* **(80)**, 1078-1091 (2020).

the Gladfelter lab, who heroically changed scientific course almost immediately after the US lockdown in March 2020 and published their article that same year. My article comprises predictions from computational modeling based on observations in their work, so conversations with Christine and Amy Gladfelter were instrumental in my design, interpretation, and refining of the simulations. Our article has the following citation:

> Ian Seim, Christine A. Roden, Amy S. Gladfelter. Role of spatial patterning of N-protein interactions in SARS-CoV-2 genome packaging. *Biophysical Journal* **120**, 2771-2784 (2021).

Biophysical Journal does not require authors of manuscripts to obtain permission for use in a dissertation.

Appendix A is a description of an article[2] that is currently under review at *eLife*. The project began when Benjamin Woods from the Gladfelter lab asked if I wanted to collaborate and create an agent-based model of the topic he had been studying experimentally, the biophysics of septin polymerization. After receiving reviews, I realized that several elements of the paper required significant re-thinking. As a result, I am currently in the process of performing several experiments and re-parameterizing my model to address all concerns. The project is therefore in a state of flux and is not ready for drafting or inclusion as a full chapter in this thesis. For this reason, I have included only a short description of the main findings of the study and the tools I have developed during the course of the project in Appendices A and C, respectively. A revised version will be resubmitted this year.

---

[2] Benjamin L. Woods, Ian Seim, Jessica Liu, Grace McLaughlin, Kevin S. Cannon, Amy S. Gladfelter. Biophysical properties governing septin assembly. *bioRxiv* https://doi.org/10.1101/2021.03.22.436414 (2021).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

a.u.	arbitrary unit
CC	coiled-coil
CD	circular dichroism
FE RNA	frameshifting element RNA
FRAP	fluorescence recovery after photobleaching
gRNA	genomic RNA
IDR	intrinsically disordered region
LLPS	liquid-liquid phase separation
mRNA	messenger RNA
N-protein	nucleocapsid protein
ORF	open reading frame
PSCP	phase separation coupled to percolation
QRR	glutamine rich region
RBD	RNA binding domain
RNP	ribonucleoprotein
RRM	RNA recognition motif
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SOD	structured oligomerization domain
TIRF	total internal reflection fluorescence
UV-CD	ultraviolet circular dichroism
WT	wild type

CHAPTER 1: INTRODUCTION

Phase separation coupled to percolation (PSCP) is a ubiquitous mechanism by which cells organize macromolecules without the use of membranes (1). The simplest manifestations of phase separation and percolation theories describe systems at thermodynamic equilibrium. Although cells are clearly non-equilibrium systems on the whole, there is evidence that an equilibrium thermodynamics framework is a valid approximation of cellular processes that take place at appropriate time and length scales, including PSCP (2). Although there has been significant theoretical research into active processes that can affect PSCP (3, 4), the majority of the works in the field, including those in this thesis, have used the equilibrium framework. Therefore, a brief description of phase separation and percolation theories at thermodynamic equilibrium is warranted.

Liquid-liquid phase separation (LLPS) is traditionally the theory that has been used to describe the formation of biological condensates (5). LLPS describes a density transition in which one or more molecular species within a solvent spontaneously overcome the entropy of mixing and self-assemble into locally dense, liquid phases that coexist with a dilute phase (5). Such self-assembly is driven by sufficiently attractive interactions between molecules. If the attraction is strong enough, the system can minimize its free energy by forming coexisting dense and dilute phases of the molecules in question, despite the entropic penalty. An important assumption in this theory is that both dense and dilute phases are simple liquids, i.e. they are fully explained by viscosities and have zero elasticity. As will be shown, this assumption is generally invalid for biological condensates.

In order for phase separation to occur, a system must undergo a sufficiently large change in a relevant thermodynamic parameter which may be the concentrations of molecules, temperature, pH, etc. (4). By definition, this change requires energy input and is therefore a non-equilibrium process. However, in practice, phase separation experiments in a biological context are often initiated by adding proteins and/or RNAs to a solvent and observing their behavior over time without further energy input. In this sense, equilibrium thermodynamics govern the evolution of the system just after the change in the relevant thermodynamic parameter(s) (here, concentrations of proteins and RNAs). It is in this sense that the field of biological condensates refers to the equilibrium thermodynamics of phase separation.

In general, the free energy of a system at the beginning of the phase separation process can be either globally unstable or metastable depending on the new values of the altered thermodynamic variables that initiated the process (4). A globally unstable free energy may, for example, arise from adding very high concentrations of molecules with sufficiently high binding affinities for each other. This system will undergo spinodal decomposition, and a dense phase will spontaneously form everywhere throughout the reaction volume, coarsening due to the drive to minimize surface tension until it reaches its final equilibrium state. A metastable free energy may arise when lower concentrations of molecules are added. In this case, dense phases form via a nucleation and growth mechanism in which molecules stochastically form clusters that are occasionally large enough to overcome surface tension. These clusters gradually grow larger into a system of droplets that again coarsen due to the drive to minimize surface tension, with a final equilibrium state being a single, spherical droplet. Since each given set of compatible macromolecules are at relatively low concentrations within cells, phase separation in a biological context almost always resembles a nucleation and growth mechanism as evidenced by the characteristic droplet morphology of the dense phase.

However, phase separated droplets within cells almost never reach their expected equilibrium state of a single, large, spherical droplet, instead remaining as an emulsion of smaller condensates (6). Although this important discrepancy can be due to active processes (7), surface adsorption (8, 9), or modification of interfacial properties (10) within cells, emulsions are also observed *in vitro* when none of these factors are present. This observation is incompatible with the assumption of LLPS that phase separated droplets are simple liquids. Instead, biological condensates are formed by large, macromolecular polymers which can transiently crosslink to form networks of interactions. When these interaction networks grow to span a system (a condensate, for example), the system has undergone a connectivity phase transition called a percolation transition (11). Due to this connectivity, the condensate obtains non-zero elasticity and is more precisely characterized as a viscoelastic fluid which exhibits solid-like behavior on time scales shorter than its viscoelastic relaxation time, and liquid-like behavior on longer timescales (12). Indeed, evidence is mounting that even the simplest reconstituted biological condensates are viscoelastic fluids (13, 14). These condensates can also have viscoelastic relaxation times that increase over time, eventually resulting in elastic responses on experimentally relevant timescales which leads to dynamical arrest before the equilibrium state predicted by LLPS is reached (13). In this case, many droplets can coexist at steady state, as is seen both in cells and in reconstituted systems *in vitro*. Since biological condensates undergo both density and connectivity transitions, the term “phase separation coupled to percolation” most accurately describes the processes underlying their formation.

In cells, biological condensates probably comprise hundreds to thousands of protein and nucleic acid species (15). Although there has been significant theoretical progress in understanding these hugely multi-component systems (16-18), full experimental characterization of these systems is untenable for *in vitro* reconstitution assays due to the number of molecular species. Fortunately, many biological condensates feature so-called "scaffold" components which are necessary and sufficient for condensate formation in *in vitro* reconstitution assays (15, 19). These relatively simple reconstituted systems have provided a wealth of insight into the principles underlying PSCP and biological condensates due to synergy between theoretical models and physical frameworks tested in experiments. The works in this thesis have leveraged this crosstalk between theory and relatively simple reconstitution experiments. However, the principles identified in these simpler systems always need to be tested in cells. Consideration of the complexities added by the cellular environment remains an essential task for fully understanding biological condensates. Nevertheless, the focus of this thesis is understanding fundamental principles underlying the self-assembly of phase separating proteins and RNAs using focused reconstitution experiments and computational models. The remainder of this chapter will lay out some of the established principles of protein and RNA interactions that are implicated in PSCP that have largely been gleaned from reconstitution experiments and computational modeling. Several of these principles are challenged and complicated by the following chapters in this thesis.

Commonly, biological condensates comprise RNAs and RNA-binding proteins with intrinsically disordered regions (IDRs) and structured oligomerization domains (SODs) (1, 20). The prevailing viewpoint is that IDRs drive protein phase separation via weak, multivalent interactions, their SODs enhance phase separation by concentrating IDRs and lowering the entropic penalty of forming a macroscopic dense phase, and that RNAs act as scaffolds that promote co-phase separation with proteins (1, 21, 22). However, these rules are overly simplistic and do not capture the full range of observed behaviors, as will be demonstrated in Chapters 2 and 3. Key variables that complicate these viewpoints are the valence of the participating molecules and the affinities of their binding sites.

Valence has been established as a critical variable in determining whether or not and to what extent biomolecules will phase separate (23). Indeed, *in vitro* experiments have demonstrated valence thresholds below which proteins and RNA molecules will not phase separate (19), and simulations have demonstrated that the "liquid network connectivity", or the time-averaged number of interaction partners for a given molecule within the dense phase, is highly correlated with dense phase stability and composition (24). Additionally, the terms "scaffold" and "client" have been used to describe phase separating molecules with high and low valence, respectively, with scaffolds driving phase separation, and clients modulating dense phase stability (19). However, the definitions of valence differ among studies, and the underlying molecular features that control valence are complex and not fully understood. Non-covalent bonds among polymers lead to transient associations whose lifetimes generally depend on binding kinetics and local concentrations of binding sites (25). Binding site patterning and polymer architectures including branching, length, and rigidity all contribute to local concentrations of binding sites (25).

The effects of binding site patterning in IDRs have been well-studied. Significant work has been done to identify binding sites within IDRs, with PSCP-promoting sites commonly referred to as "stickers", separated by solvated regions called "spacers" (1, 26, 27). The patterning of stickers along IDRs can significantly impact polymer conformations, phase behavior, and condensate material properties (27, 28). These effects have been explained in terms of discrepancies between the “apparent valence” and the “effective valence” of the IDRs (1). The apparent valence refers to the absolute count of stickers on an IDR, while the effective valence quantifies the number of effective binding sites on an IDR. The effective valence can be lower than the apparent valence if stickers are clustered. In this case, each cluster is more accurately characterized as a single binding site with a higher binding affinity relative to a single sticker (1). A computational study by Rana et al. predicts that in IDRs with sufficient clustering of stickers, PSCP can be bypassed altogether, and instead, polymers aggregate without undergoing a phase transition (29). In polymers with identical apparent valence but less sticker clustering, PSCP is observed (29). In a simple model of the SARS-CoV-2 genome interacting with nucleocapsid proteins, Cubuk et al. show that a single high affinity binding site placed among lower affinity binding sites within the disordered polymer representing the genome leads to dynamically arrested complexes of single genomes and proteins that do not coalesce into a macroscopic dense phase (30). Without this high affinity binding site, a macroscopic dense phase forms (30). This high affinity binding site may represent either a dense cluster of stickers or a distinct, high-affinity binding motif in a real biomolecule. These studies suggest that alterations to the effective valence and the binding affinities of effective binding sites within IDRs lead to a spectrum of self-assembly pathways ranging from small complexes and aggregates to macroscopic dense phases formed by PSCP.

In addition to IDRs, many phase separating proteins contain SODs that enable protein-protein binding (31). The oligomerization domains of many of these proteins are required to form condensates, including the NTF2L dimerization domain of G3BP1/2 (32), which is involved in stress granule formation, the oligomerization domain within coilin that drives Cajal

Body formation in the nucleus (33), and the PB1 domain of auxin response factors which form condensates in plants during stress (34). Engineered systems with IDRs tethered to SODs have been used to trigger PSCP in cells (21, 35, 36). In these systems, oligomerization via folded domains is thought to create macromolecular clusters with increased local concentrations of IDRs and thus higher valence and phase separation propensity (1). However, the study in Chapter 2 of this thesis shows that a transiently structured oligomerization domain within the IDR of the fungal protein Whi3 does not fit in with this framework. Instead, as the structure of the oligomerization domain is stabilized, Whi3 proteins are sequestered into inert oligomers and protein concentrations within the dense phase are lowered. This observation is reminiscent of the computational studies of IDRs demonstrating that polymers which are modified to have lower effective valence but higher affinity binding sites can self-assemble using mechanisms distinct from PSCP. In the case of Whi3, the alternative pathway is dilute phase oligomerization.

RNA and RNA-binding proteins are commonly found in biomolecular condensates (20), and the prevailing viewpoint is that RNAs act as scaffolds that promote co-phase separation with proteins (22). Proteins are known to bind RNAs using several folded domains, including RNA recognition motifs (RRM), Zinc fingers (ZnF), hnRNP K homology domains (KH), and double-stranded RNA binding motifs (dsRBMs) (37). Proteins can also bind RNA using unstructured or transiently structured sequences. Examples include RGG/RG motifs (38) and transiently structured helices within IDRs of SARS-CoV-2 N-protein (30). Positively charged regions in proteins can also non-specifically interact with negatively charged RNA. These varied modes of association lead to a wide range of reported phase behaviors when RNA is involved. Fundamentally, phase separation of multiple components, i.e. protein and RNA, is governed by distinct physical principles than those of simpler binary systems, i.e. those with only a single protein species (39, 40). For example, protein and RNA systems can have a re-entrant character, where too little or too much RNA can prevent otherwise robust phase separation (39, 40). Evidence for a functional role of re-entrance has been observed in the nucleus, where high RNA concentrations presumably limit the aggregation of prion-like proteins by keeping them soluble (41). Recent work from the Gladfelter lab has identified an RNA that solubilizes protein in an RNA sequence/structure-dependent manner. A fragment of the SARS-CoV-2 gRNA, the frameshifting element (FE RNA), dissolves pre-formed nucleocapsid protein (N-protein) condensates (42). Interestingly, a different fragment of identical length, the 5′ end RNA, promotes robust phase separation with N-protein (42). Chapter 3 of this thesis predicts that these differing modes of assembly can be accessed by altering the effective valence and the binding affinity of effective binding sites within the RNAs. Sufficiently strong, anisotropic interactions between RNA and protein molecules sequester both into inert complexes that oppose PSCP, while low-affinity, isotropic interactions promote PSCP. The rest of Chapter 3 investigates the effects of incorporating both of these binding modalities into large RNA molecules that represent the RNA genome of SARS-CoV-2. Clearly, RNA does not always act as a scaffold for protein PSCP and can instead favor behavior ranging from dissolution to condensation.

Proteins and RNAs that can undergo PSCP can have modular architectures with domains that promote PSCP coexisting with domains that promote alternative self-assembly pathways. Chapters 2 investigates the consequences of these coexisting architectures within a protein-protein interaction context and Chapter 3 explores multi-domain RNAs that balance PSCP and smaller-scale assembly with proteins.

CHAPTER 2: DILUTE PHASE OLIGOMERIZATION CAN OPPOSE PHASE SEPARATION AND MODULATE MATERIAL PROPERTIES OF A RIBONUCLEOPROTEIN CONDENSATE[3]

Introduction

Intracellular phase transitions are known to drive the formation and dissolution of membraneless biomolecular condensates (43). Prominent condensates include ribonucleoprotein (RNP) bodies such as stress granules (31, 32, 44, 45) and processing bodies (45, 46) in the cytosol and nuclear bodies such as speckles (47) and nucleoli (48). RNP condensates can form via phase transitions that are influenced by a combination of homotypic and heterotypic protein-protein, protein-RNA, and RNA-RNA interactions (31, 39, 41, 44, 49-51).

Biomolecular condensates form through phase transitions known as phase separation coupled to percolation (PSCP) (1, 52-54). The stickers-and-spacers framework provides a way to describe PSCP (1, 26, 40, 50, 54, 55). Phase separation is a density transition driven by non-specific, solvent-mediated interactions, whereas percolation is a networking transition driven by specific interactions among stickers (54, 56). Percolation transitions occur when protein and RNA molecules are topologically connected into system- or condensate-spanning networks. The solubility profiles of spacers contribute to density transitions, thereby determining whether percolation is aided by phase separation (54, 56).

---

[3] This chapter previously appeared as an article in the Proceedings of the National Academy of Sciences. The original citation is as follows:
Ian Seim, Ammon E. Posey, Wilton T. Snead, Benjamin M. Stormo, Daphne Klotsa, Rohit V. Pappu, Amy S. Gladfelter, Dilute phase oligomerization can oppose phase separation and modulate material properties of a ribonucleoprotein condensate. *Proceedings of the National Academy of Sciences USA* **119** (13) e2120799119 (2022).

Recent advances have identified the molecular features within protein (27, 57-60) and RNA sequences (20, 61) that drive PSCP. These features include but are not restricted to: the number (valence) of folded stickers (19, 53, 62) such as RNA recognition motifs (RRMs) (26, 63) and protein-protein interaction domains (43); the valence and patterning of short linear motifs (stickers) within intrinsically disordered regions (IDRs) of proteins (27, 47, 57, 59, 64); the physical properties of spacers that are interspersed between stickers (1); and linear interaction motifs, sequence-specific secondary structures, the degree of unfoldedness, and purine versus pyrimidine contents within RNA molecules (20, 44, 49, 50, 61, 65).

Proteins that drive the formation of RNP condensates via PSCP often have modular architectures featuring structured oligomerization domains (SODs), IDRs, and RRMs (31). There are several examples of proteins that contain both IDRs and SODs. Prominent among these are the dimerization domain of the Nucleocapsid protein in SARS-CoV-2 (30, 42), the NTF2L dimerization domain of G3BP1/2 (32), a scaffold for stress granule formation, the oligomerization domain within coilin that drives Cajal Body formation (33), and the PB1 domain of auxin response factors (34) which gives rise to condensates that repress transcription during stress in plants. Oligomerization via SODs tethered to IDRs has also been shown to drive phase transitions in engineered systems (21, 35, 36). Synthetic and naturally occurring SODs are thought to act as sinks that lower the entropic penalty associated with phase transitions (21). In this view, SODs enable increased multivalence and the local concentrations of stickers that in turn drive phase transitions.

It is well known that many IDRs undergo coupled binding and folding reactions either through heterotypic interactions with cognate binding partners or via homotypic interactions that lead to dimers and higher-order oligomers (66-70). An archetypal example of dimerization induced folding of an IDR is the leucine zipper motif, which has a weak intrinsic preference for alpha-helical conformations and forms coiled-coil alpha-helical dimers (70, 71) or higher-order oligomers (72). Residues that define the interface among helices contribute to the oligomerization status of these leucine zipper systems. Overall, the stability of coiled coils is thermodynamically linked to the intrinsic preference for forming alpha-helices as monomers (73). How might sequence-specific preferences for alpha-helical conformations within IDRs and their oligomerization impact the phase behavior of RNA-binding proteins? Recent studies have identified the propensity to form alpha-helical structures within the C-terminal IDR of TDP-43 (74). This has been shown to be a key driver of phase separation (74). The findings for TDP-43 are reminiscent of findings regarding the effects of oligomerization driven by SODs. Accordingly, the consensus is that oligomerization via SODs enhances the driving forces for phase separation. Here, we uncover a surprising opposition to phase separation that derives from oligomerization driven by local ordering within an IDR.

The fungal protein Whi3 forms RNP condensates with different RNA molecules (49, 75). The simplest reconstituted system has two macromolecular components namely, the Whi3 protein and an RNA molecule such as the cyclin mRNA, *CLN3*. The Whi3 protein encompasses a C-terminal RRM and a long adjacent IDR that includes a glutamine-rich region (QRR) (**Fig. 2.1A**). Sequence analysis shows that the QRR contains a 21-residue motif, which we designate as CC. This CC motif has many of the features associated with alpha-helical coiled-coil forming domains (76). This opens the door to the possibility that the phase behavior of Whi3 is influenced by this putative helix- and coiled-coil-forming domain. Our work focuses on an interrogation of the effect of the CC motif on Whi3 phase behavior. We characterized the intrinsic helicity of the CC motif, in isolation, and assessed its impact on the phase behavior of

Whi3 through comparative studies of conformational and phase equilibria of mutants that diminish (CC$^-$) or enhance (CC$^+$) helicity. Surprisingly, we find that while enhancing intrinsic helicity enhances oligomerization in the dilute phase, it lowers the density of Whi3 in the RNP condensates. The converse is true when helicity is diminished. Spectroscopic investigations help establish the importance of disorder in the QRR for driving the intermolecular interactions that promote phase separation. Additionally, using coarse-grained simulations, we show how oligomerization and clustering within the dilute phase can modulate phase behavior in agreement with experimental observations. The main finding is that dilute phase oligomers act as off-pathway sinks for Whi3 proteins, whereas disorder within the QRR promotes phase separation via homotypic interactions.

Results

The QRR of Whi3 encompasses a region with a weak intrinsic preference for alpha-helical conformations

The Whi3 protein encompasses a distinctive, 108-residue QRR that is predicted to be disordered (**Fig. 2.1A**). Sequence analysis revealed the presence of a 21-residue stretch, designated as CC, that has the characteristic features of domains that form alpha-helical coiled-coils (76, 77). These features include the absence of Pro and the presence of Leu at every seventh position (**Fig. 2.1A**). The putative CC motif appears to be an example of alpha-helical molecular recognition elements / features (α-MoREs or α-MoRFs) that are found in many IDRs (78, 79).

The sequence of the CC peptide is shown in **Fig. 2.1A**. We quantified the sequence-specificity of the alpha-helicity of the CC peptide by introducing point mutations that are likely to diminish / enhance helicity. The two designed peptides are designated as CC$^-$ and CC$^+$, respectively. The three Leu residues in the original CC peptide were replaced with Pro to generate CC$^-$, a sequence variant designed to have diminished alpha helicity. To generate a variant designated as CC$^+$, we replaced the fourth residue of each heptad repeat in the CC peptide with a Leu.

We performed ultraviolet circular dichroism (UV-CD) measurements for the three peptides (see methods for details). These spectra were collected at different temperatures at concentrations ranging from 600-900 µM; the data are shown in **Figs. 2.1B-D**. Across the temperature range of interest, the spectra for the CC peptide show two minima at 208 nm and 222 nm, respectively. The molar ellipticity is lower for the minimum at 208 nm. This feature is consistent with a heterogeneous ensemble of conformations albeit with alpha-helical character (80, 81). In contrast, and in accord with our predictions, each of the spectra for CC$^-$ show only a single minimum at ~205 nm. There is a positive peak in the 215 – 218 nm range at low temperatures and this is concordant with the presence of Pro residues within the sequence (82). Overall, the spectra for CC$^-$ are consistent with this peptide being disordered.

Finally, the spectra for the CC$^+$ peptide demonstrate the characteristic minima at 208 nm and 222 nm. In addition, at temperatures below 338 K, the molar ellipticity at 222 nm is lower than the molar ellipticity at 208 nm. The molar ellipticities between 200 nm and 240 nm become less negative with increasing temperature and the presence of an isodichroic point at 200 nm is suggestive of CC$^+$ showing a canonical helix-to-coil transition (83) as temperature increases. By fitting the molar ellipticity at 222 nm for each peptide to a two-state model, we estimated the melting temperatures for helix-to-coil transitions to be 333 K and 243 K for CC$^+$ and CC, respectively (**Fig. S2.1**, *SI Appendix*). This implies that CC populates a heterogenous conformational ensemble at physiological temperatures, whereas CC$^+$ is considerably more alpha-helical under these conditions. In contrast, CC^- has negligible preference for alpha-helical conformations.

We obtained an estimate of the secondary structure content of each peptide as a function of temperature using singular value deconvolution of the CD spectra via the CDSSTR algorithm, as implemented through the DichroWeb server (84). This analysis uses reference CD spectra of proteins with known structural content as basis sets from which to reconstruct the CD spectrum of interest. The accuracy of this approach is dependent on the extent to which the structural attributes of the protein of interest are represented in the basis set of proteins. Short peptides, coiled-coils, disordered regions, and aggregated or highly assembled states of proteins are either poorly represented or absent from these basis sets. Accordingly, we interpret the estimates of helical contents with a measure of caution. The fits are shown in **Figs. S2.2A-C**. The fractional helicities are highest at low temperatures and exhibit a single unfolding transition at high temperatures for CC and CC^+. We performed forward and reverse temperature CD scans for the CC and CC^+ peptides and did not find evidence of hysteresis in the fractional helicity for either peptide (**Figs. S2.2D, E**). Overall, the CD experiments suggest that the CC peptide has a partially helical character, CC^- has negligible helicity, and CC^+ is strongly helical.

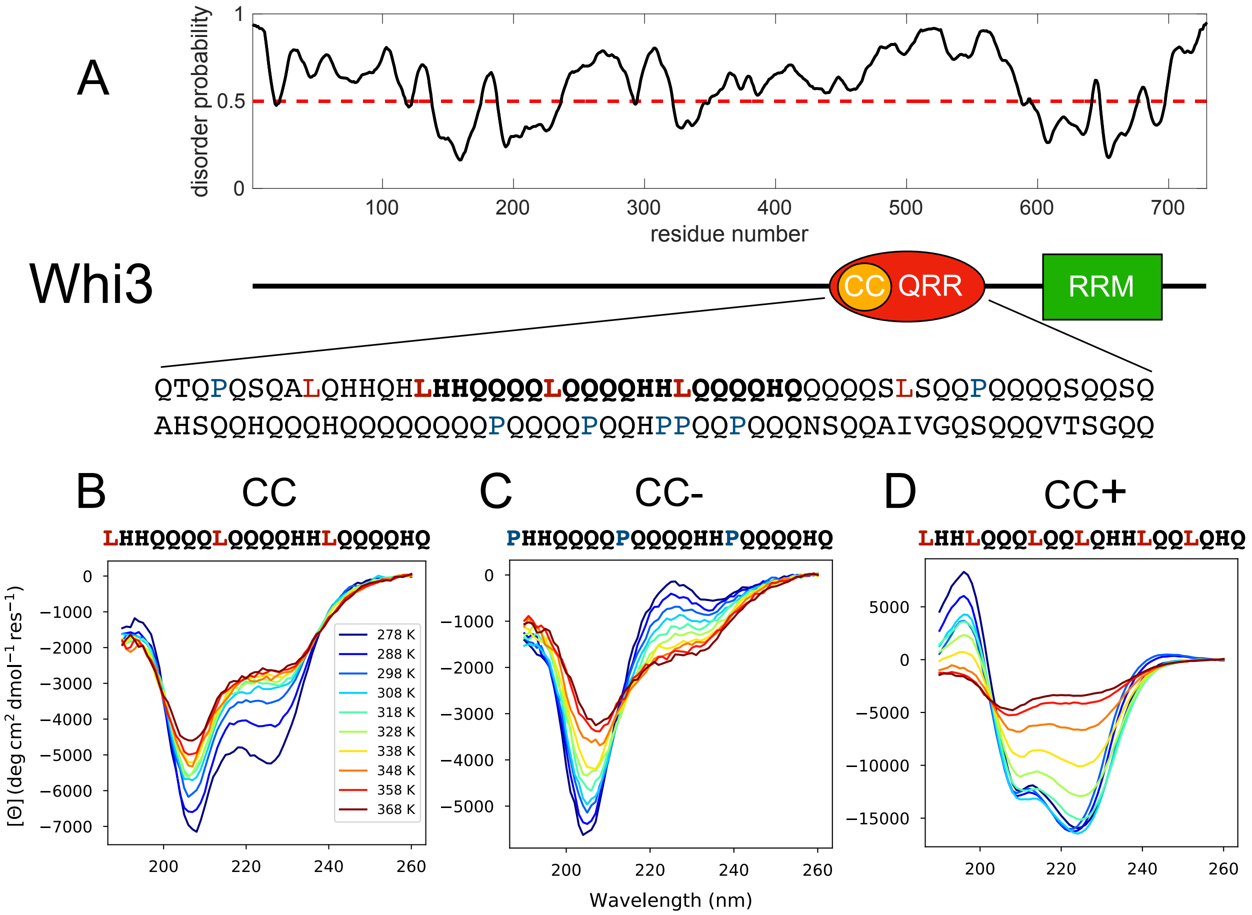


Figure 2.1: The QRR of Whi3 encompasses a putative CC motif. (A) Whi3 is predicted to be highly disordered (PrDOS) (85). The sequence includes a 108-residue QRR. (B-D) UV-CD spectra, measured at a series of temperatures, for the CC, CC$^-$, and CC$^+$ peptides, respectively. Peptide concentrations range from 600 – 900 µM.

The phase behavior of Whi3 is modulated by changes to the CC motif

To assess the impact of the CC motif on the phase behavior of Whi3, we studied the co-phase separation of Whi3 in the presence of the cyclin RNA *CLN3* (49) and compared this to the measured phase behavior obtained for three distinct variants. We expressed and purified full-length wild-type Whi3 (WT), and three variants namely, Whi3 (CC $\rightarrow$ CC$^-$), Whi3 (CC $\rightarrow$ CC$^+$), and Whi3 (ΔCC). In the Whi3 (ΔCC) protein, we deleted the 21 residues of the CC motif within Whi3. In each of our measurements, the Whi3 proteins and *CLN3* RNA molecules have fluorescent dye labels, and these are used to assess dense phase concentrations of protein and

RNA molecules. In each mixture, 5% of the Whi3 is labeled and a single *CLN3* stock is used, such that the fluorescence values for both molecules can be directly compared across experiments.

All mutants of the Whi3 protein can still undergo phase separation with *CLN3*. However, the morphology and composition of assembled droplets varied amongst constructs. Notably, changes to the CC sequence resulted in significant differences in the dense phase stoichiometry (**Fig. 2.2A**). We quantified the concentrations of Whi3 in the dense phase by measuring fluorescence intensities and comparing to dye calibration curves (**Fig. S2.3A**, *SI Appendix*). When compared to condensates formed by WT Whi3 and *CLN3*, the condensates formed by Whi3 (ΔCC) have 1.5 times higher dense phase protein concentrations and Whi3 ($CC \rightarrow CC^-$) has 2 times the protein concentration in the dense phase. In contrast, the protein concentration in condensates formed by Whi3 ($CC \rightarrow CC^+$) is 0.2 times that of WT (**Fig. 2.2B**). This suggests that increased helicity in the CC motif leads to lower concentrations of Whi3 in the dense phase. Conversely, a disordered CC motif or deletion of the CC motif increases the dense phase concentration of Whi3.

An important consequence of the altered protein concentrations is a change in the ratio of protein to RNA that depends on the CC sequence. The dense phase concentrations of *CLN3* are similar for condensates formed by WT Whi3, Whi3 (ΔCC), and Whi3 ($CC \rightarrow CC^-$). However, the dense phase concentration of *CLN3* in condensates formed by co-phase separation of Whi3 ($CC \rightarrow CC^+$) with *CLN3* is about 0.8 times the dense phase concentration of *CLN3* in the WT system. Based on these results, we hypothesized that the effects of mutations within the CC motif should be evident even in the absence of *CLN3*. To test this hypothesis, we performed phase separation assays in the absence of *CLN3*. In line with results from the Whi3-*CLN3* system, we observe robust phase separation for Whi3 ($CC \rightarrow CC^-$), intermediate levels of phase separation for WT, and weak phase separation for Whi3 ($CC \rightarrow CC^+$) (**Fig. 2.2C, D**). In fact, at equivalent bulk protein concentrations of 17 μM, Whi3 ($CC \rightarrow CC^+$) forms very few droplets, and those that form are very small (**Fig. 2.2D**). Relative to WT, the dense phase concentration increased by a factor of 1.2 for Whi3 ($CC \rightarrow CC^-$) whereas the Whi3 ($CC \rightarrow CC^+$) concentration was only 0.4 times that of WT. These results are qualitatively consistent with results from the Whi3-*CLN3* system. We conclude that homotypic interactions, driven by the CC motif, which we modulate using mutations, can alter the intrinsic phase behavior of Whi3 and its co-phase separation with *CLN3*.

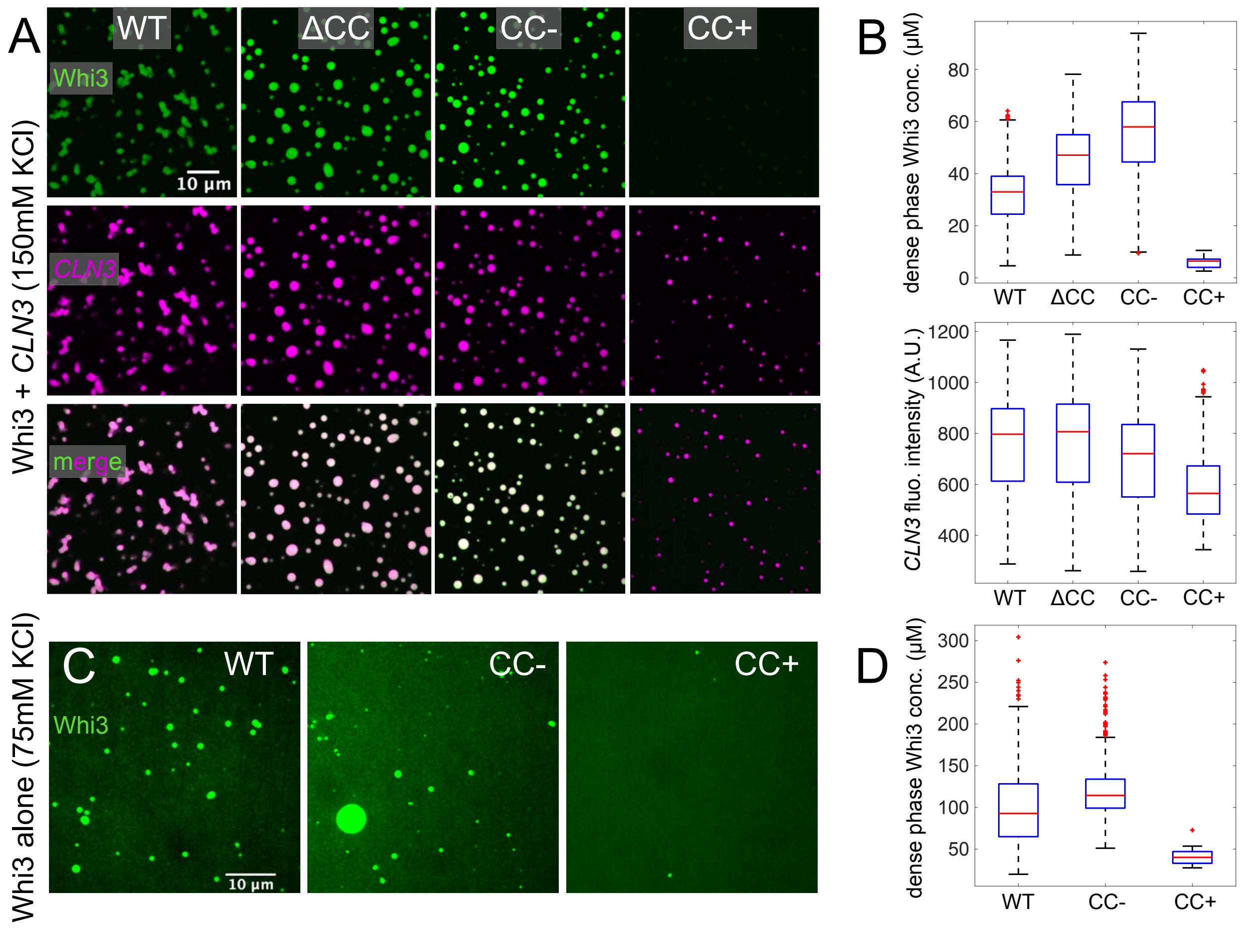


Figure 2.2: CC mutations alter Whi3-*CLN3* phase behavior. (A) Representative images of each protein at 5 hours after mixing 500 nM Whi3 with 5 nM *CLN3*. Whi3 (green) and *CLN3* (magenta) channels are contrasted identically in each image. (B) Box plots of dense phase Whi3 concentration and *CLN3* fluorescence intensity distributions of condensates in experiments corresponding to panel (A). Red lines indicate medians, edges of blue boxes indicate 25% and 75% quantiles, black lines at the edges of the dotted lines indicate the minimum and maximum observed values, and red stars indicate outliers. (C) Whi3-only phase separation at a bulk concentration of 17 µM recorded 4 hours after mixing. Images are contrasted such that fluorescence intensity is proportional to protein concentration. (D) Box plots of dense phase Whi3 concentration distributions for each condensate in experiments corresponding to panel (C).

The CC motif enables sequestration of the Whi3 protein into dilute phase oligomers

The apparent opposition between oligomerization strength and the propensity to phase separate was unexpected. To arrive at a mechanistic understanding for these observations, we modeled the CC and QRR domains of Whi3 using the LASSI stickers and spacers model (40) (**Fig. 2.3A**). For simplicity, we focused on a minimal system comprising the CC-QRR fragment. We assume that CC is a dimerization motif that can access both dimerization competent and dimerization incompetent states. This assumption is motivated by the heterogeneous conformational distributions inferred from the CD measurements. In this model, the CC dimerizes with other CC domains via anisotropic interactions. This represents interactions between folded molecules. In addition, the CC motif functions like the rest of the QRR and engages in weak, isotropic interactions representing the disordered state. To represent the effects of mutations on the helicity and dimerization of the CC motif, we implement a trade-off between the energies associated with isotropic versus anisotropic interactions involving the beads that mimic the CC motif in the model (**Fig. 2.3A**). The right-most system in **Fig. 2.3** represents the extreme case that is most like Whi3 ($CC \rightarrow CC^{+}$), with stable helices enabling strong anisotropic binding. The central system corresponds to the opposite limit, mimicking Whi3 ($CC \rightarrow CC^{-}$), with a completely unfolded CC motif that interacts purely via isotropic interactions. The WT system has a CC bead that interacts via a combination of weaker anisotropic and isotropic interaction energies.

For each system, we performed a series of LASSI simulations across a range of concentrations and temperatures. We computed coexistence curves using the approach prescribed by Choi et al. (40). The computed phase boundaries show that changes to the interactions mediated by the CC bead are sufficient to alter the driving forces for phase separation (**Fig. 2.3B**). We observe a suppression of phase separation when the isotropic interactions are weakened in favor of strengthening the anisotropic interactions. The converse is true when the isotropic interactions are strengthened over anisotropic interactions. This behavior is consistent with results from measurements that were performed in the absence of *CLN3* (**Fig. 2.2D**).

For each of the simulations, we quantified the distributions of cluster sizes in the coexisting dilute phases when phase separation occurred, and for the homogeneous system that lies outside the phase boundary. Enhancement of CC dimerization contributes to an increase in the sizes of dilute phase oligomers inside and outside the two-phase regime (**Fig. 2.3C**). Although the CC beads can only form dimers, they are connected to QRRs within the chain molecules. Accordingly, increasing the strength of anisotropic interactions mediated by the CC increases the local concentration of appended QRRs, leading to enhanced interactions among QRRs. This leads to large, dilute phase oligomers that act as sinks for Whi3 molecules. These sinks weaken the driving forces for Whi3 phase separation and may be viewed as “off-pathway” species as far as Whi3 phase separation is concerned.

To test predictions from the LASSI simulations, we measured oligomer sizes for WT, Whi3 ($CC \rightarrow CC^{-}$), and Whi3 ($CC \rightarrow CC^{+}$) at low concentrations that correspond to the one-phase regime (**Fig. 2.3D**). Using the fluorescence of single Whi3 molecules as a calibration, we recorded the distribution of oligomer sizes for each protein. Our predictions from LASSI-based simulations are supported by the observation that, despite their greatly reduced phase separation ability, Whi3 ($CC \rightarrow CC^{+}$) constructs form a heterogeneous distribution of oligomers in the dilute phase.

Taken together, our results indicate that the helicity enabled oligomerization, driven by the CC motif, can help sequester the Whi3 protein into the dilute phase, thereby impacting the driving forces for phase separation and altering the dense phase Whi3 concentration. By changing the helicity of CC, the balance of protein between the two phases is shifted. This is observed in the different Whi3-*CLN3* systems, indicating that the protein density in Whi3-RNA condensates is impacted, in part, by homotypic associations among Whi3 molecules that give rise to finite-sized clusters in the dilute phase.

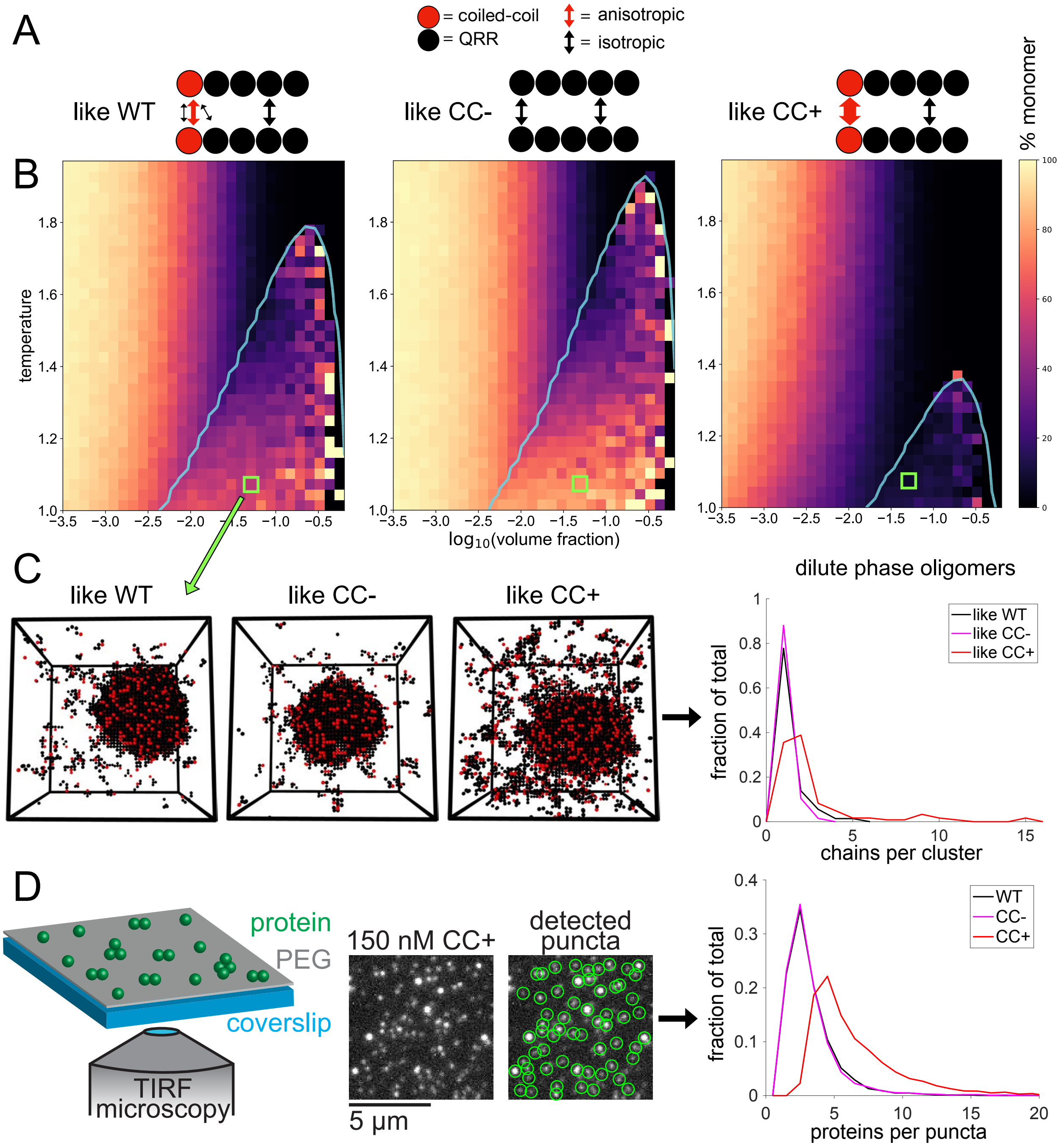


Figure 2.3: Oligomerization weakens phase separation and vice versa. (A) The CC-QRR fragment is modeled as five beads, with each bead corresponding to ~20 residues. The thickness of arrows corresponds to the strength of interactions. (B) Plots show binodals (cyan) overlaid on heatmaps indicating the percentage of the dilute phase / homogeneous phase that is monomeric. (C) Snapshots of simulations from regions corresponding to green squares in (B) are shown, and the dilute phase oligomer distributions are quantified for these snapshots. (D) Whi3 dilute phase cluster sizes at low concentrations are quantified via TIRF microscopy and particle detection.

Evidence for loss of helicity driving phase separation

Our data suggest that helix formation within the CC motif is linked to oligomerization that enables sequestration of Whi3 molecules in dilute phase oligomers that form via CC-mediated interactions and are further stabilized by interactions among QRRs. In contrast, a lack of helicity within the CC motif drives phase separation, which leads to higher protein concentrations in dense phases characterized by non-stoichiometric assemblies. Inspired by the work of Urry and coworkers (86, 87) and that of Duysens (88), we tested the hypothesis that loss of helicity through the CC motif drives phase separation. For this, we turned to UV-CD measurements and collected spectra as a function of peptide / protein concentration at room temperature. The constructs studied include the three peptides CC, CC$^-$, and CC$^+$ as well as the full-length Whi3 WT protein. Systems that undergo phase separation are known to show a strong distortion of the CD spectra that consists of a flattening of the spectra and the appearance of a new minimum at ~230 nm (88). Urry and coworkers (86, 87) noted that these features of CD spectra are related to the increased sizes of particles in solution with increasing protein concentration. Accordingly, we used UV-CD spectroscopy to answer the following questions: Do we observe the predicted distortions namely, flattening of spectra and the appearance of an auxiliary minimum near 230 nm at high peptide / protein concentrations? And are these changes, which point to the onset of phase separation, accompanied by a loss of helicity?

The results of our CD titrations are shown in **Fig. 2.4**. Data for the CC peptide show the flattening of the spectra and the appearance of a new minimum near 230 nm (**Fig. 2.4A**). These changes are manifest above a threshold concentration that is a precursor of phase separation. The onset of the distortion of CD spectra is also observed for the CC$^-$ and CC$^+$ peptides (**Figs. 2.4B, C**). The CD spectra for the CC$^-$ peptide, which has negligible helicity, show monotonic changes as a function of peptide concentration with distorted spectra being manifest at the lowest concentrations of the three peptides studied here (**Fig. 2.4B**). For CC^+, the distortion follows a loss of helicity, which is realized as the concentration goes above a threshold (**Fig. 2.4C**).

To understand the interplay of helicity and higher-order assembly in the CC and CC^+ peptides, we defined two metrics based on the concentration-dependent CD spectra. The ratio of ellipticities at 222 and 208 nm as a function of concentration provides a readout of coiled-coil formation and is a proxy for dimerization (89) (**Fig. 2.4D**). Specifically, coiled coil formation becomes more probable as the ratio of ellipticities at 222 and 208 nm exceeds one. As expected, CC^+ forms coiled coils at much lower concentrations than WT, depicted as c_{cc} in **Fig. 2.4D**. Our second metric is defined as the ratio of the ellipticities at 230 and 208 nm as a function of concentration (**Fig. 2.4E**). This provides a measure of higher-order assembly, which may include bundles of helices and / or coiled coils. These data are well-fit by a 2-state model, with each peptide showing a higher c_{m} for higher-order assembly than for coiled-coil formation. Again, CC^+ forms higher-order assemblies at lower concentrations than CC, which is consistent with data in Fig. 3 showing enhanced dilute phase oligomerization for the corresponding full-length protein. We calculated the secondary structure propensities for the spectra from each peptide using the CDSSTR algorithm (**Fig. S2.4**). CC is partially helical at low to intermediate concentrations, CC^- demonstrates negligible helicity, while CC^+ is strongly helical at all but the highest concentrations.

We can compare the extent to which each of the peptides drive phase separation through unfolded conformations by comparing the spectra collected at similar, high peptide concentrations (**Fig. 2.4F**). This comparison reveals that the extent of distortion and flattening are most pronounced for CC^- followed by CC, and then CC^+. These trends are apparent despite the enhanced higher-order assembly of CC^+, suggesting that a third level of assembly occurs in these peptides which is distinct from oligomer formation. This result is consistent with expectations that increased helicity detours peptides away from forming condensates by sequestering them into dilute phase oligomers. In contrast, the interactions among unfolded peptides appear to drive phase separation.

Finally, we asked if the observations reported for the peptides are recapitulated for the full length Whi3 protein. For this, we measured concentration dependent CD spectra for the full-length Whi3 protein (**Fig. S2.5**). At low concentrations, the contribution of the folded RRM is apparent in the CD spectrum. However, as concentrations increase, we observe the expected flattening and appearance of a new minimum at ~230 nm. We note that the assemblies observed in these CD experiments are precursors of condensates, since condensates themselves can become sufficiently large and fall out of solution. Our data suggest that unfolding is a pre-requisite for driving Whi3 phase separation. These data are in accord with the findings from the LASSI-based simulations, wherein folded states suppress phase separation via stronger, anisotropic interactions.

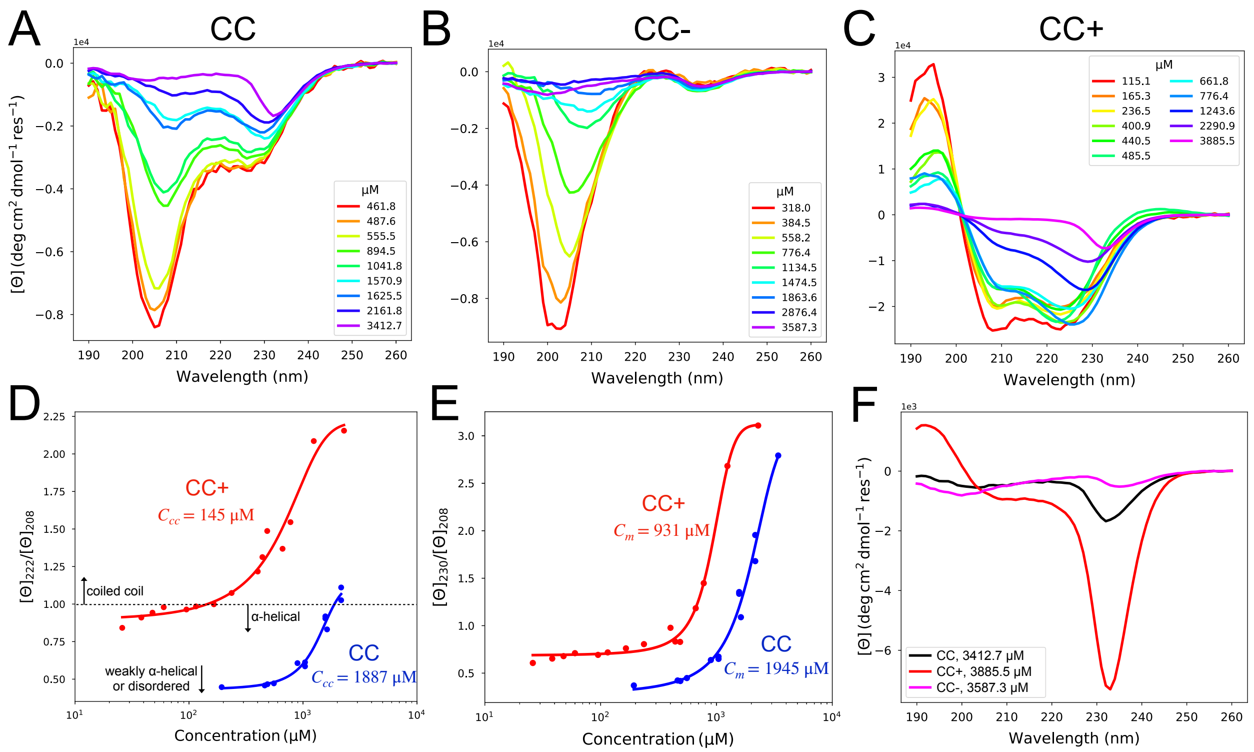


Figure 2.4: Concentration-dependent CD spectra reveal differences in conformational and assembly states across the different CC peptides. (A) CD spectra for the CC peptide. (B) CD spectra for the CC$^-$ peptide. (C) CD spectra for the CC$^+$ peptide. (D) The metric for coiled coil formation is shown as the ratio of the ellipticities at 222 nm and 208 nm from panels (A) and (C). The dotted line indicates when this ratio is one, above which coiled-coil formation is likely. (E) The metric for higher-order assembly is the ratio of ellipticities at 230 nm and 208 nm from panels (A) and (C). (F) CD spectra at the highest concentrations measured for each peptide.

Dense phase concentrations correlate with material states of condensates

In addition to changes in dense phase stoichiometry and dilute phase oligomer distributions, condensate morphologies in the Whi3-*CLN3* systems varied with time and across mutants, with each system showing distinct dynamical behaviors. An hour after mixing, the condensates are small and highly mobile. At intermediate stages of coarsening, droplets collide and partially coalesce (**Fig. 2.5A**). WT condensates have irregular shapes at the five-hour mark and later, while Whi3 (CC → CC$^-$) condensates are always more spherical and round up into spheres within four hours.

We quantified condensate shapes and sizes at different time points using a custom image analysis pipeline (see methods). The sizes of WT Whi3-*CLN3* condensates, measured in terms of their cross-sectional areas, increase as a function of time. We observe a concomitant decrease in circularity that is consistent with decreased sphericity of the condensates. Compared to WT, Whi3 (CC → CC^-) and Whi3 (ΔCC) have larger average condensate areas at early times and always have higher circularity. This is consistent with the observation that Whi3 (CC → CC^-) condensates coalesce more rapidly (**Fig. 2.5B**). The larger areas at longer time points for condensates formed by WT Whi3 and *CLN3* are partly due to the irregular shapes of these condensates, given that spheres minimize surface areas compared to other geometries.

The condensates formed by Whi3 (CC → CC^+) with *CLN3* demonstrate distinct coarsening behavior, with very small average sizes that plateau at two hours. Interestingly, their high circularity persists throughout the time periods interrogated. (**Fig. 2.5B**). This is because droplets formed by Whi3 (CC → CC^+) collide at early stages of phase separation (1-2 hrs), but instead of coalescing they either stick to one another or bounce off each other without fusing (**Fig. S2.6** in *SI Appendix*). Taken together, the data for different Whi3-*CLN3* condensates demonstrate a correlation between the dense phase stoichiometry on the one hand and morphologies and coarsening rates of condensates on the other. Specifically, higher Whi3-to-*CLN3* concentration ratios within condensates lead to more round droplets that coalesce more rapidly. We also observed that for a fixed concentration of 5 nM for *CLN3*, higher bulk concentrations of WT Whi3 lead to more Whi3 in dense phase droplets (**Figs. S2.7A, C**). This is in line with the phase behavior that is expected for a ternary mixture of RNA, protein, and solvent (**Fig. S2.8A**). Condensates with more Whi3 also round up and coarsen more quickly

(**Fig. S2.7A, B**). These data further demonstrate a correlation between dense phase stoichiometry on the one hand and morphology and coarsening rates on the other.

We performed fluorescence recovery after photobleaching (FRAP) experiments to assess the mobilities of Whi3 and *CLN3* within condensates and to uncover how internal molecular dynamics correlate with the morphological differences noted above. Two hours after mixing 1 μM Whi3 with 5 nM *CLN3*, condensates showed differential protein mobilities that were consistent with the morphological interpretations above; WT demonstrated intermediate recovery, Whi3 ($CC \rightarrow CC^-$) recovered most fully and rapidly, and Whi3 ($CC \rightarrow CC^+$) showed minimal recovery of fluorescence after photobleaching (**Fig. 2.5C**). Details of experiments, data normalization, and model fits can be found in the *SI* Appendix and **Fig. S2.9**. In contrast to the Whi3 protein constructs, the *CLN3* fluorescence did not recover to any extent (**Fig. S2.10**). This is consistent with recent observations of RNA and protein molecules featuring drastically different internal mobilities in protein-RNA condensates (50, 90). We also measured FRAP on condensates formed with WT Whi3 at different bulk concentrations mixed with 5 nM *CLN3* to test whether dense phase stoichiometry alone affects internal dynamics and material properties. Indeed, as the bulk protein concentration is lowered, dense phase protein mobility decreases, although to a lesser extent than with the CC mutant systems. This observation indicates that there are contributions of the CC region to the interactions amongst molecules within the dense phase that further modulate protein mobility (**Fig. S2.7D**). These data confirm that *CLN3* mobility is consistently lower than Whi3 mobility and that higher Whi3-*CLN3* dense phase concentration ratios promote greater protein mobility. Differences between protein mobilities across condensates formed by different Whi3 constructs are due to differences in the driving forces for phase separation that arise from mutations to the CC region. We explore these effects by modeling the impacts of dense phase stoichiometry on the dynamics of condensate aging, as discussed next.

Our data and previous work show that Whi3-*CLN3* condensates age, resulting in slowed dynamics and fibril-like structures within droplets at long times (65). To explore the dynamics of maturation, we modeled the evolution of condensates formed by the different Whi3 variant systems using the ternary Cahn-Hilliard-Cook system of equations coupled to a gelation process (91):

$$\frac{\partial \phi_i}{\partial t} = \nabla \cdot \left(M_{\phi_i}(c) \nabla \left(\frac{\partial F_{FH}}{\partial \phi_i} - \epsilon^2 \nabla^2 \phi_i \right) \right) + \sigma \cdot \zeta$$

$$\frac{dc}{dt} = M_c\big(g(\phi_p, \phi_r)c - c^2\big), \quad M_{\phi_i}(c) = \exp\big(-c/c_{\phi_i}\big)$$

Here, ϕ_i represent local protein and RNA volume fractions, c represents the local gel concentration, F_{FH} is the ternary Flory-Huggins free energy, and $g(\phi_p, \phi_r)$ defines the parabolic sol-gel contours shown in **Fig. 2.5D** (derivations in *SI Appendix*). This model provides a link between dense phase stoichiometry and aging by coupling the mobility of protein and RNA, $M_{\phi_i}(c)$, to a local gel concentration whose dynamics in turn depend on the gap between the protein and RNA volume fractions and the sol-gel line. The further the system lies beyond the sol-gel line, the faster the gel concentration grows. This lowers the mobilities of protein and RNA.

At equilibrium, condensates formed by associative polymers with sticker-and-spacer architectures are proposed to form condensate spanning networks (1, 40, 54). However, the gap between the sol-gel line and the dense phase arm of the phase boundary (92) can lead to gelation without phase separation (93, 94) and this corresponds to dynamically arrested phase separation.

If the gap between the sol-gel line and the dense phase arm of the phase boundary is large (92), then gelation without phase separation is readily realized via dynamical arrest of phase separation (55, 92). This results in a combination of irregular morphologies and ultraslow relaxation to spherical morphologies (50, 95). Thus, the material state and resulting dynamics of condensates can be tuned by the dense phase stoichiometry and the gap between the sol-gel line and the dense phase arm (92).

We can tune dense phase stoichiometries by changing the magnitude of Whi3 homotypic interactions in the Flory-Huggins free energy (**Figs. S2.8B, C** in *SI Appendix*). The Flory-Huggins free energy represents averaged, non-specific interactions among components, so the Whi3 homotypic interaction in this context encompasses contributions from both the CC and QRR. The strength of the homotypic interactions that drives phase separation is implicitly linked to the degree of unfoldeness of the CC motif and the QRR in Whi3. Accordingly, in the FH model, when the Whi3 homotypic interaction strength is low, the dense phase arm is furthest beyond the sol-gel line, and droplet dynamics are readily arrested, as in the Whi3 ($CC \rightarrow CC^{+}$) system (**Figs. 2.5D, E**). Alternatively, when the Whi3 homotypic interactions strength is high, the gap between the dense phase concentration and sol-gel line is narrowed, and the dynamics are minimally affected (**Figs. 2.5D, E**). We observed similar effects when altering only Whi3 bulk concentration in the model and fixing interaction energies. Here, high Whi3 bulk concentrations lead to high dense phase Whi3 concentrations and coarsening over longer times (**Figs. S2.7E, F**, *SI Appendix*). These results are a natural consequence of a ternary phase boundary intersecting the sol-gel line. They represent the influence of stoichiometry on gelation.

Since the model does not include Brownian motion of droplets, coalescence does not contribute to coarsening. Instead, coarsening is driven by Ostwald ripening (96), which is hindered as gelation arrests phase separation. Thus, droplet circularity is not affected by gelation in this model, but droplet sizes are affected, whereby smaller sizes indicate increased gelation that arrests phase separation. We confirmed this by performing simulations at different homotypic interaction strengths and bulk concentrations without the gelation term, an approximation that corresponds to the scenario of phase separation without gelation. In this scenario, we do not find significant differences in average droplet sizes (**Fig. S2.11**, *SI Appendix*). Overall, the results from the modeling are consistent with the experimental observations that a higher dense phase Whi3 to *CLN3* concentration ratio correlates with increased mobility of proteins within condensates. The modeling also highlights the importance of dense phase stoichiometries in impacting the sequence-specific gaps between sol-gel lines and dense phase arms of phase boundaries (92).

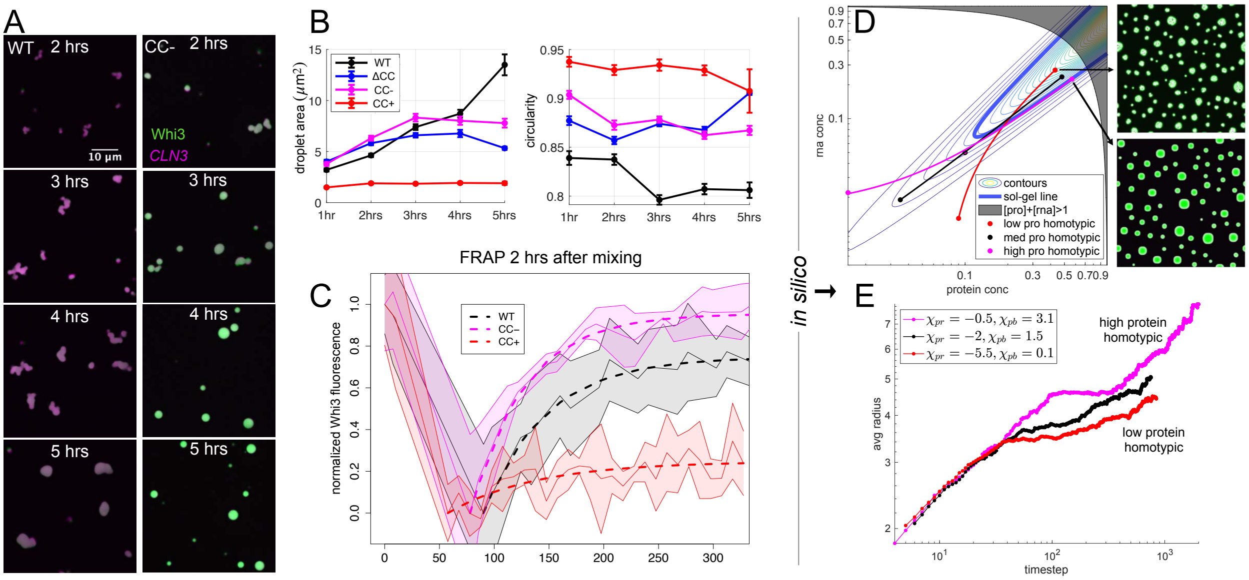


Figure 2.5: Droplet dynamics are linked to dense phase stoichiometries. (A) Representative images at each time point for WT and Whi3 (CC → CC$^-$) proteins at bulk concentrations of 500 nM Whi3 with 5 nM *CLN3*. (B) Quantification of average droplet area and circularity for experiments corresponding to panel (A). Error bars are standard error of the mean. (C) FRAP curves for Whi3-*CLN3* mixtures performed 2 hours after mixing. FRAP curves for 5-8 droplets for each protein were recorded; dashed lines indicate model fits to the mean, and the shaded areas indicate +- 1 standard deviation. (D) Parabolic sol-gel contours with the sol-gel line in bold blue are overlaid by dilute, bulk, and dense phase equilibrium concentrations of simulated systems with different protein homotypic interaction energies. For fixed bulk concentrations of protein and RNA, alteration of the protein homotypic interaction energies results in dense phase concentrations that lie at varying depths beyond the sol-gel line. Simulation snapshots after steady state has been reached in the low protein homotypic and high protein homotypic interaction energy systems are shown. (E) The average radii of droplets from simulations in panel (D) are shown as a function of time. The legend provides values for the interaction parameters in the Flory-Huggins free energy that determine whether the systems encode low, medium, or high protein homotypic interaction energies.

Discussion

We have identified a 21-residue motif, designated as CC, that lies within the Whi3 QRR. In WT Whi3, the CC motif is likely defined by an equilibrium between helical and non-helical conformations. Mutations that alter the helicity of this motif can be used to tilt the equilibrium toward conformations that are more helical or less helical. Altering the helical propensity of CC through mutations results in trade-offs of order and disorder at the molecular level. Interestingly, these differences have profound effects on the phase behavior of the Whi3-*CLN3* system. Interactions among Whi3 molecules with stable helical CC motifs lead to a sequestration of proteins into dilute phase oligomers. This leads to a lowering of protein concentrations in the dense phases of RNP condensates. It also contributes to a faster aging of RNP condensates because the gap between the sol-gel line and dense phase arm of the phase boundary is widened. In contrast, increased disorder within the CC motif and disorder within the remainder of the QRR weakens dilute phase oligomerization and drives Whi3 molecules into RNP condensates. These condensates experience slower aging due to a minimized gap between the sol-gel line and the dense phase arm of the phase boundary.

Interestingly, we find that the Whi3 (ΔCC) construct behaved most like the Whi3 ($CC \rightarrow CC^{-}$) construct. Specifically, we observe higher protein concentrations in the dense phase for Whi3 (ΔCC) when compared to WT at similar bulk concentrations of each protein and hence lower bulk volume fractions of Whi3 (ΔCC) compared to WT. Clearly, the effect of removing the CC is strong enough to compensate for a small length difference that should weaken phase separation of Whi3 (ΔCC). Deletion of the CC motif would imply that the QRR seldom samples helical conformations, and phase separation is driven via interactions among disordered regions. For WT, we hypothesize that helicity from the CC is sufficient to tip the balance toward interactions among helical domains that sequester Whi3 proteins into dilute phase oligomers.

Interactions among Whi3 and *CLN3* drive phase separation, primarily through heterotypic interactions that involve a disordered QRR and a folded RRM. Homotypic interactions mediated by oligomerization via short helical domains detour the molecules away from co-phase separation with RNA. The LASSI simulations suggest that the valence of stickers within the QRR contribute to homotypic interactions, which are potentially weakened by competing heterotypic interactions with *CLN3* in the ternary system as indicated by lower dense phase concentrations in the ternary system (**Fig. 2.2B**) relative to those in the binary system (**Fig. 2.2D**). The question is why do we observe clusters that span a range of sizes in the dilute phase

(**Figs. 2.3C, D**)? The clusters are likely to involve a combination of stoichiometric assemblies mediated mainly by interactions among helical motifs, and heterogeneous distributions of clusters that are mediated by the CC motif and the QRR. These heterogeneous distributions are manifestations of PSCP, whereby pre-percolation clusters of different sizes are expected to be present in the sol that coexists with the dense phase (53, 55, 97).

The sequence motifs we have identified within the Whi3 QRR are a common feature of proteins with QRRs. Indeed, other groups have shown that Leu, Pro, and His are especially enriched within and around QRRs, with Leu being typically clustered at N-terminal regions and Pro being prevalent at C-terminal regions (98, 99). Here, we observe similar positioning of these residues within the Whi3 QRR (**Fig. 2.1A**). In accord with our findings, Kokona and Fairman have shown that formation of stable coiled coils suppresses polyQ driven insolubility via aggregation and the formation of solid-like inclusions by peptides that mimic the exon 1 encoded region of huntingtin (100). Interestingly, Ford and Fioriti have demonstrated that QRRs, coiled-coils, and RNA-binding domains commonly co-occur, especially in neuronal proteins that are implicated in both phase separation and neurogenerative diseases caused by toxic, amyloid assemblies (101).

Relation to other systems with oligomerization domains and IDRs

Many proteins that have been identified as drivers of phase separation contain both structured oligomerization domains and IDRs. Reports to date have shown that oligomerization promotes phase separation (21, 31, 32, 34, 35, 39, 44, 74, 102). A notable exception to these findings is the prediction of a “magic-number effect” that occurs for rigid domains with sufficiently strong, specific interactions, whereby molecules become trapped in stoichiometric assemblies (103). Similar effects are realizable, irrespective of the strength of inter-domain interactions, if the linkers between the interacting domains are too short (54). In the model that shows the magic number effect, binding partners with an integer multiple of one-to-one binding sites can saturate interactions with each other, favoring the formation of network terminating oligomers in the dilute phase. The experiments we summarize here show the phenomenon of dilute phase oligomers sequestering the Whi3 protein out of the dense phase, although this behavior does not appear to be a magic-number effect. It is possible that the CC motif may prefer a specific oligomerization state, which on its own could result in stable oligomers with saturated binding sites. However, Whi3 also mediates homotypic interactions through the rest of its QRR that presumably has no fixed valence. It also participates in heterotypic interactions with RNA via its RRM. These interactions contribute complexity to the system that makes it difficult to interpret in the context of a magic number effect, although the phenomenological similarity between the theory of Xu et al. (103) and the effects of increasing the CC stability (rigidity) and interaction energy are intriguing.

The presence of structural motifs embedded within IDRs may provide an explanation for the effects of disease-related point mutations, as has been shown for TDP-43 (74). If such mutations perturb structural motifs and their associations, then drastic shifts in phase behavior and material states are possible even with point mutations (104). Importantly, the effects of the CC motif that we have uncovered in this study are distinct from those in TDP-43, highlighting the need for context-dependent studies of embedded structured domains. Identification of such motifs within phase separating systems will be an essential task for improved understanding of the driving forces that contribute to the formation of biomolecular condensates.

Materials and Methods

UV-Circular Dichroism

Peptides were purchased in pure form from GenScript with N-terminal acetylation and C-terminal amidation. All peptide sequences included an N-terminal tryptophan to enable concentration measurements, followed by a glycine, and then the sequence of interest. CD measurements were carried out using a Jasco 810 spectropolarimeter scanning from 260 nm to 190 nm, with a data pitch of 1 nm and a bandwidth of 1 nm. Four to six accumulations were averaged for each spectrum with a scanning speed of 50 nm/min and a two second response time. Additional details regarding the CD measurements are described in the *SI Appendix.*

Assembly and imaging of Whi3-*CLN3* condensates *in vitro*

Labeled and unlabeled stocks of Whi3 protein (WT, CC+, CC-, ΔCC) and labeled *CLN3* mRNA were stored in 200 µL, 33 µL, and 10 µL aliquots respectively at -80˚C until just before experiments. Aliquots were thawed on ice and protein samples were spun down at 13,200 rpm for 5 minutes to remove any aggregates. The concentration of labeled and unlabeled protein was then determined by measuring absorbance at 280 nm. The labeled protein concentration was corrected using a factor of 0.09 to account for the Atto 488 absorbance, and labeling efficiency was determined as the ratio of the dye concentration to the corrected protein concentration. Labeled and unlabeled Whi3 proteins and *CLN3* mRNA were diluted into buffer containing 50 mM HEPES at pH 7.4 and 150 mM KCl to desired concentrations and labeled fractions of protein and mRNA. The volumes of labeled and unlabeled protein stock needed to simultaneously obtain a desired total protein concentration and labeled fraction were determined using the following formulas: $u = \frac{VC}{[u]+\lambda[l]}$, $l = \lambda u$, $\lambda = \frac{L[u]}{(l_L - L)[l]}$, where u is the desired volume of unlabeled protein (µL), l is the desired volume of labeled protein (µL), V is the final reaction volume (200 µL for phase separation assays, 50 μL for FRAP assays), C is the desired final protein concentration (μM), $[u]$ is the measured concentration of the unlabeled protein stock (μM), $[l]$ is the measured concentration of the labeled protein stock (μM), l_L is the measured labeled fraction of the labeled protein stock, and L is the desired protein labeled fraction (5% for phase separation assays, 10% for WT and CC^- and 20% for CC^+ in FRAP assays). Whi3 proteins were prepared at three concentrations, 250 nM, 500 nM, and 1 μM. *CLN3* mRNA was prepared at a concentration of 5nM for every condition. For phase separation assays, 200 μL of protein and mRNA solutions were loaded into glass-bottom imaging chambers (Grace Bio-Labs) that were pre-coated with 30 mg/mL BSA (Sigma) for 30 mins to prevent protein adsorption to the surfaces of the well. The surfaces were washed thoroughly with buffer prior to addition of protein and mRNA solutions. Labeled and unlabeled Whi3 stocks were added first to imaging wells, followed by *CLN3*, and the final solution was mixed by pipetting without introducing bubbles. Imaging chambers were either placed in a 25C incubator until imaging or immediately transitioned to a temperature-controlled microscope stage (Tokai Hit from Incubation System for Microscopes) set to 25C. For each desired timepoint, confocal z-stacks of imaging wells were acquired. Three separate z-stacks within each well were taken at each time point at random locations. Fluorescence images in two channels were acquired using 488 nm and 561 nm lasers to visualize Atto488-labeled Whi3 and Cy3-labeled *CLN3* respectively. An Atto 488 dye calibration curve was created by imaging 200 μL wells with varying concentrations of dye. Experiments were replicated three times.

LASSI simulations

Simulations were performed using LASSI and run on the Longleaf computing cluster at UNC-Chapel Hill. Each simulation was run independently on a single compute node with 4GB

RAM. Interaction energies that were used for each of the three systems studied (WT, CC+, CC-), are shown in **Table S2.2** *SI Appendix*.

Acknowledgments

This work was supported by grants from the US National Institutes of Health (5R01NS056114 and R01NS089932 to RVP, R01BM081506 to ASG, F32-GM133123-01A1 to BMS, F32-GM136055 to WTS, and T32-GM8570-25 to IS), the Air Force Office of Scientific Research (FA9550-20-1-0241 to ASG and RVP), the HHMI faculty scholars (to ASG). We are grateful to Furqan Dar and Kiersten Ruff for helpful discussions. Furqan Dar provided guidance and assistance with LASSI simulations.

SI Appendix

UV-CD measurements

A demountable quartz cuvette with a path length of 0.1 mm was used for all measurements. The buffer used in all measurements was 20 mM HEPES, pH 7.4 with 150 mM NaF, filtered with a 0.2 μm PES filter. Fluoride was used in place of chloride as the anion to avoid the signal interference caused by chloride. For concentration-dependent CD measurements, a high concentration stock (3-4 mM peptide) was prepared by dissolving lyophilized peptide in a small volume of buffer. Subsequent samples were obtained by dilution of the stock solution using the same buffer. For CD measurements acquired as a function of temperature, the temperature was varied from 5C to 95C or from 95C to 5C at a rate of 5C per minute, with a one minute pause before each full spectrum measurement made at 10C intervals. Single wavelength (either 208 or 222 nm) measurements were made at 5C intervals. Peptide concentrations were prepared in the 600-900 μM range. CD spectra were converted from machine units (mdeg) to mean residue ellipticity ([θ], deg cm^2 $dmol^{-1}$ $residue^{-1}$) using the equation $[\theta] = \frac{\theta}{(N-1)LM}$, where θ is machine units, N is the number of amino acids in the peptide, L is path length and M is molar concentration. Peptide concentrations were determined by measuring absorbance spectra using an Implen NP80 nanophotometer. Molar concentration values were calculated using an extinction coefficient of 5500 M^{-1} cm^{-1} at 280 nm. In **Fig S2.1**, mean residue ellipticity at 222nm was fit for each peptide with the following 2-state model:

$$[\theta]_{222nm} = \frac{A_1 - A_2}{1 + e^{(m(T_m - T))}} + A_2$$

Deconvolution of CD data

CD Data sets were converted to units of mean residue ellipticity as described above and submitted to the Dichroweb server (84) (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) for deconvolution using the CDSSTR algorithm (105) and basis sets 4, 7 (106), and SP175 (107). These are the basis sets that have been optimized for the wavelength rage of 190-240 nm. We explored algorithms other than CDSSTR, including CONTIN and SELCON3 (108), but CDSSTR consistently provided the best fits for our data, so the other algorithms were not used further. We note that the output of the analysis is broken down into two helical categories and two beta strand categories (109), but for the sake of simplicity we present the results in terms of a single helical estimate and a single beta strand estimate, each of which is the sum of the respective sub-categories. Reported secondary structure content estimates represent the mean and standard deviation of the values resulting from analysis using each of the three basis sets (4, 7, SP175).

Bacterial expression and purification

Whi3 constructs were transformed into BL21 competent cells (New England Biolab C2527I). Cells were grown at 37C with shaking and appropriate selection until an OD600 of 0.6 was reached. Then cells were shifted to 18C and induced with 1mM IPTG (Fisher I2481C25).

Cells were grown for 18hr and then pelleted by centrifugation at 13000rcf. The Pellet was resuspended in 25ml Lysis buffer (1.5M KCl, 20mM Imidazole, 5mM beta-mercaptoethanol, 50mM Hepes pH7.4) along with 1 tablet Pierce Protease inhibitor (Thermo-Fisher Scientific A32965) and 15mg Lysozyme (Fisher BP535-10). Cells were lysed at 4C with gentle rocking for 1hr and then briefly sonicated on ice. Lysate was clarified by centrifugation at 50000rcf for 30min. Clarified lysate was mixed with 0.8mL of Cobalt resin (Thermo Fisher Scientific PI89965), column was washed with 5 column volumes of lysis buffer and then eluted with 3ml of elution buffer (150mM KCl, 200mM Imidazole, 5mM beta-mercaptoethanol, 50mM Hepes pH7.4). Protein was dialyzed into storage buffer (150mM KCl, 50mM Hepes pH 7.4) using a slide-a-lyzer cassette with a 20,000MW cutoff (Thermo-Fisher Scientific 87735). Protein was labelled using an atto-488-NHS ester (Sigma Aldrich 41698) and then dialyzed again to remove uncongugated dye.

RNA *in vitro* transcription

Plasmid was linearized with XbaI and NotI restriction enzyme (New England Biolabs R0145S, R0189S) and gel purified (QIAGEN 28706). 100ng of gel purified DNA was used as a template for in vitro transcription (NEB E2040S) carried out according to the manufacturer’s instructions with the addition of Cy3 (Sigma PA53026) labeled UTP to each reaction. Following incubation at 37C for 18 hours, in vitro transcription reactions were treated with DNAseI (NEB M0303L) according to the manufacturer’s instructions. Following DNAse treatment, reactions were purified with 2.5M LiCL precipitation. Purified RNA concentrations were quantified using nanodrop and verified for purity and size using a denaturing agarose gel and Millenium RNA ladder (ThermoFisher Scientific AM7151).

Assembly and imaging of Whi3-only condensates *in vitro*

Whi3 proteins were concentrated using 30 kDa molecular weight cutoff centrifugal filters (Amicon) and diluted into 50 mM HEPES pH 7.4 and 2 mM TCEP buffer at the specified protein concentrations such that the final KCl concentration was 75 mM. Atto488-labeled protein was included such that approximately 4-6% of protein was dye-labeled. Protein solutions were loaded into glass-bottom imaging chambers (Grace Bio-Labs). Wells were pre-coated with 30 mg/mL BSA (Sigma) to prevent protein adsorption to the well surfaces, and washed thoroughly with experiment buffer prior to addition of protein solutions. Wells were incubated for 2.5 h at room temperature to allow droplets to form and sediment onto the surface of the imaging chambers. Confocal z-stacks of droplets were acquired, starting below the coverslip surface, with a z-spacing of 0.2 μm. Imaging was repeated 4 h post-mixing. At least five z-stacks of each condition were acquired, and all experiments were replicated three times per condition.

Fluorescence recovery after photobleaching (FRAP)

Silicone gaskets were laid onto plasma-cleaned coverglass and served as the reaction compartments for imaging of Whi3-CLN3 droplets. Reaction wells were passivated with BSA for 15 mins and then washed thoroughly with buffer (150 mM KCl, 50 mM HEPES pH 7.4, 5 mM BME) before adding buffer, Whi3, and CLN3. 50 μL volumes were used in each well, and components were assembled as described in the section: Assembly and imaging of Whi3-CLN3 droplets in vitro in the Methods section of the main text. 1 μM Whi3 and 5 nM CLN3 were used for all experiments. The coverglass was placed in a petri dish to prevent evaporation and incubated at 25C for 2 hours before imaging. During imaging, 4-8 droplets were photobleached for each experiment, and a similar number of similarly-sized droplets in each field of view were left unbleached for use in later correction of bleaching due to imaging. The data were processed in FIJI using the Time Series Analyzer which allows for selection of ROIs corresponding to both bleached and unbleached droplets. The following analysis was carried out for each of the Whi3 and CLN3 channels. The fluorescence signals for bleached and unbleached droplets were averaged separately to obtain the following average fluorescence intensities as a function of time: $\hat{F}(t)$, the averaged bleached droplet fluorescence, and $\hat{B}(t)$, the average unbleached droplet fluorescence. A corrected signal was calculated by dividing the bleached droplet signal at each timepoint by the unbleached droplet signal: $\hat{F}_{corrected}(t) = \frac{\hat{F}(t)}{\hat{B}(t)}$. The fluorescence before bleaching is defined as $F(0) = \hat{F}_{corrected}(t = 0)$, and the fluorescence just after bleaching is defined as $F_0 = \hat{F}_{corrected}(t = t_{bleach})$. We normalized $\hat{F}_{corrected}(t)$ by $F(0)$ and F_0 using the following formula: $N(t) = \frac{\hat{F}_{corrected}(t) - F_0}{F(0) - F_0}$. We fit $N(t)$ to the following formula: $N(t) = A\left(1 - e^{-t/\tau}\right)$. We calculated the standard deviation of the original fluorescence curves, $sd\big(F(t)\big)$, and propagated the error through the normalization process. This amounts to dividing the standard deviation at each time by the factor $\hat{B}(t)(F(0) - F_0)$. $N(t)$ plus and minus the propagated standard deviation and with the model fit overlaid is shown in the main text in **Fig. 2.5C**, and the raw data with normalizations and model fit parameters are shown in **Fig. S2.9**.

Calculating protein concentration from fluorescence calibration curves

We used dye solution standards to determine protein concentrations. Briefly, dye solution standards were created by diluting Atto488 dye to concentrations ranging from 0-10 or 0-20 µM in the same buffer used for droplet assembly. Solutions were loaded into imaging wells and at least three confocal z-stacks of dye standards, starting below the coverslip surface, were acquired using the same laser power and exposure settings used for droplet imaging. For Whi3-only phase separation assays (**Figs. 2.2C, D**, two different laser settings were used for WT and CC mutant droplets, hence the two plots in **Fig. S2.2B**. Dye intensity was determined by computing the mean intensity of the dye image frame with the highest intensity value. For experiments including both Whi3 and CLN3, all proteins were labeled at 5%. For Whi3-only phase separation assays, estimated labeled populations of the constructs were WT: 5.95%, CC-: 4.15%, and CC+: 4.28%. Fluorescence values in condensates are converted to dye concentrations using the fits in **Fig. S2.2** and divided by labeled population for each construct to estimate protein concentrations.

Analysis of *in vitro* condensates

Images were analyzed using a custom FIJI macro and the output data was further analyzed in Matlab to perform statistics and plotting. For each image, the FIJI macro loops through each z-slice in the mRNA channel for Whi3-CLN3 experiments, and the single Whi3 channel for Whi3-only experiments. The mRNA channel is used for Whi3-CLN3 experiments because the mRNA intensity is constant across experimental conditions, while the protein fluorescence intensity varies widely. For each slice, the image is duplicated, and on the duplicate a Yen threshold (110) on the fluorescence intensity is performed, and a watershed operation (111) is run on the resulting binary image. Using masks from the binary image, the "Analyze Particles" operation within FIJI is run on the original image and records the area, mean fluorescence intensity, and circularity for objects larger than 50 pixels in area. The area of all detected objects is summed and recorded. The slice with the largest total area for each image is chosen as the optimal slice, and statistics in both CLN3 and Whi3 channels are recorded and used for further analysis. For Whi3-only experiments, only the single channel is used.

In Matlab, Whi3 fluorescence intensities are converted to concentrations using the calibration curves. Data from three imaging frames per replicate per experimental condition at each time point were aggregated across three replicates. Data are plotted either as boxplots at single time points, or as line plots with means plus and minus the standard error of the mean over time.

Determination of proteins per puncta on passivated surfaces

Glass coverslips were passivated with polyethylene glycol (PEG) by coating the glass with a layer of poly-L-lysine (PLL) conjugated to PEG. PLL-PEG was made according to a previous protocol (112). Briefly, amine-reactive PEG-SVA (succinimidyl valerate) was added to a 40 mg/mL mixture of PLL in 50 mM sodium tetraborate pH 8.5 at a molar ratio of one PEG per five lysine subunits. The mixture was stirred continuously for 6 h at room temperature and buffer exchanged into 50 mM HEPES pH 7.4 and 150 mM KCl using Centri-Spin size exclusion columns (Princeton Separations). Imaging wells were made by placing silicone gaskets onto oxygen plasma-treated coverslips. Wells were coated for 20-30 min with PLL-PEG diluted tenfold in 50 mM HEPES pH 7.4 and 75 mM KCl buffer. After coating, the well was washed repeatedly with HEPES buffer to remove excess PLL-PEG. Whi3 proteins were then diluted in 50 mM HEPES pH 7.4 and 75 mM KCl buffer and added to PEG-coated coverslips at concentrations of 100 or 150 nM. Proteins were allowed to settle on the surface for approximately 10-20 min, and images of puncta were acquired using TIRF microscopy.

Images of puncta were cropped to the center 400x400 pixels (center 1/9) of the original images to ensure even illumination across the field of view. Diffraction-limited protein puncta were detected using cmeAnalysis particle detection software (113), which fit puncta to a two-dimensional Gaussian function with standard deviation, σ, determined from the TIRF microscope point-spread function (PSF).

Briefly, the PSF σ was determined by acquiring images of 100 nm TetraSpeck fluorescent microspheres (Invitrogen) on a coverslip surface using the same TIRF settings used in experiments. The fluorescence intensity profiles of diffraction-limited microspheres were fitted to the two-dimensional Gaussian function using Wolfram Mathematica 12 software. The PSF σ was taken as the average from seven fits.

The cmeAnalysis software reports the amplitude, "A," of the Gaussian fit over the local background intensity, "c." The A values were accepted as valid if they came from diffraction-limited puncta and were significantly above the local c values. The number of proteins per puncta was computed from A values by comparing to the intensity of single protein molecules, described in the next section.

Determining the intensity of single protein molecules

Silicone gaskets were placed on oxygen plasma-treated coverslips and Atto488-labeled Whi3 was added at a concentration of 50-100 pM. Images of single protein molecules adhered to the coverslip surface were acquired with TIRF microscopy using the same TIRF and laser power settings used in experiments. Images were cropped to the center 400x400 pixels of the original images and the diffraction-limited puncta of single proteins were detected using cmeAnalysis software, following a similar procedure described in the previous section. "A" values were pooled from 5-10 image frames and binned into a histogram, yielding a distribution with one clear peak corresponding to the average intensity of a single protein molecule. This value was then used to calculate the number of dye-labeled proteins per puncta. Finally, the total number of proteins per puncta was determined by adjusting for the known fraction of dye-labeled to unlabeled protein in the experiment. Single-step photobleaching experiments were performed to validate our approach for determining single protein molecule intensity. Specifically, we prepared samples of dilute, glass surface-adhered proteins as described above and acquired time-lapse movies of protein bleaching using TIRF microscopy. The intensities of puncta over time, after background subtraction, were analyzed using ImageJ software. This approach revealed that proteins frequently bleached to the level of the local background in a single, clear step. We found that the initial, average intensities of such puncta were in agreement with particle detection-based estimates of single molecule intensity.

Additional details regarding the LASSI simulations

We used the open-source version that we have made available on GitHub (https://github.com/Pappulab/LASSI). We added analysis routines to enable the calculation of specific quantities that are unique to the current work.

Each chain consists of a single CC bead connected to four QRR beads, with a correspondence of approximately 20 amino acids per bead. 2000 chains were used in each simulation. For each system, 35 volume fractions were run, logarithmically spaced between volume fractions of 0.0001 to 0.79. For each volume fraction, 35 temperatures were run, linearly spaced between 1 and 2 (arbitrary units). Temperature scales interaction energies as $\frac{\varepsilon}{T}$. For each temperature and volume fraction, each simulation ran for 1e9 steps, after 5e6 steps of thermalization during which the temperature is set to 1000 and polymers are forced to the center of the simulation box. Two replicates are run for each simulation. As described in (40), the last half of trajectories for each simulation across the two runs are analyzed and averaged to produce a global density inhomogeneity metric, $\bar{\rho}$. Above $\bar{\rho} = 0.025$, the system is phase separated, as detailed in (40). Binodals in this work are drawn along contours corresponding to $\bar{\rho} = 0.025$.

Dilute phase oligomer distributions were calculated with custom scripts that used the ovito python module (114). The final frame of each trajectory corresponding to a single run for each simulation condition was used for analysis. A cluster analysis was performed on the particle positions, where clusters within 2 units of each other are considered to be in the same cluster (this allows for all possible adjacent lattice positions with a 3D lattice). If the system was phase separated ($\bar{\rho} > 0.025$), the largest cluster, which corresponds to the dense phase, is removed from further analysis. The proportion of chains that are monomeric in the resulting dilute phase is calculated as the total number of chains not in a cluster divided by the total number chains in the dilute phase. Plots were created using matplotlib in python.

Computing phase diagrams from the Flory-Huggins free energy

To begin to understand the effects of homotypic interactions on phase behavior, we studied the Flory-Huggins free energy, a commonly used mean-field model of phase separation in polymer solutions (115). Our system consists of two polymers, protein and RNA, and the buffer, so it is characterized by the ternary form of the free energy:

$$F_{FH}(\phi_p, \phi_r) = \frac{\phi_p}{N_p}\ln\phi_p + \frac{\phi_r}{N_r}\ln\phi_r + (1 - \phi_p - \phi_r)\ln(1 - \phi_p - \phi_r) + \chi_{pr}\phi_p\phi_r$$

$$+ \chi_{pb}\phi_p(1 - \phi_p - \phi_r) + \chi_{rb}\phi_r(1 - \phi_p - \phi_r)$$

where ϕ_p, ϕ_r are the concentrations of protein and RNA, respectively, and the third component, buffer, is expressed as $\phi_b = 1 - \phi_p - \phi_r$ due to the conservation of mass. Each interaction term is defined as: $\chi_{ij} \sim \varepsilon_{ij} - \frac{\varepsilon_{ii}+\varepsilon_{jj}}{2}$, where ε_{ij} denotes the interaction energy between species i and j, and is negative for attractive forces by convention. We compute phase diagrams for the system by finding dense and dilute phase concentrations that result in equivalent chemical potentials. For a given set of parameters, the system will phase separate if the chemical potentials of the dense and dilute phases are equal for each component, i.e.:

$$\left.\frac{\partial F_{FH}}{\partial \phi_p}\right|_{\phi_p=\phi_p^L} = \left.\frac{\partial F_{FH}}{\partial \phi_p}\right|_{\phi_p=\phi_p^H}$$

$$\left.\frac{\partial F_{FH}}{\partial \phi_r}\right|_{\phi_r=\phi_r^L} = \left.\frac{\partial F_{FH}}{\partial \phi_r}\right|_{\phi_r=\phi_r^H}$$

$$\left.\frac{\partial F_{FH}}{\partial \phi_b}\right|_{\phi_b=\phi_b^L} = \left.\frac{\partial F_{FH}}{\partial \phi_b}\right|_{\phi_b=\phi_b^H}$$

Using the conservation of mass relation from above, this system simplifies to a system of three equations with the unknowns $\phi_p^L, \phi_p^H, \phi_r^L, \phi_r^H$. For a fixed value of ϕ_p^L, for example, we can solve the system by minimizing the following residual function (116):

$$R = \sum_i \left(\left.\frac{\partial F_{FH}}{\partial \phi_i}\right|_{\phi_i=\phi_i^L} - \left.\frac{\partial F_{FH}}{\partial \phi_i}\right|_{\phi_i=\phi_i^H} \right)^2$$

For a given set of parameters, we solve for phase diagrams using a custom global minimization routine on the residual function defined above in Matlab.

In ternary systems, two species will phase separate together out of the third if their heterotypic interactions dominate over each of their homotypic interactions. We are interested in protein and rna condensing together, so the above requirement is on their interaction term, χ_{pr}. The constraint that heterotypic interactions dominate requires that $|\varepsilon_{pr}| > \left|\frac{\varepsilon_{pp}+\varepsilon_{rr}}{2}\right|$ which implies that $\chi_{pr} < 0$. We indeed find that this is a necessary condition to produce closed loop phase diagrams. We investigated the effects of altering protein and rna homotypic interactions by fixing all other parameters and setting χ_{pb} and χ_{rb} accordingly (**Fig. S2.5**). When homotypic interactions are equal, the phase diagram is symmetric about the line at which protein and RNA concentrations are equal (**Fig. S2.5A**). As either protein or rna homotypic interactions are increased, the phase diagram lies more closely along that axis and its tie lines become more aligned with that axis (**Figs. S2.5B, C**). It is evident that one can increase the dense phase concentration of a component in a ternary system by either increasing the bulk concentration of that component, or by increasing the homotypic interaction energy of that component.

Deriving the ternary Cahn-Hilliard-Cook equation coupled to gelation

Since our system comprises protein, RNA, and buffer, a model well-suited to study the dynamics of phase separation is the Cahn-Hilliard-Cook system of equations:

$$\frac{\partial \phi_i}{\partial t} = \nabla \cdot \left(M(\phi) \nabla \frac{\partial F}{\partial \phi_i} \right) + \sigma \cdot \zeta$$

where $\phi_i = \phi_p, \phi_r$, $M(\phi)$ is a concentration dependent mobility, ζ is a mean-zero Gaussian noise with variance σ, and $\frac{\partial F}{\partial \phi_i}$ is the functional derivative of the following free energy functional:

$$F(\boldsymbol{\phi}, \nabla \boldsymbol{\phi}) = \int_{\Omega} \left(F_{FH}(\boldsymbol{\phi}) + \nabla \boldsymbol{\phi}^T \Gamma \boldsymbol{\phi} \right) d\Omega$$

where F_{FH} is the ternary Flory-Huggins free energy. We assume that Γ is a diagonal matrix of constants, ε^2, for simplicity. Since we are only interested in 2-phase behavior in which protein and RNA condense together, their cross gradient terms should have either neutral or negative energies, i.e. there should be no interfaces formed between RNA and protein. To simplify the model, we assume that this energy is 0 such that all interactions are encoded in the χ_{ij}. Taking the functional derivative of F, we arrive at the following form:

$$\frac{\partial \phi_i}{\partial t} = \nabla \cdot \left(M(\phi_i) \nabla \left(\frac{\partial F_{FH}}{\partial \phi_i} - \varepsilon^2 \nabla^2 \phi_i \right) \right) + \sigma \cdot \zeta$$

We introduce a gelation process to the above system of equations as described in the text. To define the log-parabolic sol-gel contours, $g(\phi_p, \phi_r)$, we have the following:

$$g(\phi_p, \phi_r) = \frac{r'\left(\frac{p'}{p_0}\right)^{-w\log\left(p'/p_0\right)} - r_0}{\max\limits_{p+r=1} g}$$

$$p' = p_0\left(\frac{p}{p_0}\right)^{\cos(\theta)} \Big/ \left(\frac{r}{r_0}\right)^{-\sin(\theta)}$$

$$r' = r_0\left(\frac{p'}{p_0}\right)^{\sin(\theta)} \Big/ \left(\frac{r}{r_0}\right)^{\cos(\theta)}$$

where θ defines the rotation angle in log-log space of the function.

Numerical solution of the Cahn-Hilliard-Cook equation coupled to gelation

The introduction of the gelation coupling to the Cahn-Hilliard equation introduces a mobility term which varies in time and space, thus making the entire equation non-linear. For this reason, we use the scheme first developed in (117). In the Fourier domain, our system of equations becomes:

$$\frac{\partial \tilde{\phi}_i}{\partial t} = ik \cdot \left(M(c)\left[ik'\left(\frac{\partial \tilde{F}_{FH}}{\partial \phi_i} + \varepsilon^2 k^2 \tilde{\phi}_i \right) \right]_r \right)_k$$

where the subscript $i = p, r$ denotes the protein and rna concentrations, respectively, F_{FH} is the ternary Flory-Huggins free energy, the tilde and the *k* subscript denote the Fourier transform, and the *r* subscript denotes the inverse Fourier transform. We stabilize the equation by adding the following linear fourth-order term to both sides: $A\varepsilon^2 k^4 \tilde{\phi}_i$, where $A = \frac{1}{2}\max\left(\chi_{ij}\right)$. Discretizing in time, we now have:

$$\frac{\tilde{\phi}_i^{t+1} - \tilde{\phi}_i^t}{\Delta t} + A\varepsilon^2 k^4 \tilde{\phi}_i^{t+1} = A\varepsilon^2 k^4 \tilde{\phi}_i^t + ik \cdot \left(M(c^t)\left[ik'\left(\frac{\partial \tilde{F}_{FH}^t}{\partial \phi_i} + \varepsilon^2 k^2 \tilde{\phi}_i^t \right) \right]_r \right)_k$$

Solving for $\tilde{\phi}_i^{t+1}$,

$$\tilde{\phi}_i^{t+1} = \frac{(1 + A\Delta t\varepsilon^2 k^4)\tilde{\phi}_i^t + \Delta t i k \cdot \left(M(c^t) \left[i k' \left(\frac{\partial \tilde{F}_{FH}^t}{\partial \phi_i} + \varepsilon^2 k^2 \tilde{\phi}_i^t \right) \right]_r \right)_k}{1 + A\Delta t\varepsilon^2 k^4}$$

we finally obtain our desired solution by taking the inverse Fourier transform and adding the noise term:

$$\phi_i^{t+1} = \left(\tilde{\phi}_i^{t+1}\right)_r + \sigma \cdot \zeta$$

Numerical implementation of the various spatial derivatives was guided by (118). We used an implicit time scheme to discretize the gelation equation:

$$\frac{c^{t+1} - c^t}{\Delta t} = M_c g\left(\phi_p^{t+1}, \phi_r^{t+1}\right) c^{t+1} - M_c c^t c^{t+1}$$

Solving for c^{t+1}:

$$c^{t+1} = \frac{c^t}{1 - \Delta t M_c g\left(\phi_p^{t+1}, \phi_r^{t+1}\right) + \Delta t M_c c^t}$$

Due to numerical difficulties associated with large values of the polymerization parameters N_p and N_r in the Flory-Huggins free energy, we used an approximation that was numerically tractable. Large values of N_i result in numerically unresolvable small values for the dilute phase concentrations, so we smoothly interpolated the free energy near 0 with the case when $N_i = 1$ using the following function, $\hat{F}_{FH}$:

$$\begin{aligned} \hat{F}_{FH} = {} & \left[\frac{\phi_p}{N_p}\left(1 - e^{-\alpha\phi_p}\right) + \phi_p e^{-\alpha\phi_p} \right] \ln \phi_p + \left[\frac{\phi_r}{N_r}\left(1 - e^{-\alpha\phi_r}\right) + \phi_r e^{-\alpha\phi_r} \right] \ln \phi_r \\ & + \left(1 - \phi_p - \phi_r\right) \ln\left(1 - \phi_p - \phi_r\right) + \chi_{pr}\phi_p\phi_r + \chi_{pb}\phi_p\left(1 - \phi_p - \phi_r\right) \\ & + \chi_{rb}\phi_r\left(1 - \phi_p - \phi_r\right) \end{aligned}$$

All simulations were run with $\Delta t = 0.0005$ for 4e6 timesteps, $\varepsilon = 2$, $N_p = N_r = 30$, $\alpha = 30$, and $\sigma = 0.0001$. The gelation parameters were $c(t = 0) = \phi_p/1000$, $c_{\phi_p} = c_{\phi_r} = 0.01$, and $M_c = 0.2$.

Figures

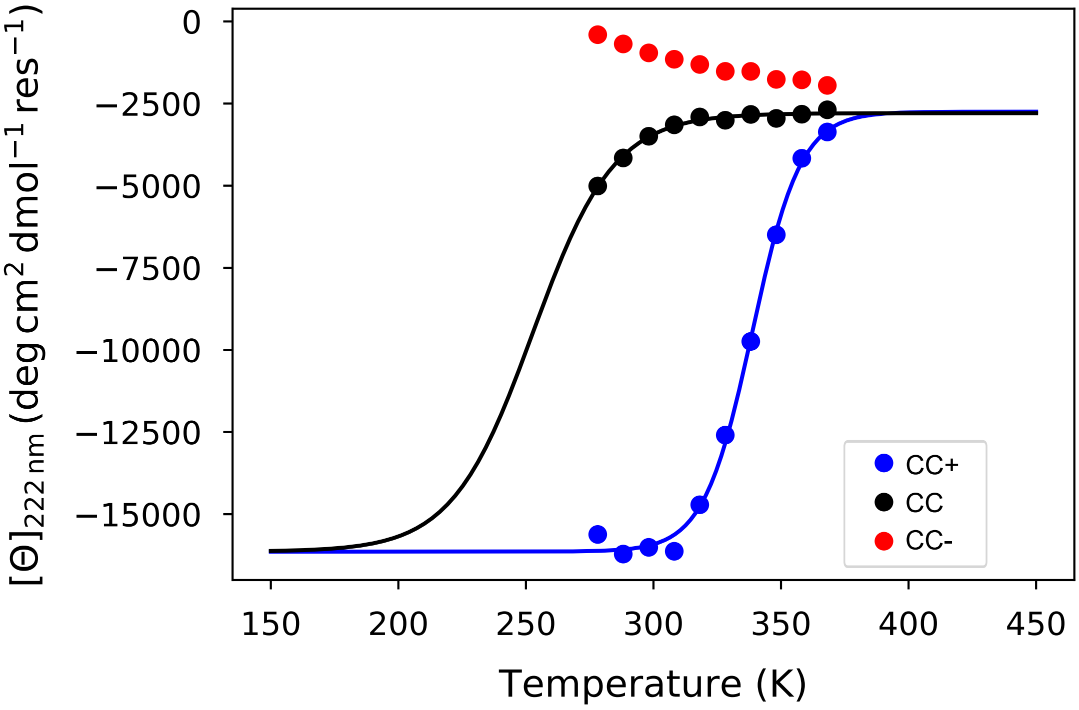


Figure S2.1: Two-state model fits to the mean residue ellipticity at 222nm for each of the three peptides.

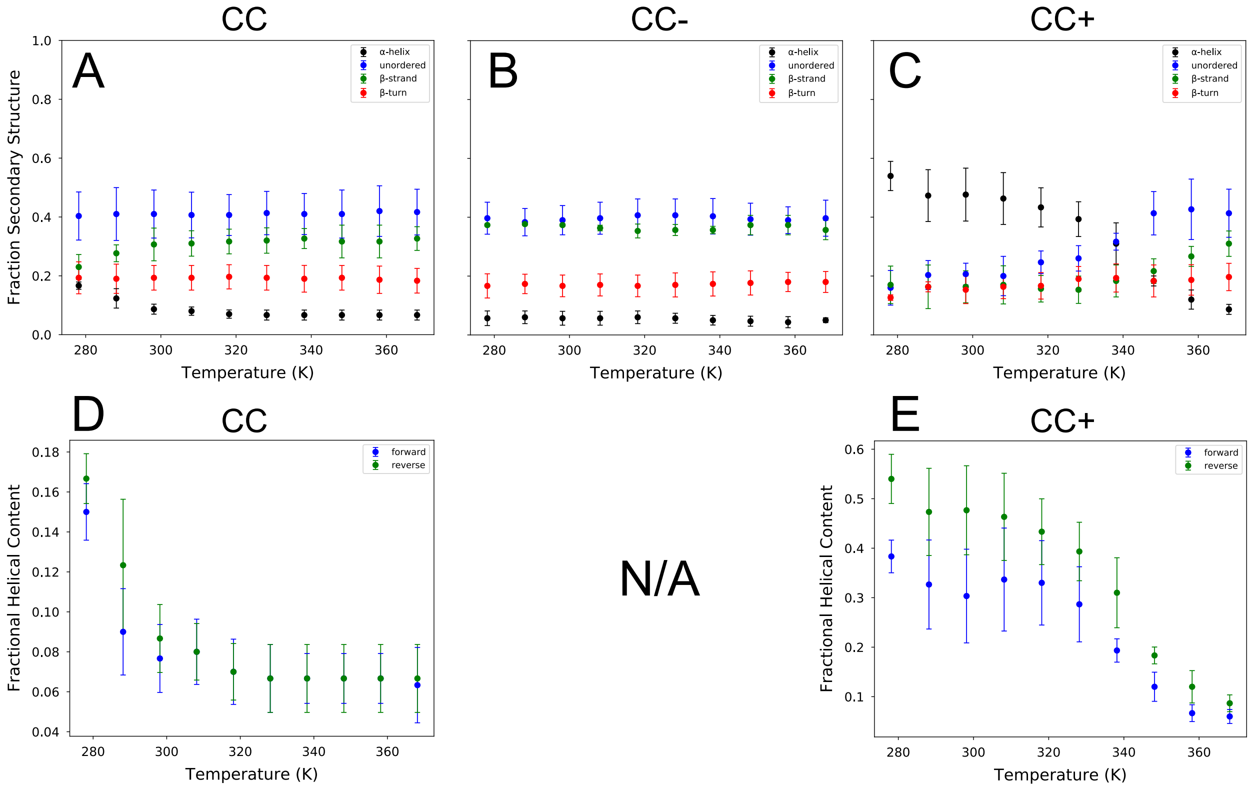


Figure S2.2. Secondary structure contents of each peptide for temperature-dependent CD.. In all plots, secondary structure was determined by deconvolution using the CDSSTR algorithm and basis sets 4, 7, and SP175, as implemented by the Dichroweb server. Filled circles and error bars represent the mean and standard deviation, respectively, of the analysis resulting from these three basis sets. See SI Methods for details. (A) Secondary structure content for the WT peptide. (B) Secondary structure content for the CC- peptide. (C) Secondary structure content for the CC+ peptide. (D) Fractional helical content for the WT peptide for forward and reverse temperature CD scans. (E) Fractional helical content for the CC+ peptide for forward and reverse temperature CD scans.

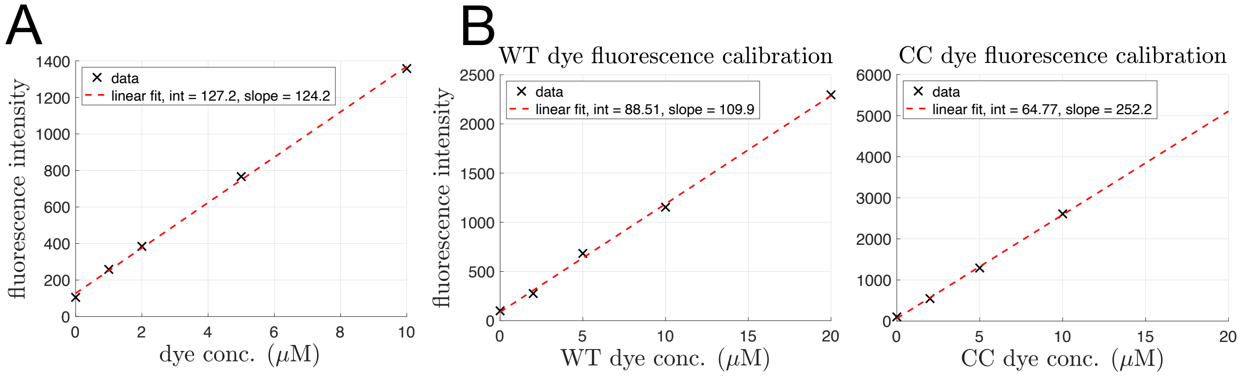


Figure S2.3. Dye fluorescence calibration curves for measuring dense phase Whi3 concentrations.. Fluorescence intensities of Whi3 in experiments fall within the ranges covered by the calibration curves.

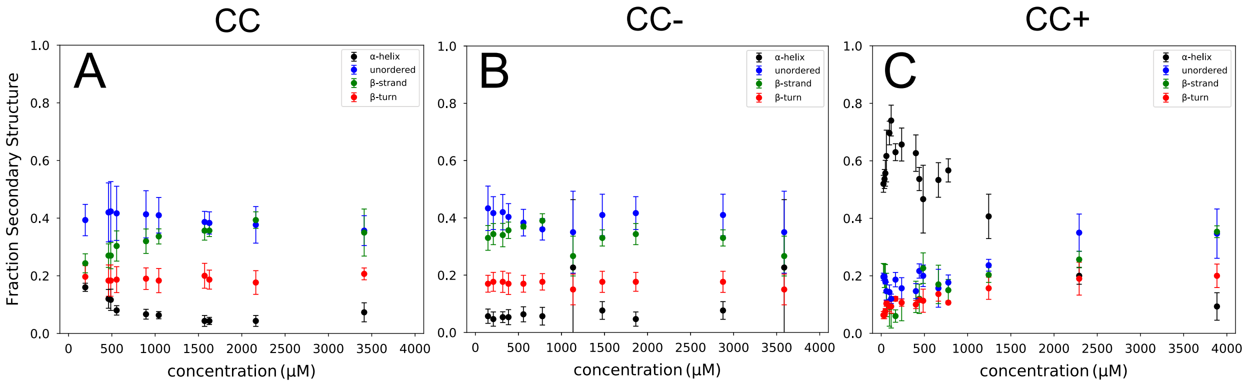


Figure S2.4. Secondary structure contents of each peptide for concentration-dependent CD. In all plots, secondary structure was determined by deconvolution using the CDSSTR algorithm and basis sets 4, 7, and SP175, as implemented by the Dichroweb server. Filled circles and error bars represent the mean and standard deviation, respectively, of the analysis resulting from these three basis sets. See SI Methods for details. (A) Secondary structure content for the WT peptide. (B) Secondary structure content for the CC- peptide. (C) Secondary structure content for the CC+ peptide.

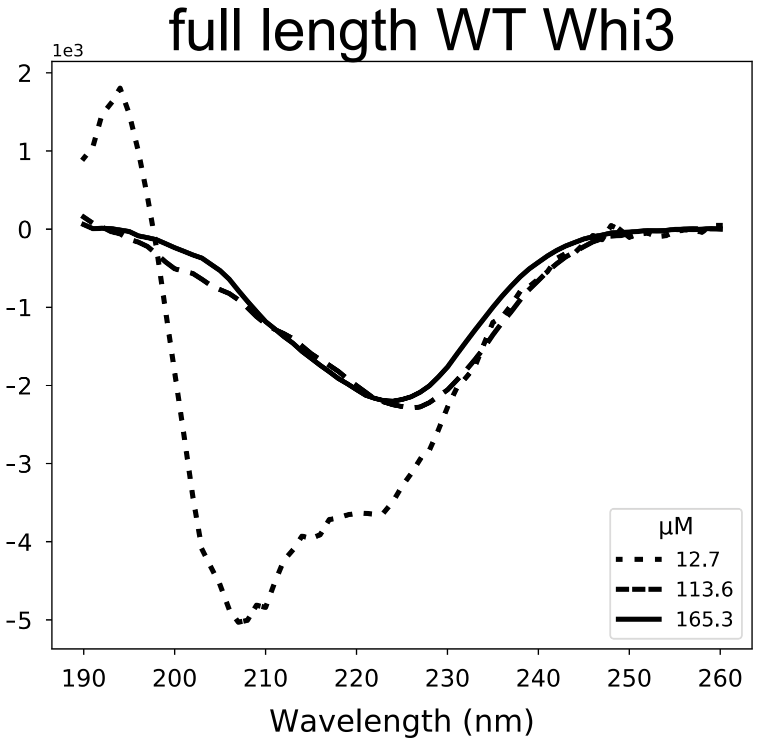


Figure S2.5: CD spectra are shown for full-length WT Whi3 at three concentrations.

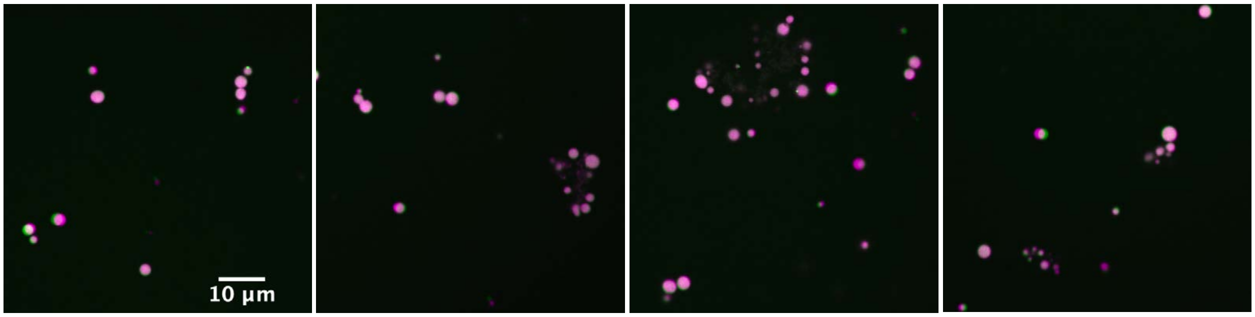


Figure S2.6. Whi3 CC+ structures five hours after mixing 1µM Whi3 with 5nm CLN3. Images are contrasted identically to maximize visualization and do not reflect protein or RNA concentrations as in the main text.

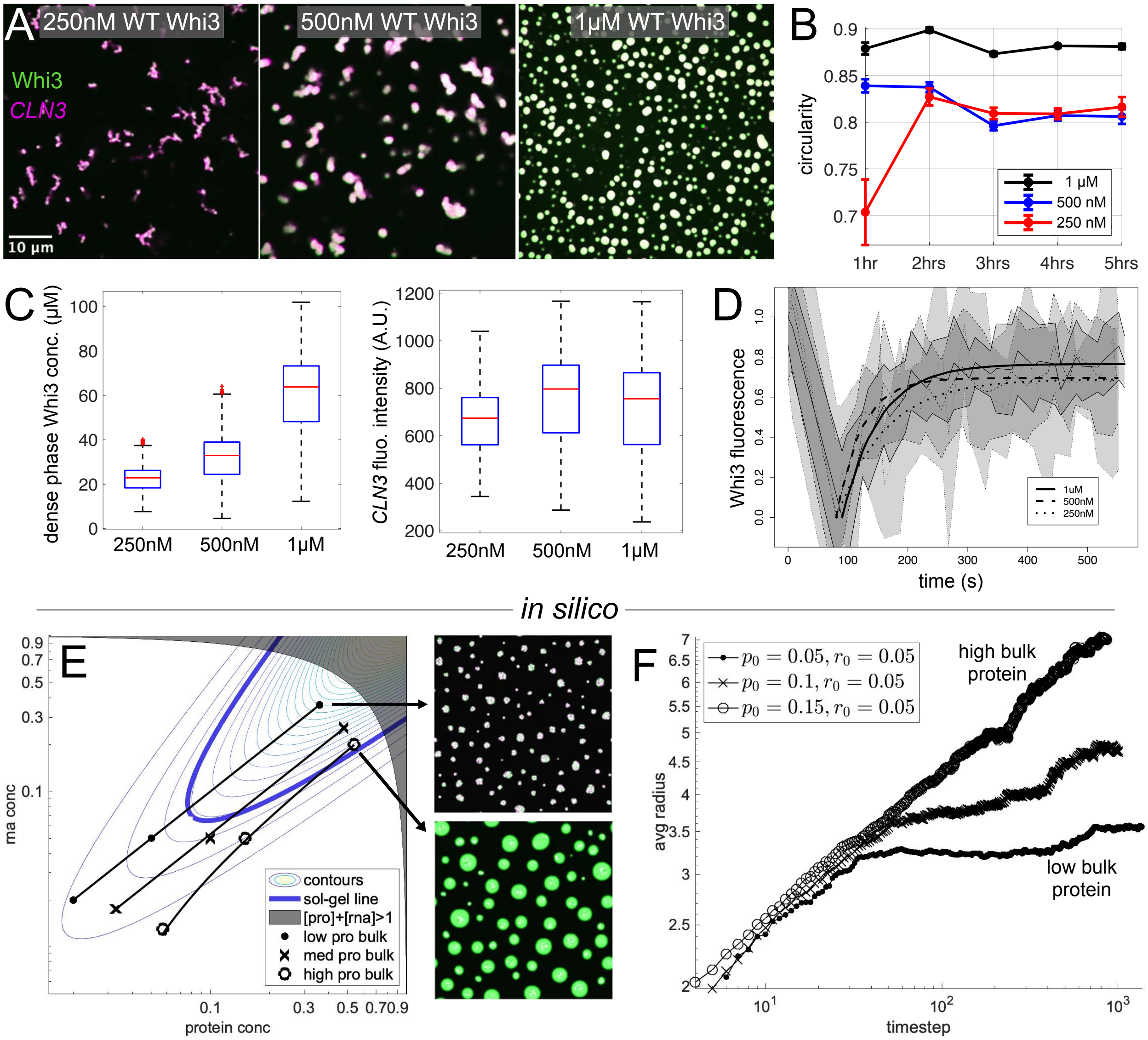


Figure S2.7. Bulk Whi3 concentration controls dense phase stoichiometry and dynamics. (A) Representative images of WT Whi3 at different bulk concentration five hours after mixing with 5nM CLN3. Whi3 (green) and CLN3 (magenta) channels are merged and contrasted identically. (B) Quantification of average droplet circularity for 3 replicates of experiments corresponding to panel (A). Error bars are standard error of the mean. (C) Dense phase Whi3 concentration and CLN3 fluorescence intensity calculated for each condensate in experiments corresponding to panel (A). (D) Fluorescence recovery after photobleaching (FRAP) curves for WT Whi3 with CLN3 at different bulk concentrations of Whi3 at around 2.5 hours after mixing. Solid lines show averages over 5-8 droplets each, with shaded regions indicating +- 1 standard deviation. (E) Parabolic sol-gel contours with the sol-gel line in bold blue are overlaid by dilute, bulk, and dense phase equilibrium concentrations of simulated systems initialized at different bulk protein concentrations but fixed interaction energies. Altered fixed bulk concentrations of protein result in dense phase concentrations that lie at varying depths beyond the sol-gel line. Simulation snapshots after steady state has been reach in the low bulk protein and high bulk protein concentration systems are shown. (F) The average radii of droplets from simulations in panel (E) are shown as a function of time. The legend provides values for the bulk concentrations of protein and RNA that determine whether the systems encode low, medium, or high bulk protein concentrations.

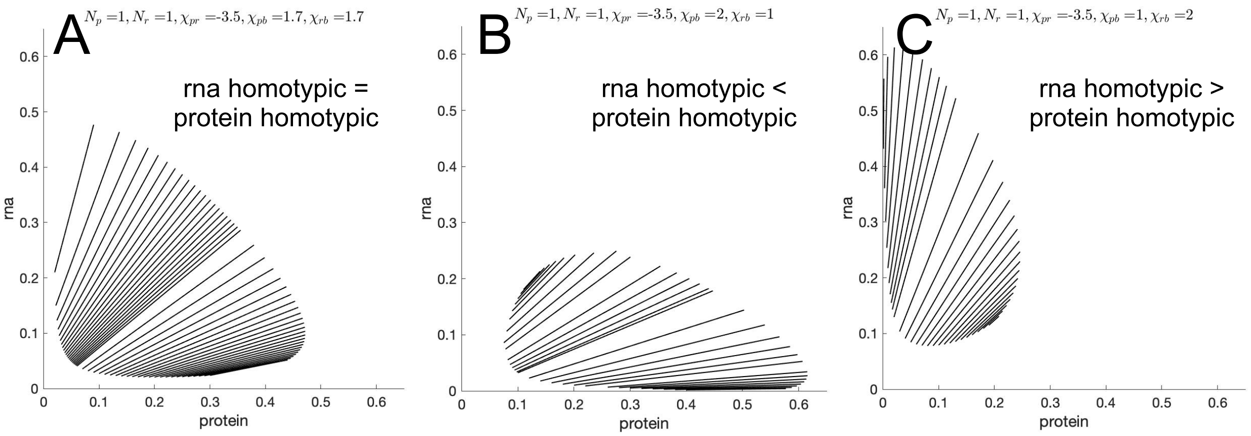


Figure S2.8. The balance of homotypic interactions determines the slopes of tie lines in FH ternary phase diagrams. (A) When protein and rna homotypic interactions are equivalent, a phase diagram that is symmetric about the [rna]=[pro] line results. As more protein or rna is introduced to the system, the slopes of tie lines more closely match the respective axes. (B) When protein homotypic interactions dominate, the phase diagram is altered, and tie lines become more aligned with the protein axis. (C) When rna homotypic interactions dominate, the phase diagram is altered, and tie lines become more aligned with the rna axis. Note that since protein and rna homotypic interaction parameters are switched relative to panel (D), the phase diagram is simply reflected over the line [rna]=[pro].

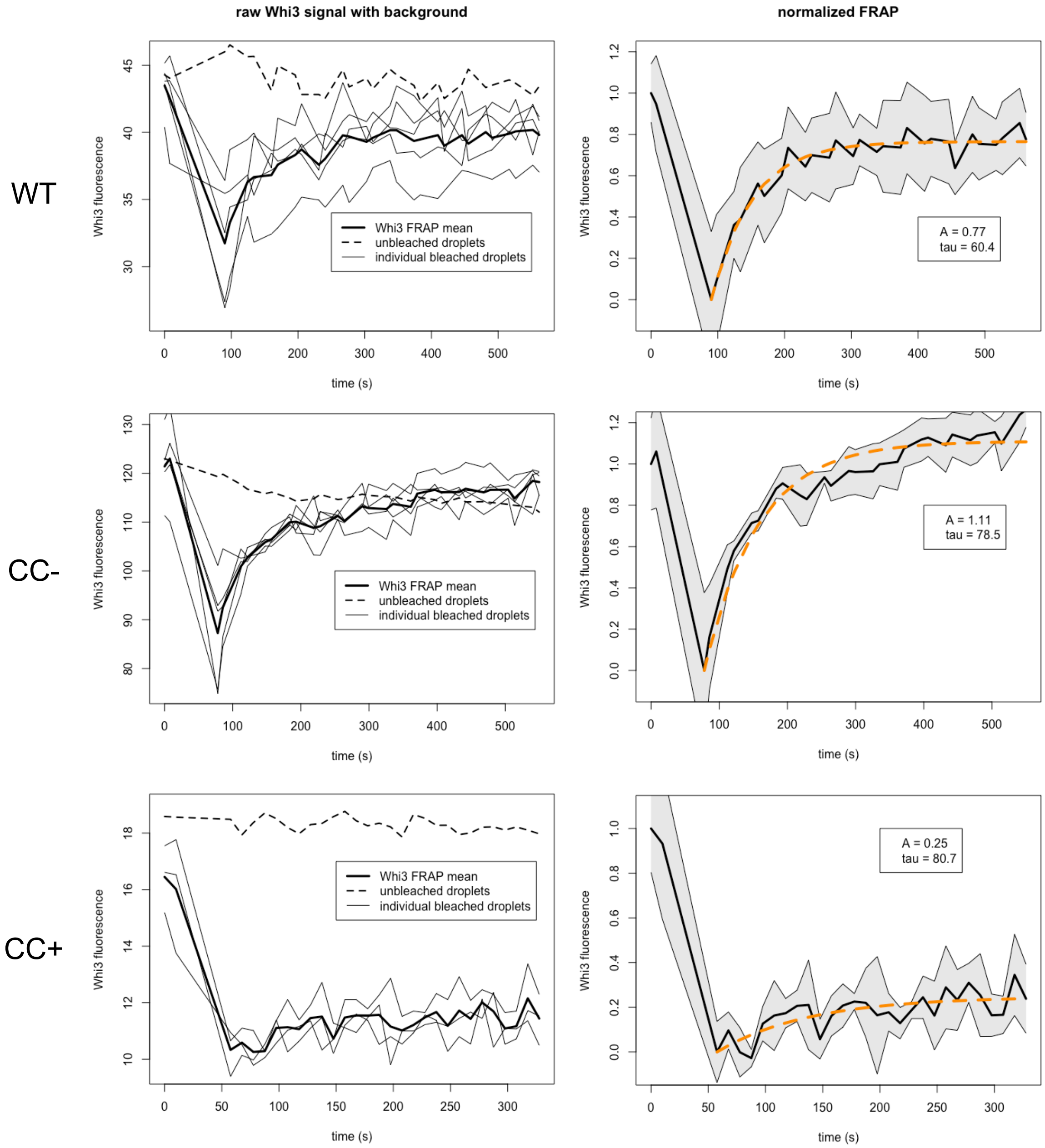


Figure S2.9. FRAP experiment raw data, normalizations, and model fits. In the left column, for each experiment, the raw Whi3 fluorescence signal for each bleached droplet is shown, along with their mean and the mean fluorescence signal of the unbleached droplets in the same field of view. In the right column, the normalized signal is shown along with the model fit and fit parameters.

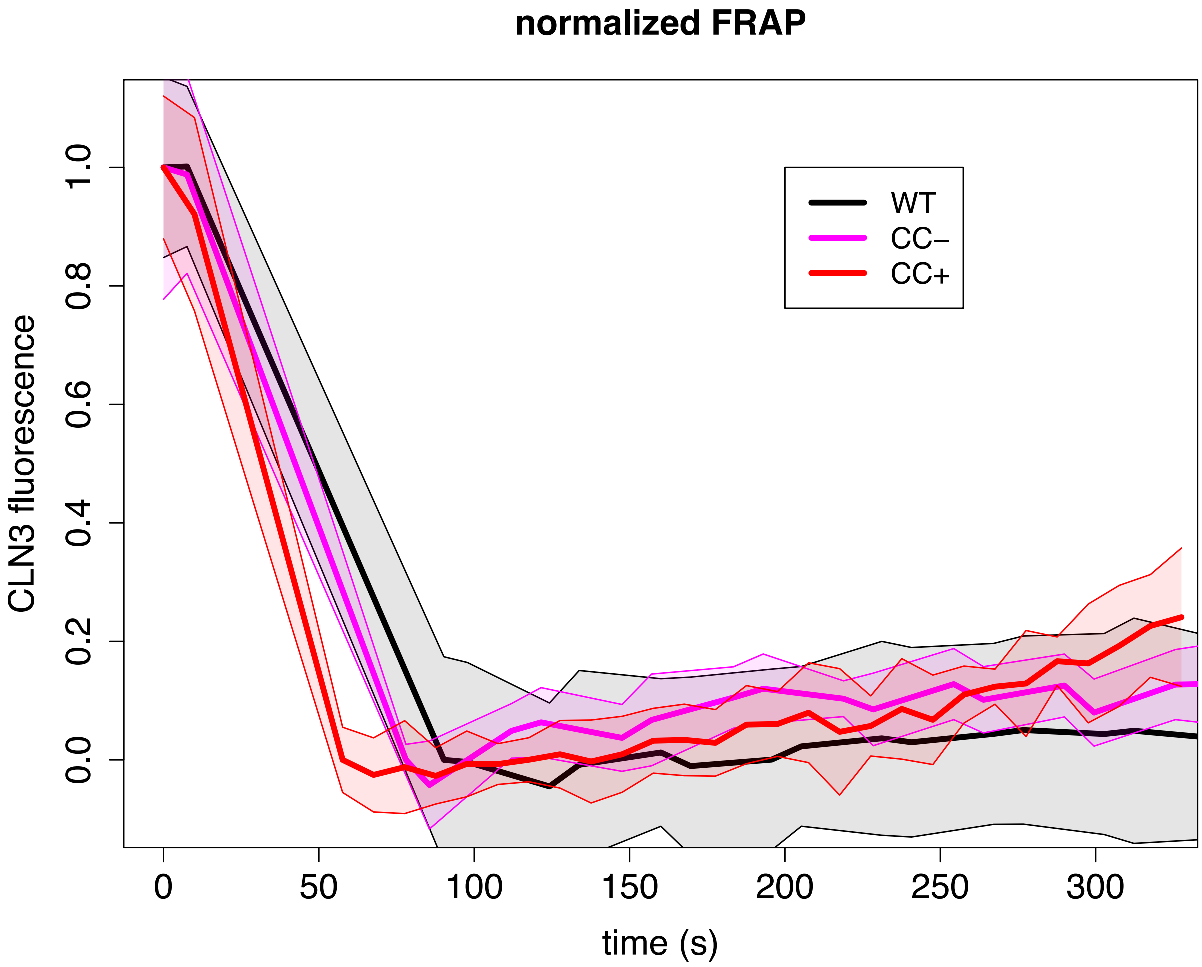


Figure S2.10: Normalized fluorescence recovery after photobleaching (FRAP) curves plus and minus 1 standard deviation for CLN3 mixed with Whi3 CC variants 2 hours after mixing.

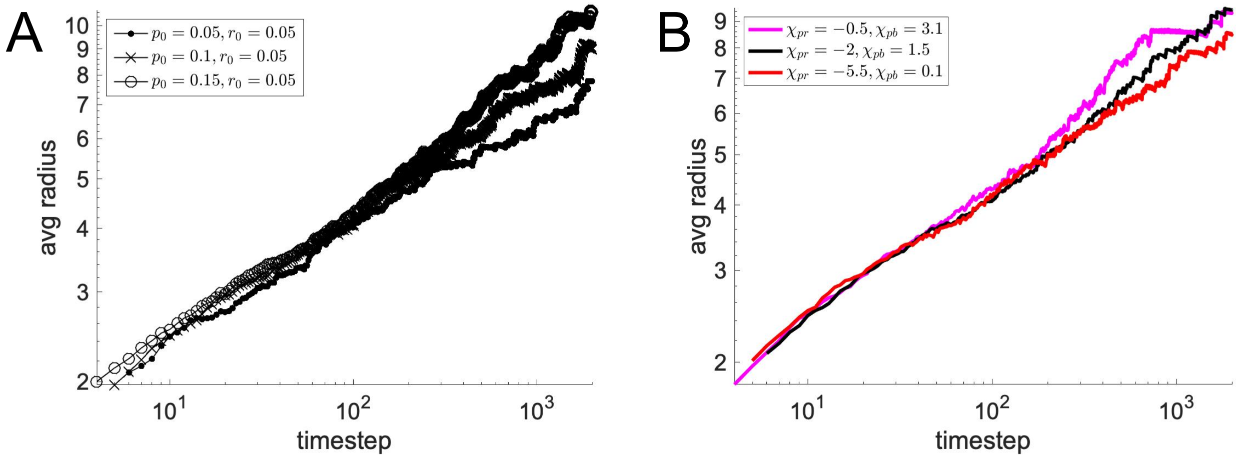


Figure S2.11: Average droplet size is not significantly affected without gelation. (A) Average droplet size as bulk concentration of protein is increased. At long times when each system is a single droplet, the average droplet radius will differ. (B) Average droplet size as bulk concentrations are fixed, but interaction parameters change.

Primer Name	**Construct**	**Sequence**
AGO3201	WT, CC-, CC+, ΔCC	ctttaagaaggagatatacatatgcaccatcatcatcatcatatgtcgctggttaacagt
AGO3024	WT, CC-, CC+, ΔCC	aagcttgtcgacggagctcgtcaagatttgccgaaggc
AGO3086	ΔCC	ATGCTGGTGATGCTGGAGTGCTTGggattgtggc
AGO3090	ΔCC	gccacaatccCAAGCACTCCAGCATCACCAGCATcagcagcaacagtctctttc
AGO3085	CC-	gGGTGGTGCTGCTGCTGCTGGgGTTGCTGCTGCTGGTGATGCggatgctggtgatgctgg
AGO3084	CC-	ccGCATCACCAGCAGCAGCAACcCCAGCAGCAGCAGCACCACCc-CCAGCAGCAGCAGCAC
AGO3087	CC+	tGCAGCAGCAACTCCAGCAGCtGCAGCACCACCTCCAGCAGCtGCAGCACCAGCAG
AGO3088	CC+	aGCTGCTGGAGGTGGTGCTGCaGCTGCTGGAGTTGCTGCTGCaGGTGATGCAAatgctgg
AGO3001	CLN3	TAATACGACTCACTATAGGGAGAGTCTGCATACCAAAGATCAGCCGCTTGC
AGO3002	CLN3	GTATGCTAGCGTAGTTTCTTGACC

Table S2.1: Whi3 and CLN3 constructs

WT isotropic	CC	QRR
CC	-0.1	-0.1
QRR	-0.1	-0.2
CC+ isotropic	CC	QRR
CC	0	0
QRR	0	-0.2
CC- isotropic	CC	QRR
CC	-0.2	-0.2
QRR	-0.2	-0.2

Table S2.2: LASSI simulation isotropic energies

WT anisotropic	CC	QRR
CC	-2	0
QRR	0	0
CC+ anisotropic	CC	QRR
CC	-4	0
QRR	0	0
CC- anisotropic	CC	QRR
CC	0	0
QRR	0	0

Table S2.3: LASSI simulation anisotropic energies

CHAPTER 3: THE ROLE OF SPATIAL PATTERNING OF N-PROTEIN INTERACTIONS IN SARS-COV-2 GENOME PACKAGING[4]

Introduction

Biomolecular condensation is a simple and versatile way for cells to spatially and temporally control biochemistry. It is now clear that a wide variety of compartments likely form using the process of liquid-liquid phase separation (LLPS) which leads to a condensation of specific components out of bulk cytosol or nucleoplasm (119). The protein components of condensates tend to contain intrinsically disordered or low complexity sequences and RNA-binding domains (120). Many condensates also contain nucleic acids, and indeed RNA can promote phase separation in many instances (20). However, the contributions of specific RNA sequences and structures in condensate assembly, contents, and material properties is poorly understood (20).

Viruses present a powerful system to examine sequence specificity for both proteins and nucleic acids in phase separation because of their highly compact genomes and limited protein coding genes. Indeed, reports have emerged for VSV (121), respiratory syncytial virus (122), rabies (123), measles (124), and HIV (125) components showing the capacity to undergo LLPS.

SARS-CoV-2 is a positive strand RNA virus that has an exceptionally large genome of ~30kb which is selectively packaged into a relatively small capsid estimated to be ~100 nm in diameter (126). How the genome is selectively packaged while excluding sub-genomic RNAs

---

[4] This chapter previously appeared as an article in Biophysical Journal. The original citation is as follows: Ian Seim, Christine A. Roden, Amy S. Gladfelter, Role of spatial patterning of N-protein interactions in SARS-CoV-2 genome packaging. *Biophysical Journal* **120**, 2771-2784 (2021).

generated by the virus and the host transcriptome and sufficiently compressed to fit into a virion is not yet understood. The necessity of a packaging signal for SARS-CoV-1 is still not clear, although one sequence has been found to be sufficient but not necessary to package RNA (127). To our knowledge, no packaging signal has been identified for SARS-CoV-2.

The nucleocapsid protein of SARS-CoV-2 undergoes LLPS (30, 42, 128-130), and our work found that this occurs in an RNA sequence-specific manner with different regions of the genome (42). Remarkably, RNAs of the same length can either promote or limit phase separation depending on the sequences. The sequences with differing behavior also show distinct patterns of binding of N-protein with LLPS-promoting sequences having discrete patterned N-protein interactions, while RNA sequences that limit phase separation are uniformly coated in N-protein. The regions that promote phase separation are in the 5′ and 3′ ends of the genome, prompting us to speculate that phase separation could be relevant to packaging, as these LLPS-promoting sequences are present specifically on the whole genome and would not be together on sub-genomic or host RNAs.

Here, we developed a coarse-grained model to test the hypothesis that phase separation could be a relevant process for selecting and compacting a single genome. Our goal was to examine how the linear location of different RNA sequences in the genome generates spatially segregated and condensed RNA molecules. We first explored fragments of the SARS-CoV-2 genome that have opposing phase behavior when mixed with N-protein, as shown in (42). Specifically, the 5′ and 3′ ends of the genome promote phase separation, while the frameshifting element (FE) and central regions of the genome solubilize N-protein (**Fig. 3.1A**). We next examined the spatial patterning of these opposing elements within a full genome model and quantified its effects on phase separation, packaging of single genomes, genome compaction, and genome cyclization. We found that in this model, localization of LLPS-promoting features to the 5′ and 3′ ends of the genome is sufficient to drive LLPS-based single genome packaging and genome compaction, and is necessary for genome cyclization. Addition of clustered LLPS-promoting features throughout the genome further enhanced all of these metrics.

Methods

LASSI model parameterization

Simulations were performed using LASSI (40) and run on the Longleaf computer cluster at UNC-Chapel Hill and on the Comet XSEDE cluster at the San Diego Supercomputer Center (131). Each simulation was run independently on a single compute node with 4GB RAM. The following default parameter sets were used for all simulations (**Tables 3.1, 3.2**):

Isotropic Binding Energies	RBD1	RBD2/dimer	5′ end RNA	FE RNA	5′ gRNA terminus	3′ gRNA terminus
RBD1	-0.5	0.0	-1.2	0.0	0.0	0.0
RBD2/dimer	0.0	0.0	0.0	0.0	0.0	0.0
5′ end RNA	-1.2	0.0	1.0	1.0	1.0	1.0
FE RNA	0.0	0.0	1.0	1.0	1.0	1.0
5′ gRNA terminus	0.0	0.0	1.0	1.0	1.0	0.0
3′ gRNA terminus	0.0	0.0	1.0	1.0	0.0	1.0

Table 3.1: Isotropic binding energies

Anisotropic Binding Energies	RBD1	RBD2/dimer	5′ end RNA	FE RNA	5′ gRNA terminus	3′ gRNA terminus
RBD1	0.0	0.0	0.0	0.0	0.0	0.0
RBD2/dimer	0.0	-3.0	0.0	-5.0	0.0	0.0
5′ end RNA	0.0	0.0	0.0	0.0	0.0	0.0
FE RNA	0.0	-5.0	0.0	0.0	0.0	0.0
5′ gRNA terminus	0.0	0.0	0.0	0.0	0.0	-3.0
3′ gRNA terminus	0.0	0.0	0.0	0.0	-3.0	0.0

Table 3.2: Anisotropic binding energies

Negative energies indicate attraction and positive energies indicate repulsion. The appropriate subsets of these interactions were used for the simulations involving N-protein alone, N-protein and 5′ end RNA, N-protein and FE RNA, and all spatial rearrangements of beads in genomic RNA (gRNA) simulations. We performed parameter sweeps over many of these energies for the N-protein only system, the 5′ end and FE RNA fragments with N-protein systems, and the N-protein with gRNA systems, explained further in the main text and Supplemental Material (**Figs. S3.2-S3.4, S3.7, S3.8**). Simulations of N-protein alone were run at 15 temperatures linearly spaced between 1 arbitrary unit (a.u.) and 2 a.u., and 15 concentrations logarithmically spaced between 1e-4 and 1e-0.3, each for 1e9 total time steps with 5e6 timesteps of thermalization. Temperature scales interaction energies, ε, as $\frac{\varepsilon}{T}$. During thermalization, the temperature is raised to 1000 a.u., anisotropic interactions are inactivated, and all chains are pushed to the center of the system, as described in (40). For systems with short chains, like the 6- and 2-bead chains representing 5′ end and FE RNA fragments and N-protein, respectively, this step allows for a more rapid convergence to equilibrium if the system will phase separate. For very long chains, this step may kinetically trap the system in a clustered state. For systems with N-protein and 5′ end and FE RNA fragments, simulations were run at a single temperature, 1 a.u., for 1e9 timesteps and 5e6 steps of thermalization. For genomic RNA (gRNA) with N-protein systems, simulations were run at a single temperature, 1 a.u., for 10e9 timesteps with no thermalization due to kinetic trapping when thermalization was used. These simulations equilibrated to similar minimum energies with or without thermalization, but had altered packaging properties (**Fig. S3.5, S3.6**). We concluded that since gRNA chains in thermalized systems gradually disengaged from the initial central cluster over the course of simulations that their equilibrium states were not fully clustered at the center, as observed for simulations without thermalization (**Fig. S3.5**) and that they were instead kinetically trapped. Thus, we did not use thermalization for gRNA systems. For all ternary systems, concentrations of each component were chosen to be logarithmically spaced across the ranges of interest, with hundreds of ([protein], [RNA]) coordinates sampled for each system (**Fig. S3.1**). For N-protein with 5′ end and FE RNA, the maximum number of N-protein and RNA chains were used such that the total beads in the system never exceeded 40000 and the target stoichiometries were satisfied. For simulations with gRNA and N-protein, 10 gRNA molecules were always used, and the number of N-protein chains was altered to match target stoichiometries. At the lowest gRNA and highest N-protein volume fractions, the number of gRNA molecules was gradually scaled down to 1 due to computational limitations on the sizes of the systems. At each coordinate, 2 independent simulations were run with and without interactions, for a total of 4 simulations at each volume fraction coordinate.

Simulation analysis

Analysis was performed using scripts within LASSI and custom scripts that relied upon the ovito python module (114). LASSI outputs a global density inhomogeneity value, $\bar{\rho}$, for each simulation, which is used to determine whether phase separation has occurred. $\bar{\rho}$ is calculated using the pair distribution function for all beads in the system by default, as described in (40). We used this default calculation for N-protein-only systems and the 5′ end and FE RNA fragments with N-protein systems. However, for gRNA systems, we did not observe a dilute phase of gRNA chains and concluded that phase separation of gRNA in this context was not meaningful (more detail in main text). Instead, we calculated a new density inhomogeneity metric, $\bar{\rho}(N)$, that uses the pair distribution function for only N-protein beads in the system. Using this limited pair distribution function, we calculated $\bar{\rho}(N)$ analogously to $\bar{\rho}$ as described in

(40). Contours at $\bar{\rho} = 0.025$ were used to determine phase boundaries for N-protein only and $5'$ end and FE RNA fragments systems. This value of $\bar{\rho}$ was shown in (40) to universally indicate the onset of phase separation. We used $\bar{\rho}(N) = 0.02$ for gRNA systems as this value better aligned with sharp transitions in $\bar{\rho}(N)$ as a function of volume fraction and resulted in smoother phase boundaries. We used the ovito module to calculate clusters using a maximum cutoff of $3^{1/2}$, the maximum distance between two interacting particles in a cubic lattice. The radius of gyration for each gRNA molecule was calculated according to the equation in the text. Ovito was also used to calculate end-to-end distances of genomic RNAs. We counted genomes as cyclized if their terminal beads were within $3^{1/2}$ units of each other. For all analysis of packaging metrics in gRNA systems, 100 simulation snapshots were used from the last half of simulations for each of 2 runs and averaged over time and runs. Snapshots were recorded starting at 5e9 steps every 5e7 steps and used for analysis. For $5'$ end and FE RNA fragments with N-protein systems, 50 simulations snapshots were recorded starting 2e7 steps after thermalization completion and used for analysis. For all systems, $\bar{\rho}$ or $\bar{\rho}(N)$ and clustering and packaging metrics were interpolated to a grid linearly spaced between 1e-4 and 1 with a discretization of 80 points along the protein and RNA axes. Interpolation was performed using interpolate.griddata from scipy. Plots were made using matplotlib.

Results

Simulations recapitulate known phase behavior and binding patterns of N-protein with FE and $5'$ end RNA

We first sought to characterize the phase behavior of different regions of the viral genome by focusing on simulations involving N-protein with either the first 1000 nucleotides of the genome ($5'$ end RNA) or 1000 nucleotides around the frameshifting region located between ORF1A and ORF1B (FE RNA). Simulations were carried out using LASSI, which employs stickers-and-spacers representations of polymers and generates full phase diagrams by performing Monte Carlo simulations at many temperatures and volume fractions of components (40). Simulations take place on a cubic lattice, and only a single particle can occupy a given lattice site at one time. In this work, polymer beads are connected by implicit linkers that do not occupy space but rather guarantee that adjacent beads within chains are always in adjacent lattice sites. The first consideration was how to represent the N-protein and RNA in a coarse-grained manner based on existing data. The N-protein has two RNA binding domains (RBDs), a dimerization domain which overlaps with RBD2, and three intrinsically disordered regions (IDRs) (**Fig. 3.1A**). RBD1 is conserved across multiple betacoronavirus genomes and has previously been shown to interact with the conserved sequences and structures in the 5′ UTR (132). We previously demonstrated that a single point mutant within RBD1, Y109A, greatly reduces N-protein phase separation and changes N-protein interactions with 5′ end RNA, while only minimally affecting protein binding and phase behavior with FE RNA (42). Since FE RNA primarily solubilized N-protein, and deletion of the RBD2/dimerization domain blocked N-protein co-phase separation with RNA (128), we postulated that FE may block N-protein phase separation by specifically interacting with RBD2 and preventing N-protein dimerization. Thus, we hypothesize that N-protein RBD1 primarily binds to 5′ end RNA, while RBD2 primarily binds to FE RNA.

Based on these data, we represented the N-protein as simply as possible using two spheres. The first sphere participates in weak isotropic interactions with other N-proteins (representing the association of the IDRs) and with the 5′ end RNA (representing RBD1). The second sphere participates in strong anisotropic interactions with other N-proteins to capture dimerization and the interaction with the FE RNA via RBD2 (**Fig. 3.1A**). These latter anisotropic interactions operate under the assumption that N-protein dimerization competes with binding to FE RNA, since anisotropic binding in this model is one-to-one. To model charge effects, the RNA molecules experience an isotropic repulsive force amongst themselves, both within chains and among distinct RNA chains. Using these specifications, we sought to qualitatively reproduce the phase behavior among these molecules established experimentally in (42). We found that, relative to N-protein phase separation on its own (**Fig. 3.1B**), addition of 5′ end RNA promotes enhanced phase separation across a wide range of protein and RNA concentrations (**Fig. 3.1C**), while addition of FE RNA does not promote phase separation and solubilizes N-protein at sufficiently high concentrations of RNA (**Fig. 3.1D**).

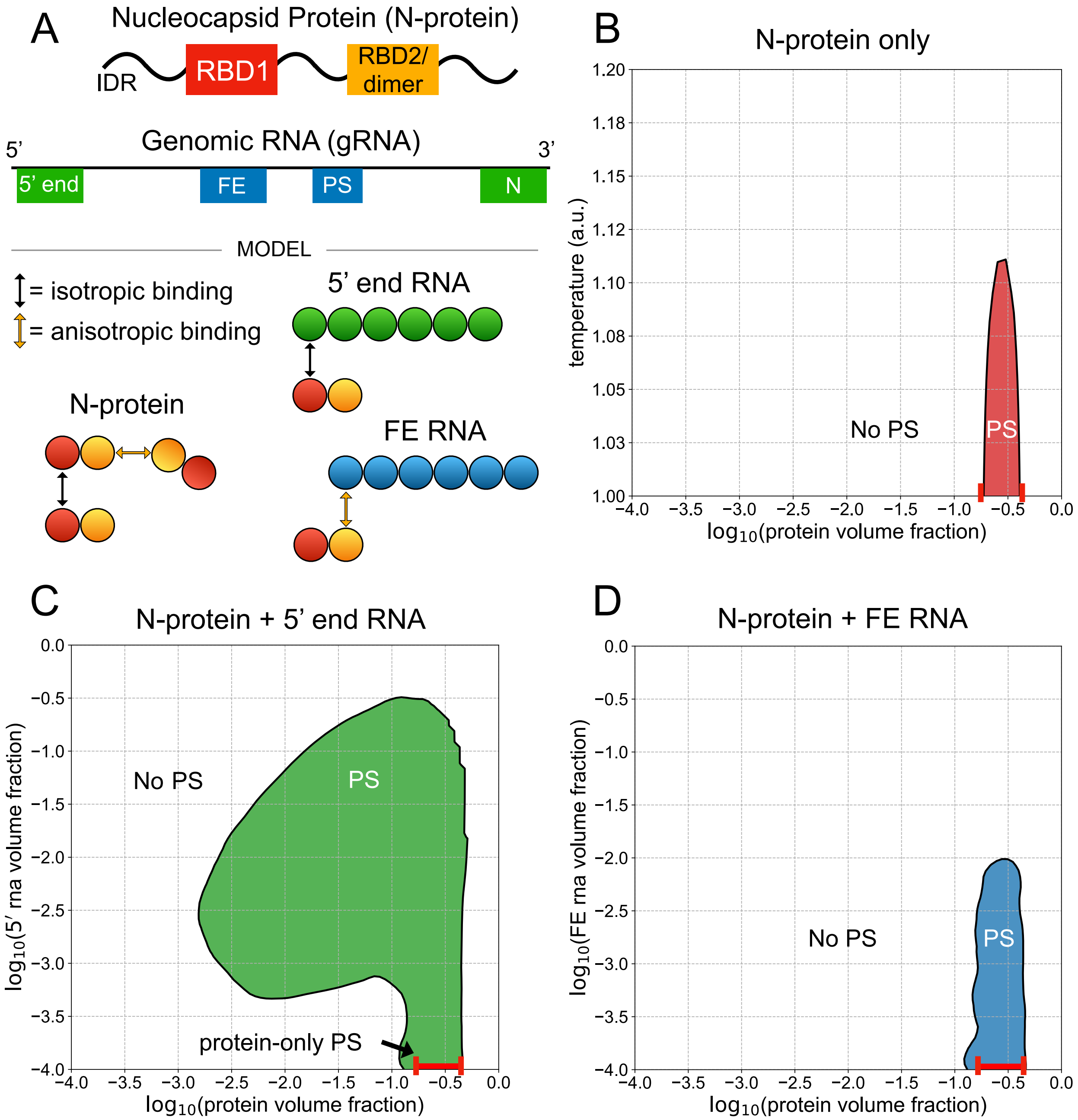


Figure 3.1: 5′ end and FE RNA with N-protein have opposing phase behavior. (A) N-protein is represented as a two-bead chain, with the first bead participating in isotropic homotypic interactions, and the second bead participating in anisotropic homotypic interactions. Both 5′ end and FE RNA segments are roughly three times larger than N-protein and are represented by six beads each. N-protein interacts with all 5′ end RNA beads via isotropic binding with its first bead, and it interacts with all FE RNA beads via anisotropic binding with its second bead. This interaction with FE competes with N-protein dimerization. (B) N-protein phase separates (PS) in a narrow concentration and temperature range on its own. (C) N-protein with 5′ end RNA at temperature 1 a.u. phase separates across a wider concentration range than on its own. (D) N-protein with FE RNA at temperature 1 a.u. is solubilized at sufficiently high FE RNA concentrations.

In addition to opposing phase behavior, the binding of N-protein to 5′ end and FE RNA as a function of N-protein concentration was shown to be distinctly patterned based on protein crosslinking in (42). While FE RNA is uniformly coated with protein across a wide range of protein concentrations, 5′ end RNA has a few discrete binding sites and is only gradually coated more with protein as the protein concentration is increased. The simulations report a similar distinct pattern of protein interactions. For a fixed RNA concentration and at low bulk protein concentrations, 5′ end RNA is initially coated with very little protein (**Fig. 3.2A**). As the bulk protein concentration is increased, 5′ end RNA sharply transitions to binding large amounts of protein (**Fig. 3.2A**). In contrast, FE RNA binds more protein at low protein concentrations, relative to 5′ end RNA, and experiences a more gradual transition to high amounts of bound protein (**Fig. 3.2B**). In all, the distinct protein binding behavior of these two RNA elements is consistent with that found in (42) (**Fig. 3.2C**). Thus, the simulation results are consistent with the experimental system and support that the coarse-grained representations and binding energies are reasonable approximations of the actual system.

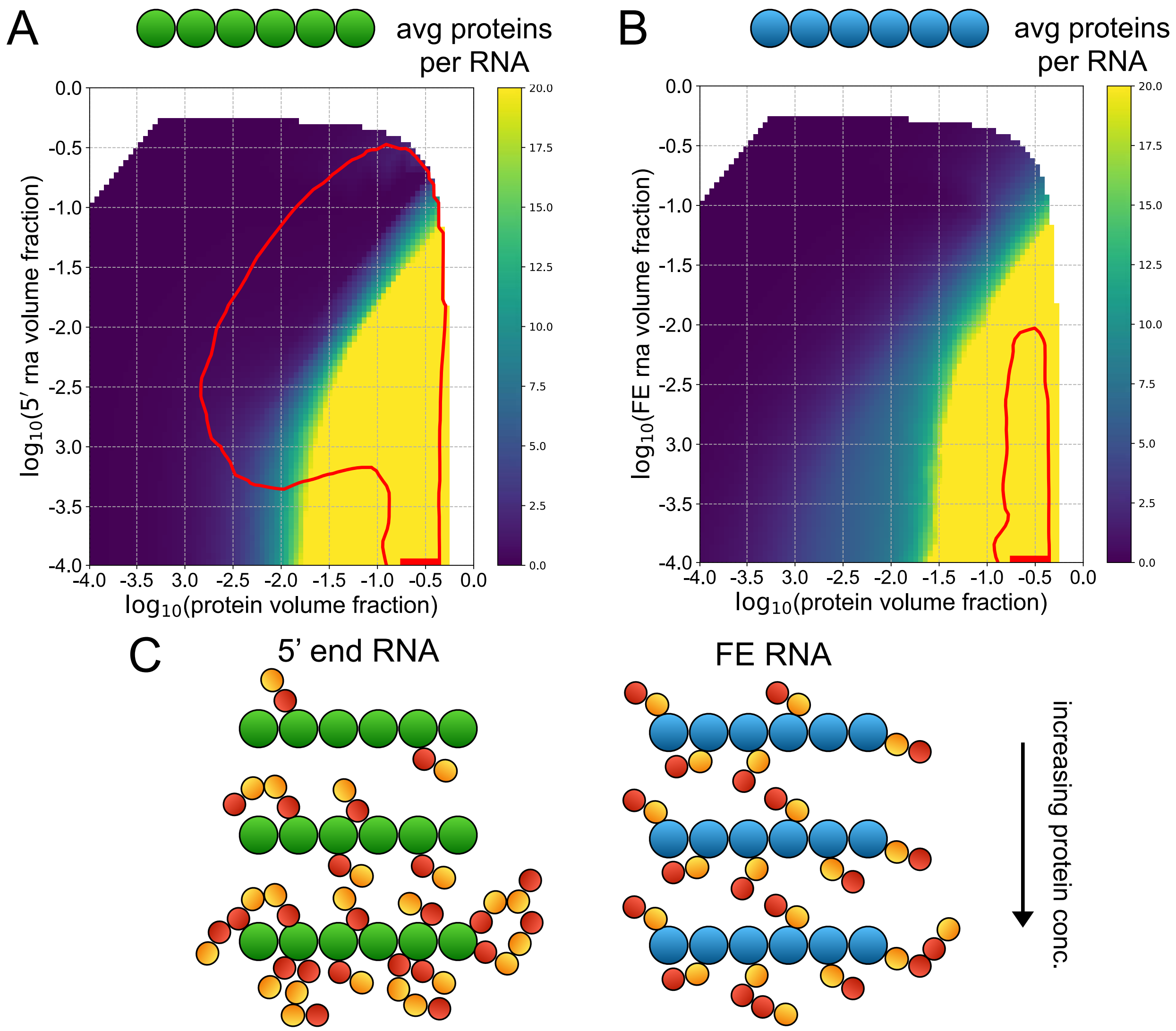


Figure 3.2: 5′ end and FE RNA have distinct N-protein binding behavior. (A) and (B) Phase boundaries for 5′ end RNA with N-protein (A) and FE RNA (B) with N-protein are shown in red. The heatmaps indicate the average number of protein chains per RNA chain for each cluster identified in each simulation. If more than one RNA chain is in a cluster, the ratio of protein chains to RNA chains within that cluster is reported. (C) For any given fixed RNA concentration, as the protein concentration is increased, 5′ end RNA shows a sharp transition to highly bound protein, while FE shows a more gradual transition, with more proteins bound at low protein concentrations relative to the 5′ end system.

We ran extensive parameter sweeps for these systems to determine the contributions of specific interactions. For N-protein alone, the isotropic binding energy greatly influenced phase behavior (**Fig. S3.2**). Dimerization was necessary for phase separation for low isotropic binding energies, including the default energy used for all following simulations (**Fig. S3.2**).

Interestingly, the strength of the dimerization interaction was non-monotonically related to phase separation propensity, with very high dimerization energies starting to weaken phase separation (**Fig. S3.2**). We also studied the effects of altering the strength of binding between N-protein and 5′ end and FE RNA. The strength of the isotropic binding interaction between N-protein and 5′ end RNA strongly influenced co-phase behavior (**Fig. S3.3**). In contrast, the strength of the anisotropic binding between N-protein and FE RNA did not alter phase behavior until very high energies, when a new arm of the phase diagram emerged (**Fig. S3.3**). We did observe differences in the average number of N-proteins bound to FE outside of the phase envelope, with low anisotropic energies leading to less binding (**Fig. S3.3**). We doubled the magnitude of the RNA-RNA repulsive isotropic energy and observed small effects on the phase behavior of N-protein with both 5′ end RNA and FE RNA (**Fig. S3.4**). Finally, we increased the isotropic binding energy between N-proteins and observed phase behavior in the context of 5′ end and FE RNA. Since N-protein alone phase separates across a wider range of volume fractions under this condition, we see a corresponding widening of the phase diagrams across these concentrations when mixed with both RNA fragments (**Fig. S3.4**). Additionally, more FE RNA is required to solubilize N-protein since the N-proteins are more stably phase separated (**Fig. S3.4**). The most sensitive interactions for these systems appear to be the isotropic binding between N-proteins and the isotropic binding of N-protein with 5′ end RNA.

Whole genome simulations reveal effects of phase separation on single genome packaging, genome compaction, and genome cyclization

The large size (30kb) of the genome makes it challenging to synthesize *in vitro* for experiments. We therefore were eager to use this simulation space to ask questions about how the different RNA-sequence elements will behave when present *in cis* on the same polymer, as they are found in the native virus. To address if specific arrangements of RNA encoded features could be sufficient for packaging a single genome, we utilized the same representations of the 5′ end and FE RNA and the N-protein described above and assembled a system that represents the N-protein and the entire viral genome (gRNA) (**Fig. 3.3A**). In addition to the 5′ end of the gRNA, the 3′ end was also found to promote phase separation with N-protein and shared similar protein binding behavior and RNA sequence features (**Fig. 3.1A**) (42). Central regions of the gRNA that were studied behaved similarly to the FE RNA, and further, the internal portions of the genome were predicted to be more similar to the FE than 5′ end RNA (42). Taking into account the relative sizes of nucleotides and amino acids, the gRNA is roughly 90x larger than N-protein, so our representation involves 180-bead chains for each genome. An additional strong, anisotropic interaction between the terminal beads of each gRNA chain is added to represent known nucleotide complementarity between the 5′ and 3′ ends of the gRNA and propensity to crosslink *in vivo* (133).

Due to the length of gRNA chains, we were only able to include 10 chains per simulation, with hundreds to thousands of N-proteins. Interestingly, we did not observe dilute phases of gRNA chains under any conditions; there were no well-defined clusters including all gRNA chains, and instead, all gRNA chains were coated with N-protein. N-protein demonstrated a well-defined dilute phase and a dense phase on and around gRNA chains. For these reasons, we determined that phase separation of gRNA in this context was not well-defined, and regarded gRNA instead as a surface upon which N-protein phase separation occurred. To quantify this behavior, we altered the definition of the density inhomogeneity metric, $\bar{\rho}$, originally defined as a function of the pair distribution functions between all beads in a given simulation as explained in (16). Since we observed a density transition only of N-protein in gRNA-containing systems, we defined a new metric, $\bar{\rho}(N)$, which depends only on the pair distribution function of N-proteins in the system. Thus, the following phase diagrams describing gRNA-containing systems are no longer ternary phase diagrams as in **Figs. 3.1** and **3.2** with well-defined dilute phases of both components. They instead indicate the density inhomogeneity of N-protein as a function of the volume fraction of N-protein and the surface upon which they condense, gRNA.

For a given N-protein concentration, phase separation of N-protein in the gRNA system occurs over a smaller range of RNA concentrations relative to the 5′ end system (**Fig. 3.1C**, **Fig. 3.3B**). This result is consistent with experiments in which 5′ end and FE RNAs were combined *in trans* which led to more limited phase separation compared to 5′ end RNA alone (42). However, phase separation of N-protein can occur at lower volume fractions of N-protein and gRNA due to the length of the gRNA chains, which is also consistent with experiments that were performed with RNA purified from infected cells that contained gRNA (42). We quantified the average amount of protein bound to single genomes and found that the contour delineating an equal volume fraction of protein and gRNA per cluster aligns well with the high-RNA concentration edge of the phase envelope (white dotted line, **Fig. 3.3B**). The area to the right of this contour indicates bulk concentrations of N-protein and gRNA that lead to a majority of volume fraction per cluster occupied by N-protein. This region includes almost all of the phase separating regime and likely captures the most relevant stoichiometries of gRNA and N-protein during infection in host cells and virion assembly (127).

During virion assembly, single genomes must be packaged within a capsid built of structural proteins and N-protein (134), so we also quantified how many gRNA chains were in each phase separated cluster. We defined a simple metric to quantify single gRNA packaging; the number of gRNA-containing clusters in the system is divided by the total number of gRNA chains. Therefore, the metric is 1 when single-genome packaging is perfect, and approaches 0 as multiple genomes are clustered together. A zeroth-order effect due to excluded volume in the absence of other interactions is shown in **Fig. S3.9A** and demonstrates that single packaging becomes impossible with prohibitively high volume fractions of N-protein or gRNA. In the presence of interactions, our analysis indicates that, for the most part, N-protein phase separation hinders packaging of single gRNAs in clusters (**Fig. 3.3C**). However, at higher N-protein volume fractions in the concave region below the phase envelope, single packaging is robust. It appears that coexisting phases of N-protein condensed on gRNA and a dilute phase in solution promotes clustering of gRNA, but that a more uniform, high density of N-protein throughout the system effectively keeps gRNA chains separated.

If arranged linearly, the ~30kb genome has an end-to-end length of roughly 10000nm. However, during virion assembly, this genome must be packaged into a viral particle with a diameter of about 100nm, representing an immense compaction challenge. We reasoned that N-protein binding could provide a simple mechanism for gRNA compaction. To quantify compaction, we measured the average radius of gyration of gRNA in each simulation (135),

$$R_g^2 = \frac{1}{N}\sum_{k=1}^{N}(\vec{r}_k - \vec{r}_{\text{mean}})^2$$

where $N = 180$ is the number of monomers in a chain, $\vec{r}_k$ is the position of monomer k, and $\vec{r}_{\text{mean}}$ is the average position of monomers in the chain. R_g is computed for each gRNA in a simulation, and the average across gRNAs, $\langle R_g \rangle$, is reported for each simulation. Again, there is an effect due solely to excluded-volume interactions that leads to compaction of gRNA chains under more crowded conditions (**Fig. S3.9**). Therefore, we evaluated the effects of N-protein binding interactions using the fold change in the radius of gyration in simulations with interactions relative to those without interactions, $\log_2\left(\langle R_g \rangle / \langle R_g^0 \rangle\right)$. We found that N-protein binding indeed leads to more compact genomes to the right of the equal protein-gRNA volume fraction contour (dotted white line), and more extended genomes to the left of it (**Fig. 3.3D**). As seen with the single genome packaging metric, the most robust compaction occurs in the concave region of the phase diagram with a high N-protein density (**Fig. 3.3D**).

Genome cyclization is important for replication of many RNA viruses (136), and there is recent *in vivo* evidence of cyclization of the SARS-CoV-2 genome (133). We thus sought to characterize the potential role of phase behavior in genome cyclization in our model. We hypothesized that the similar RNA features at the 5′ and 3′ ends of the genome would promote cyclization when facilitating phase separation of N-protein. We defined a cyclized genome as one whose terminal beads occupy adjacent lattice sites. Averaging over the final half of each simulation, we quantified the percentage of genomes that met this criterion. Strikingly, we see that up to 40-50% of genomes are cyclized for volume fractions to the right of the equal N-protein gRNA volume fraction contour and within the phase boundary (**Fig. 3.3E**). Thus, it appears that sufficient binding of N-protein is essential for cyclization of gRNA molecules, with phase separation providing additional efficiency.

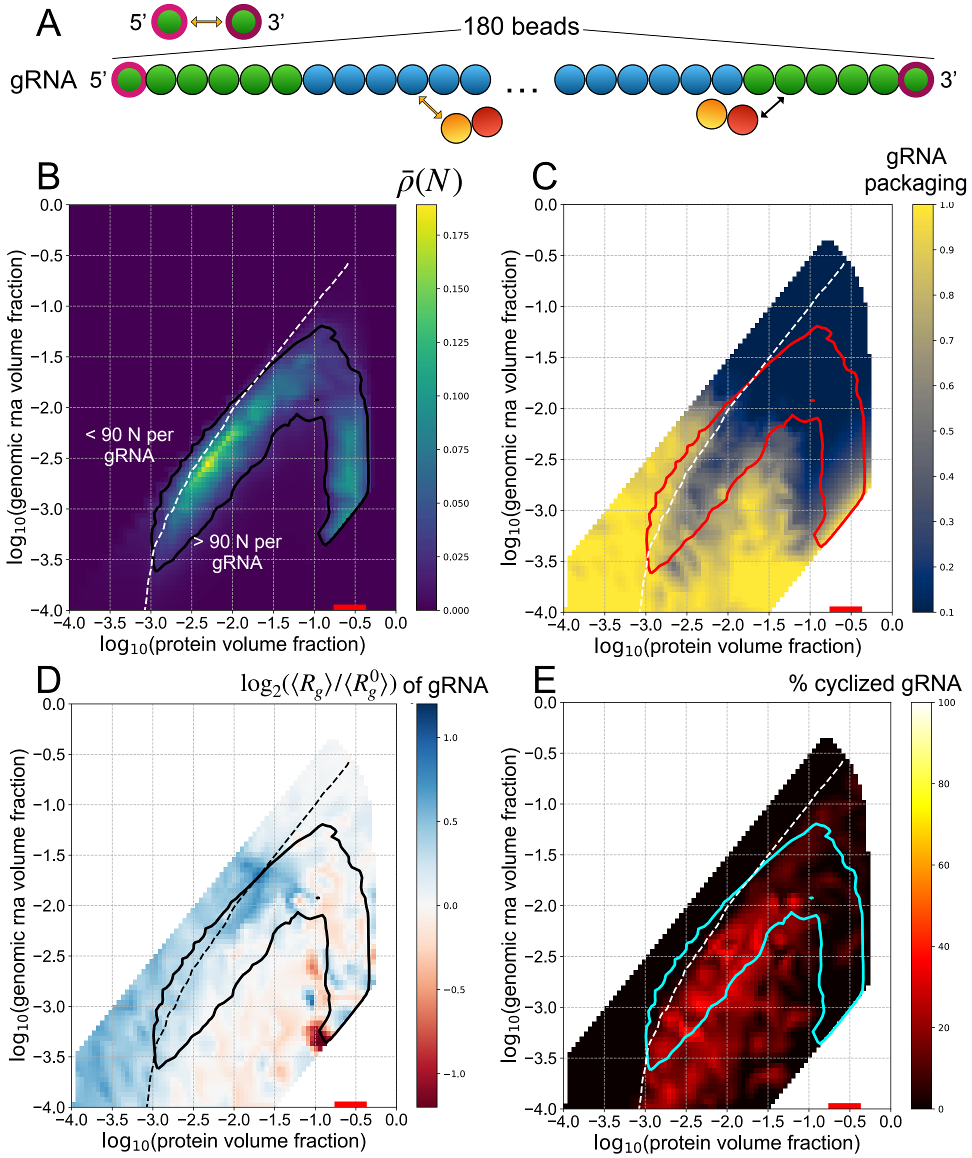


Figure 3.3: gRNA phase separates with N-protein which limits single genome packaging but promotes genome compaction and cyclization. (A) Genomic RNA (gRNA) is represented as a chain with 180 beads. The terminal 6 beads on the 5′ and 3′ end are 5′ end-like beads, and the rest are FE-like beads. An additional anisotropic interaction among the terminal beads is added to represent known nucleotide complementarity. (B) The phase boundary is drawn along the contour $\bar{\rho}(N) = 0.02$ and is shown in black. The heatmap indicates the value of $\bar{\rho}(N)$ at different volume fractions of N-protein and gRNA. The white dotted line indicates the contour along which an equal volume fraction of N-protein and gRNA is found within a given cluster, i.e. 90 N-proteins for each gRNA. A heatmap indicating the average number of proteins bound to each gRNA is shown in **Fig. S3.10**. This contour is shown in all subsequent panels. (C) The phase boundary is shown in red. The heatmap shows the single gRNA packaging metric, which is the number of clusters containing gRNA divided by the total number of gRNA chains in each simulation. A value of 1 represents perfect single-genome packaging. This metric is shown for an analogous system without attractive interactions in **Fig. S3.9A**. (D) The phase boundary is shown in black. The heatmap shows the fold change of the average radius of gyration of gRNA chains, $\langle R_g\rangle$, over the average radius of gyration of gRNA in a system without attractive interactions, $\langle R_g^0\rangle$. A phase diagram with a heatmap of $\langle R_g^0\rangle$ is shown in **Fig. S3.9B**. (E) The phase boundary is shown in cyan. The heatmap shows the percentage of cyclized genomes. Each gRNA chain is categorized as cyclized if its terminal beads are in adjacent lattice positions.

We also performed parameter sweeps for this gRNA-containing system, focusing on the two most sensitive interactions identified in the N-protein with 5′ end and FE RNA parameter sweeps; these are the isotropic interaction between N-protein and 5′ end RNA and the isotropic interaction between N-proteins. With an increased binding energy between N-protein and the 5′ end RNA beads in the gRNA, we saw a modest extension of the phase boundary (**Fig. S3.7A**). Single genome packaging (**Fig. S3.7B**) and compaction (**Fig. S3.7C**) are relatively unaffected with respect to WT, but surprisingly, cyclization is much less efficient (**Fig. S3.7D**). It appears that while N-protein binding to the ends of gRNA chains can facilitate their co-location, very strong binding may limit the ability of the ends of the gRNA chains to come into contact due to interference from N-protein. For the system with increased isotropic interactions among N-proteins, we observed a widening of the phase boundary corresponding to the broader range of volume fractions over which N-protein alone phase separates under this condition (**Fig. S3.8**). Single genome packaging is unaffected (**Fig. S3.8B**), compaction is minimally enhanced in the concave region below the phase envelope (**Fig. S3.8C**), and genome cyclization is unaffected (**Fig. S3.8D**).

Maximal 5′ end-like RNA content in gRNA chains leads to optimal packaging

Our previous work experimentally examined relatively small regions of the genome with regard to the ability to promote N-protein phase separation. While it was clear that the 5’ and 3’ ends of the genome were both highly structured and this was associated with LLPS-promoting activity, it was less definitive that most of the interior of the genome was most similar to FE. Therefore, we next asked how the total LLPS-promoting content vs. the solubilizing RNA content altered N-protein phase separation and genome packaging. We created gRNA chains composed entirely of either 5′ end-like beads or FE-like beads (**Fig. 3.4A**). We found that a genome of purely 5′ end-like beads led to greatly enhanced phase separation of N-protein (**Fig. 3.4B**), robust gRNA compaction below the equal volume fraction contour (**Fig. 3.4C**), and nearly perfect genome cyclization below the equal volume fraction contour and within the phase envelope (**Fig. 3.4D**). In contrast, gRNA composed entirely of FE beads limited single packaging at high volume fractions of N-protein and gRNA (**Fig. 3.4B**), led to gRNA expansion for most conditions (**Fig. 3.4C**), and did not promote genome cyclization (**Fig. 3.4D**). It appears that maximal content of LLPS-promoting sequence is ideal for packaging gRNA by N-protein alone. However, the genome must encode many features other than optimal binding with N-protein and is thus constrained to have a limited amount of LLPS-promoting sequence. Given a limited supply of LLPS-promoting sequence, we next asked what the optimal spatial patterning would be to promote packaging.

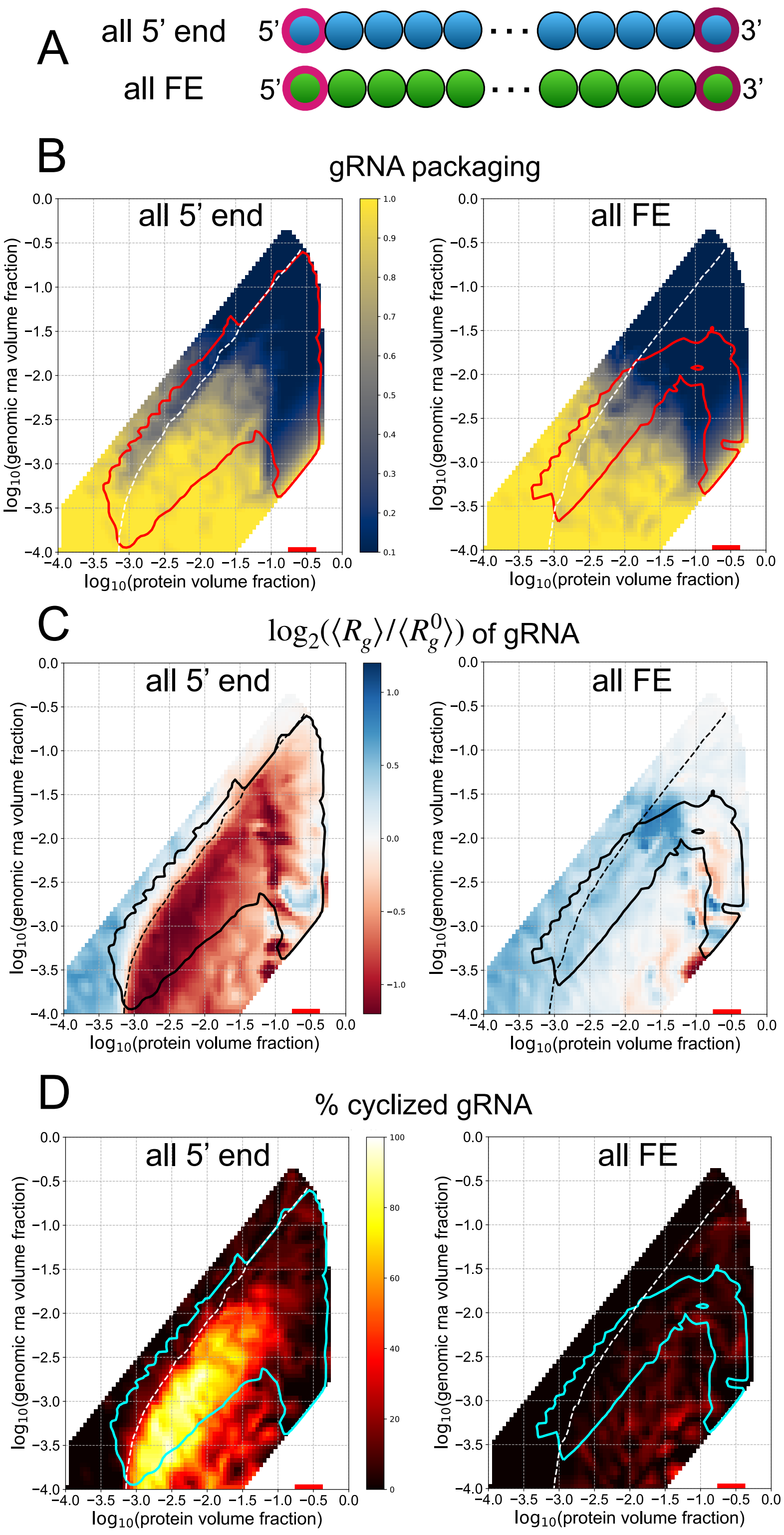


Figure 3.4: LLPS-promoting sites enhance genome packaging. (A) Schematic showing gRNA chains with only 5′ end-like beads and only FE-like beads. (B) The phase boundaries are shown in red. Heatmaps show the single genome packaging metric for each gRNA mutant. (C) The phase boundaries are shown in black. Heatmaps show the fold-change in radius of gyration for each gRNA mutant. (D) Phase boundaries are shown in cyan. Heatmaps show the genome cyclization metric for each gRNA mutant.

Spatial patterning gRNA mutants can enhance single genome packaging but limit compaction and prevent genome cyclization

Since the 5′ and 3′ ends of the gRNA were both found to promote phase separation with N-protein, we hypothesized that this spatial arrangement of phase separation-promoting elements at the ends of the genome may be relevant to packaging. We investigated the importance of the arrangement of phase separation-promoting sequences on the ends of the gRNA by designing mutants where these regions are rearranged. We created three mutants, all of which retain 12 5′ end-like beads and the anisotropic interaction among their terminal beads. The 5′ end-like beads are repositioned either in the middle of the genome (middle), uniformly throughout (uniform), or on one end (end) (**Fig. 3.5A**). The phase boundaries of each of the mutants remain relatively unchanged with respect to the WT system, suggesting that the total amount and not the spatial patterning of 5′ end and FE RNA beads determines the bulk volume fractions at which N-protein phase separation occurs (**Fig. 3.5**). However, we observed differences in the genome packaging metrics relative to WT and among the mutants.

The uniform and middle gRNA systems are more efficient than WT at packaging single genomes into clusters, and end gRNA behaves similarly to WT (**Fig. 3.5B**). Thus, it appears that dispersed or centrally located phase separating elements within the gRNA are preferred for single genome packaging by N-protein alone. However, uniform gRNA does not significantly compact upon phase separation, while middle and end gRNA compact similarly to WT, suggesting that sufficiently clustered LLPS-promoting sequences are important for compaction (**Fig. 3.5C**).

Given the positioning of the wildtype LLPS-promoting sequences at the ends of the genome, we postulated that rearrangement of the location of these sequences would have the strongest impact on genome cyclization. To this end, we also quantified genome cyclization for these mutants, and found that none of them were able to cyclize genomes (**Fig. 3.5D**). Importantly, each of these systems maintains an intrinsic bonding capability between the terminal beads of its gRNA chains. However, since the chains are so large, they cannot efficiently locate each other during the course of the simulation. Thus, the localization of phase separation to the 5′ and 3′ ends with N-protein in the WT system is necessary for the positioning of the genome ends for binding and cyclization (**Figs. 3.3E**, **3.5D**).

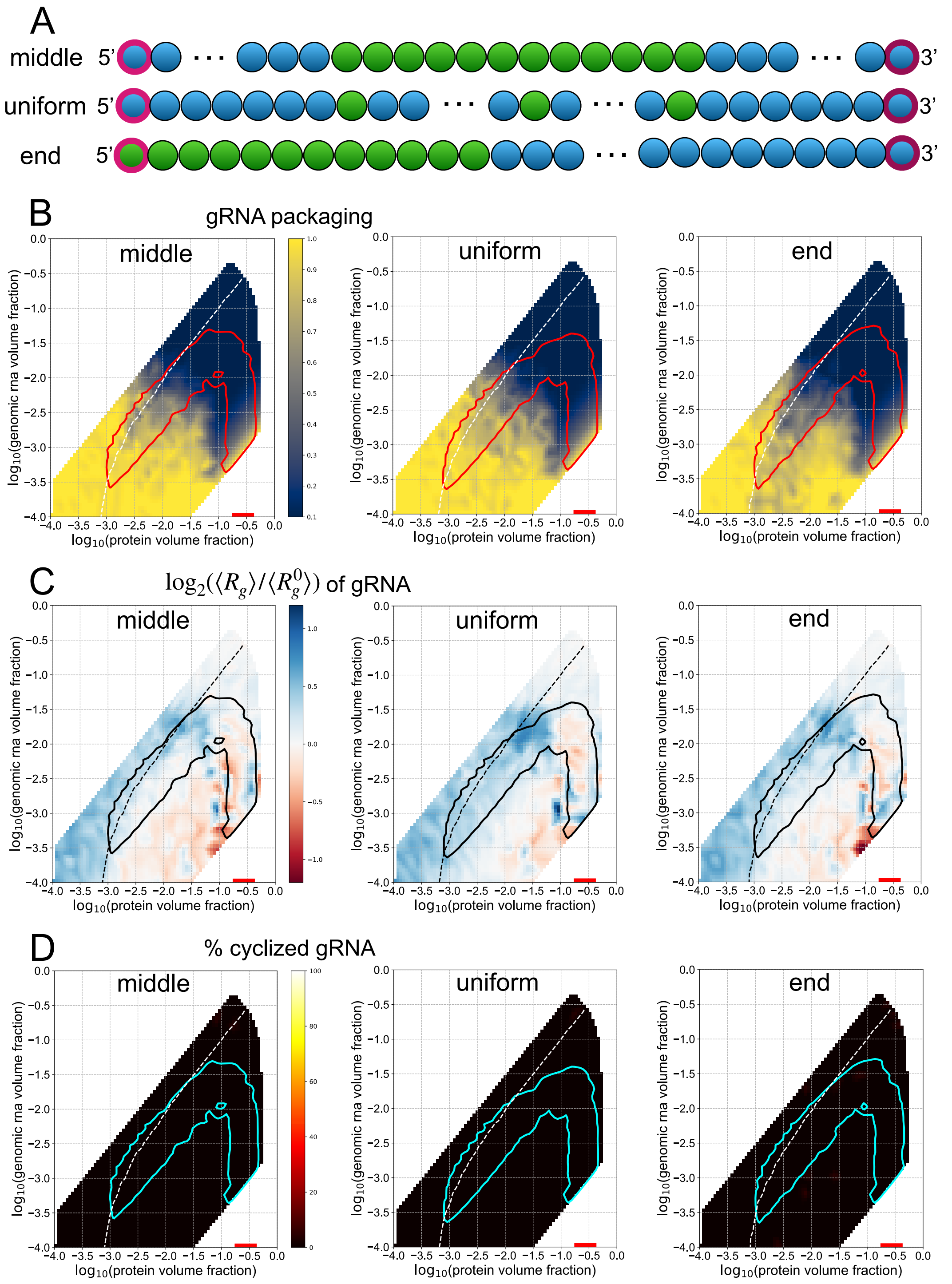


Figure 3.5: Spatial patterning gRNA mutants show altered genome packaging metrics. (A) Three spatial patterning genome mutants were constructed. Each mutant retains the same number of 5′ end-like and FE-like beads as in WT. The terminal beads maintain their anisotropic interaction, regardless of their identity as 5′ end-like or FE-like. (B) Phase boundaries are shown in red for each spatial patterning mutant system. The white dotted line indicates the contour along which an equal volume fraction of N-protein and gRNA is found within a given cluster and is included in all subsequent panels. The heatmaps show the single gRNA packaging metric. (C) Phase boundaries are shown in black. The heatmaps show the fold change in the average radius of gyration of gRNA chains. (D) Phase boundaries are shown in cyan. The heatmaps show the percentage of cyclized genomes.

Optimal gRNA design with limited LLPS-promoting sequence

We know from the studies above that increased 5′ end-like RNA content enhances all packaging metrics. With a limited supply, it appears that positioning of these beads at the ends of genome is essential for cyclization, clustered beads are important for compaction, and uniformly spaced or centrally located beads can promote single packaging. With this understanding, we hypothesized that a genome could evolve that would optimally function according to all of these metrics, given a limited supply of LLPS-promoting beads. We designed a gRNA which has the WT arrangement at the ends, with 12 additional groups of 3 5′ end-like beads uniformly spaced throughout its length (**Fig. 3.6A**). We found that N-protein phase separation occurs over a broader range of concentrations, and that all studied genome packaging metrics are enhanced, relative to WT. Single genome packaging is more preferable within the concave region below the phase envelope (**Fig. 3.6B**). Genome compaction is greatly enhanced below the equal N-protein gRNA volume fraction contour (**Fig. 3.6C**). Genome cyclization is also enhanced (**Fig. 3.6D**).

For this optimized gRNA design, there appears to be a concentration regime corresponding to the concave region below the phase envelope where single genome packaging, compaction, and cyclization can all occur efficiently. These results raise two predictions concerning the SARS-CoV-2 virion assembly process. First, there exists an optimal concentration range of gRNA and N-protein that promotes virion assembly. Indeed, a sufficiently high protein/rna ratio has been shown to be necessary for self-assembly of viral capsids *in vitro* (137). Second, phase-separation promoting gRNA sequences may be located not only at the 5′ and 3′ ends, but also distributed in clusters throughout the genome to enhance genome packaging.

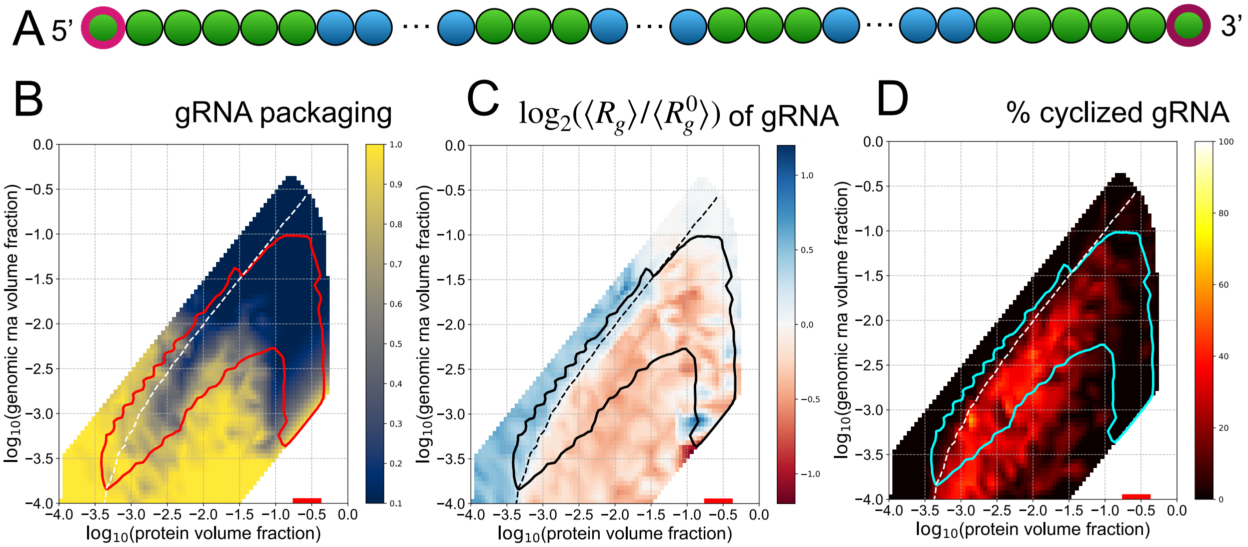


Figure 3.6: WT and uniform elements optimize genome packaging when combined. (A) The optimal gRNA pattern for singular and compact packaging has 6 5′ end-like beads at each end, with 12 additional groups of 3 5′ end-like beads distributed uniformly throughout the rest of the chain. (B) The phase boundary is shown in red. The white dotted line indicates the contour along which an equal volume fraction of N-protein and gRNA is found within a given cluster and is included in all subsequent panels. The heatmap shows the single gRNA packaging metric. (C) The phase boundary is shown in black. The heatmap shows the fold change of the average radius of gyration of gRNA chains relative to a system with only excluded volume interactions. (D) The phase boundary is shown in cyan. The heatmap shows the percentage of cyclized genomes.

Discussion

It is clear that protein and RNA elements of SARS-CoV-2 can engage in phase separation with N-protein (30, 42, 128-130), but the functional consequences of this physical chemistry capacity for viral replication remain elusive. In this study we sought to explore how the spatial patterning of phase separation-promoting or -inhibiting RNA elements in the genome could facilitate the specificity and singularity of packaging the genome. Using coarse-grained simulations rooted in empirical observations, we find that single genome packaging is most efficient when binding sites are centrally located or distributed throughout the genome. However, the arrangement of phase separation promoting sequences in clusters is critical for genome compaction, and the positioning of these elements at both ends is necessary for cyclization.

Robustness of the model

The model of N-protein interactions with the gRNA fragments, 5′ end and FE RNA, was constructed to recapitulate several features of the experimentally characterized system. Importantly, we recover the following features: N-protein dimerizes and phase separates weakly on its own (**Fig. 3.1B**); N-protein phase separation is enhanced when mixed with 5′ end RNA (**Fig. 3.1C**); N-protein is dissolved when mixed with sufficient amounts of FE RNA (**Fig. 3.1D**); the binding patterns of N-protein on both 5′ end and FE RNA in the dilute phase are qualitatively consistent with those measured experimentally (**Fig. 3.2**). These behaviors arise because of the choice of the class of interactions among components, either isotropic or anisotropic. A key prediction of this model is that 5′ end RNA promotes co-phase separation with protein using isotropic interactions by allowing for large-scale interaction networks to form, while FE RNA abrogates phase separation by clustering N-protein into small, inert protein-RNA complexes via anisotropic interactions. These choices led to phase behavior and dilute phase properties that were consistent with experimental results and are a central prediction of this work. We performed a wide range of parameter sweeps over the binding energies for each of these interactions in **Figs. S3.2-3.4** and found that the important features recovered are robust to changes in magnitude of binding energies; rather, the class of interaction (isotropic versus anisotropic) is what defines the behavior in these systems.

Biophysical interpretation of the different protein binding modes leading to opposing phase behavior

An essential feature of our model is the different binding mode of N-protein with 5′ end and FE RNA. We found that weak, isotropic interactions with 5′ end RNA promote phase separation, while strong, anisotropic interactions with FE RNA lead to N-protein solubilization. Intrinsically, the valence of the isotropic interactions is 26 in a cubic lattice, while the valence is

1 for the anisotropic interactions. Additionally, since the anisotropic interactions have a higher binding energy than the isotropic ones, they last for longer, which effectively compounds the difference in valence between the 5′ end and FE beads. For the FE RNA system, this design allows for more N-protein binding at lower protein concentrations, but a less significant increase in protein binding as the protein concentration is increased. On the other hand, 5′ end RNA experiences a cooperative binding effect with N-protein, leading to greatly increased protein binding as protein concentrations are increased. The valence difference between 5′ end and FE RNA also includes competition between N-protein-FE RNA binding and N-protein dimerization, since a single bead in the model is responsible for both interactions. Therefore, while N-protein is bound to FE RNA, it can no longer dimerize, but while it is bound to 5′ end RNA, it is free to dimerize. We suspect that this competition also compounds the cooperative effect of the N-protein binding to each of these different RNAs and contributes to the distinct behaviors of these different polymers. The agreement between experimental results and our model suggests that there may be some underlying differences in the N-protein-RNA interactions in the 5′ end-like and FE-like regions in the gRNA that lead to distinct protein binding behavior. We hypothesize that FE-like regions have different N-protein binding kinetics than 5′ end-like regions, which will need to be explored experimentally in future work.

Boomerang shape of the phase boundaries

It has been previously reported that phase boundaries in ternary systems that are purely driven by heterotypic interactions are roughly elliptical in log-log space (40). However, the phase diagrams of the ternary systems studied in this work have distinct shapes. The 5′ end RNA with N-protein phase diagram resembles an elliptical shape combined with a high N-protein concentration arm that corresponds well to the shape of the FE RNA with N-protein phase diagram. Each of these high N-protein concentration arms correspond to the N-protein concentration range where phase separation occurs for N-protein alone at temperature 1 a.u., the temperature at which all ternary systems were studied (**Figs. 3.1B-D**). The same superimposed shape is even more apparent in the gRNA with N-protein phase diagrams, with N-protein phase boundaries that resemble boomerangs (**Figs. 3.3B-E**). We hypothesize that these phase diagrams arise from the union of two regimes of phase behavior that are driven by distinct forces. The elliptical portion that extends to low N-protein and RNA concentrations demonstrates phase separation that is driven by heterotypic interactions between RNA and N-protein, which aligns with reports in (40). On the other hand, the high concentration N-protein portion of the phase diagrams indicates phase separation that is driven by N-protein homotypic interactions. Here, RNA can partition into the dense phase, but it is neither necessary for phase separation, nor is it the driver. Therefore, the FE RNA with N-protein phase diagram consists only of this high N-protein concentration regime, which is eventually capped at high enough RNA concentrations (**Fig. 3.1D**). As more FE RNA is present to sequester N-protein out of solution, the concentration of the available pool of N-protein is effectively decreased and phase separation can no longer occur. For systems with 5′ end RNA present, the phase diagrams have a complex reentrant character, passing in and out of the phase separating regime for certain fixed RNA concentrations as N-protein concentration is changed (**Figs. 3.1C, 3.3B-E**). We speculate that this reentrant behavior is due to the interplay of phase separation driven by heterotypic or homotypic interactions, with intermediate regimes where neither are strong enough to drive phase separation. It is interesting to speculate whether such rich phase behavior may exist in other multicomponent systems, specifically those that involve long nucleotide polymers and RNA- or DNA-binding proteins.

Compatibility with the single packaging signal model

A recent model of single genome packaging has emerged as part of a study of the SARS-CoV-2 N-protein (30). Results from Cubuk et al. suggest genomes with a single packaging signal are much more efficiently packaged, and multiple viruses such as HIV employ this strategy. Our results confirm this finding in that restriction of LLPS promoting elements to the ends of the genome (**Fig. 3.5**) is not as efficient as a single central packaging signal and multiple peppered LLPS promoting elements in the center (**Fig. 3.6**). Cubuk et al showed that for a 2-bead representation of N-protein that experienced isotropic attraction to itself and a long, 61-bead RNA molecule, large phase-separated clusters would form. However, if a much stronger binding site was added to the center of the RNA chain, mimicking a hypothetical packaging signal, N-protein and RNA chains would instead form kinetically trapped clusters that only very slowly coalesced into a single phase-separated droplet. This model is simpler than the one presented here, but it is reminiscent of the distinct effects seen here between 5′ end RNA and FE RNA. In our model, the most important effective difference between the 5′ end and FE RNA beads is their valence, with 5′ end RNA having a much higher valence than FE RNA. The model in (30) includes only a difference in binding energy, since all beads interact isotropically. However, the higher binding energy beads have a lower effective valence than the low energy beads at a given timescale, since they participate in bonds with fewer partners. In line with our results, the presence of lower-valence binding sites sequesters N-protein into clusters, opposing large-scale phase separation. Indeed, other groups have shown that a high enough valence is required for phase separation to occur *in silico* (24) and *in vitro* (19). In our model, however, most of the binding sites on the gRNA are of a low-valence character, which is distinct from the single, low-valence, packaging signal site in (30). Despite this difference, both models provide evidence that such low-valence sites are essential for packaging tasks required during virion assembly, and that runaway phase separation must be tempered via alternative self-assembly pathways.

Why would SARS-CoV-2 and other betacoronaviruses use a relatively inefficient packaging methodology with LLPS? One possible reason could be that packaging and condensation may be acting in direct competition with other viral processes such as translation, and N-protein LLPS may block or slow ribosomal read-through. In line with this hypothesis, Fmr1 LLPS has previously been demonstrated to repress translation of co-condensing RNA (60). By restricting LLPS promoting elements to the 5' and 3' ends of the genome, SARS-CoV-2 could allow for efficient packaging while ensuring viral protein production can proceed unencumbered.

The phase behavior of systems with very short and very long chains

For our systems with N-protein and gRNA, we never observed a dilute phase of gRNA and instead saw that the dense phase of N-protein was condensed on gRNA chains. We hypothesize that this effect is due to the large size of gRNA and their low abundance (10 chains in simulations). The Flory-Huggins free energy (115) predicts that as polymer length increases, the dilute phase volume fraction of that polymer decreases. As chain length increases in a fixed volume, the expected dilute phase volume fraction will eventually pass below the volume fraction occupied by a single chain, leading to disappearance of the dilute phase. When a small number of such large chains are combined with many thousands of shorter chains, an interesting blend of properties can arise. The shorter chain may still exhibit thermodynamically well-defined phase separation, although the dense phase now occurs on the long chains which effectively become a surface upon which phase separation is favored. This situation may be a relatively common one in biology. For example, within a host cell, the number of SARS-CoV-2 gRNA chains in a volume relevant for N-protein phase separation may not be much more than the 10 studied here (126). The difference in scale between thousands of proteins and dozens of very large nucleic acid polymers presents challenges to existing physical frameworks for ternary phase separation and will require theoretical innovations to rigorously understand.

Generalization to other viruses and systems with long RNAs or DNA

Our models developed here are sufficiently coarse-grained to speculate that they may be applicable to other viruses and systems that involve long nucleotide chains and proteins. Components from several viruses have been shown to undergo phase separation, raising the possibility that spatial patterning of specific LLPS-promoting RNA or DNA sequences may have evolved to promote optimal genome packaging in other viruses in addition to SARS-CoV-2.

Many cellular phase separated bodies involve long RNAs or DNA and proteins that bind them. A particularly relevant example for our modeling is paraspeckles. Paraspeckles are highly ordered, condensed nuclear bodies that require the presence of the long noncoding RNA, NEAT1. NEAT1 was recently shown to contain distinct functional domains, one of which is repetitive in its sequence and is necessary for paraspeckle formation (138). Paraspeckles also require several proteins, most of which contain RNA-binding domains and disordered regions (139). Recent work has shown that the central region of NEAT1 is necessary and sufficient for paraspeckle formation, and that it initiates assembly by binding several proteins (138). Specifically, the proteins NONO and SPFQ must first bind NEAT1, dimerize, and promote further polymerization via coiled-coil domains with other proteins for paraspeckle assembly to continue (139). There are thus many parallels with the SARS-CoV-2 system studied here; specific spatial patterning of protein binding elements along the RNAs is essential, and the protein partners must be able to dimerize/oligomerize for further assembly.

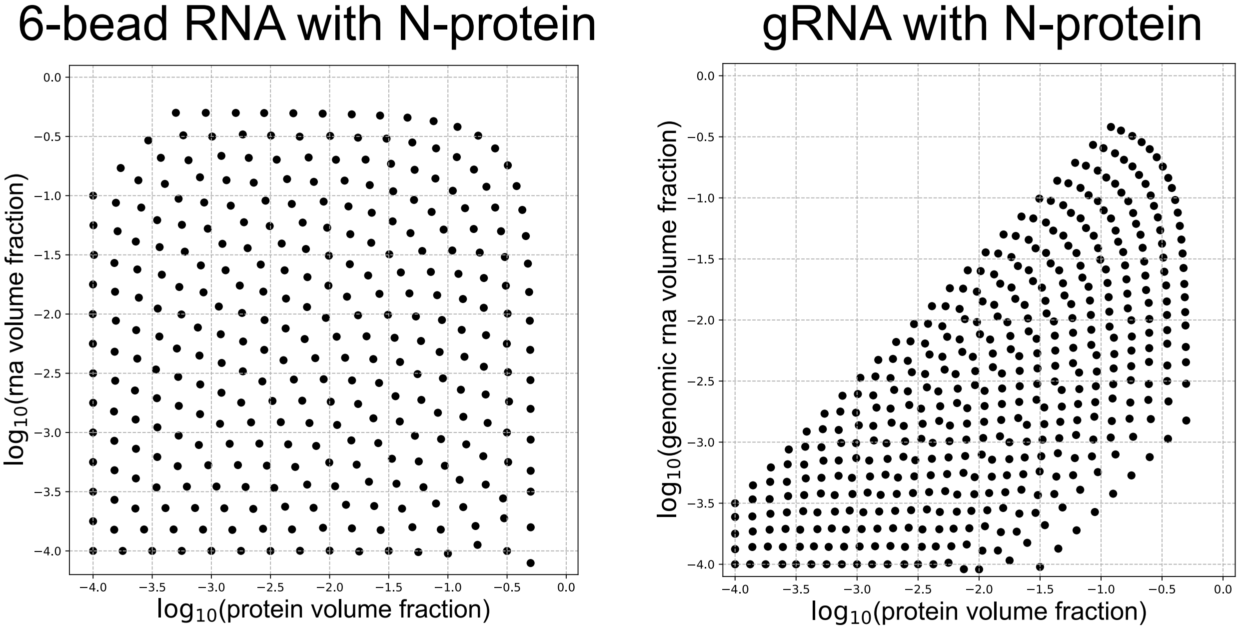
Conclusion

Identification of specific RNA sequences that promote ordered phase separated bodies via protein binding will not only advance our understanding of viruses, but also the many diverse cellular bodies and regions that contain long RNAs or DNA.

Acknowledgements

We thank Mikayla Feldbauer for helpful discussions and code drafting and the Pappu lab for initial introductions to working with LASSI, and Furqan Dar for technical guidance when modifying LASSI. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. The Comet cluster at the San Diego Supercomputer Center was used through allocation nca106. We thank UNC Chapel Hill campus champion Mark Reed for helping I.S. access XSEDE resources on short notice. This work was supported by a FastGrant to ASG and C.A.R. is supported by NIH T32 CA 9156-43, F32GM136164, and L'OREAL USA for Women in Science Fellowship, and I. S. and A. S. G. were supported by NIH R01 GM081506 and the Air Force Office of Scientific Research (grant FA9550-20-1-0241). A. S. G. serves on the scientific advisory board to Dewpoint Therapeutics. The authors declare no conflicts of interest.

Supplemental Figures



6-bead RNA with N-protein

gRNA with N-protein

Figure S3.1: (protein, gRNA) volume fractions used in ternary simulations. 4 simulations were run at each point for respective systems as described in methods. Specifically, a given point represents a certain volume fraction of both the protein and RNA components in a system. The total number of chains for each point is set as specified in the methods, and the number of each protein and RNA chain is set to satisfy the corresponding volume fractions at each point.

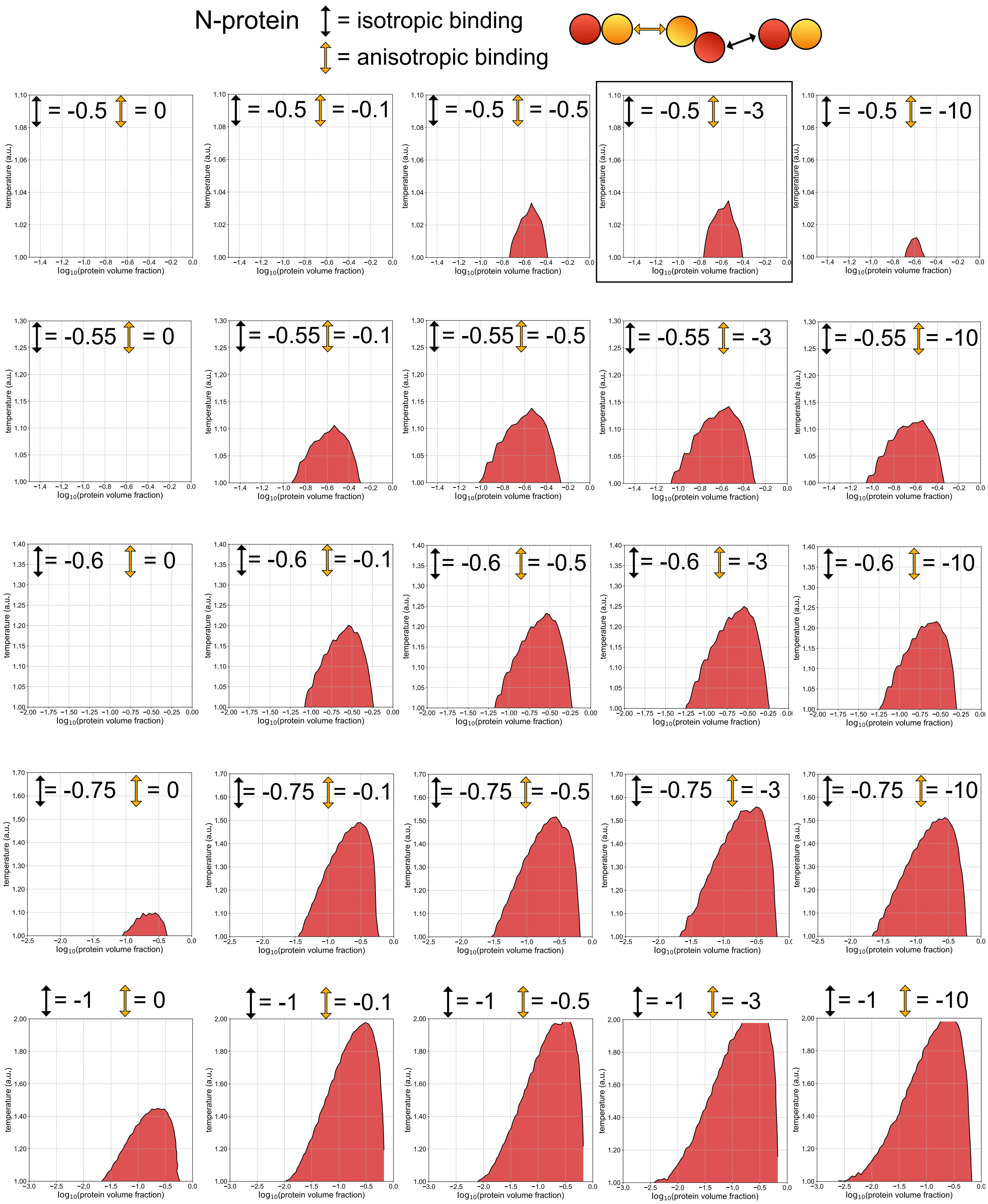


Figure S3.2: Sweeps over protein-protein interaction energies for the protein-only system. An independent set of simulations was performed to generate phase diagrams in each panel shown. Along rows, the isotropic energy is fixed and the anisotropic energy changes, and across columns the anisotropic energy is fixed and the isotropic energy changes. The default N-protein energies are indicated by the black rectangle.

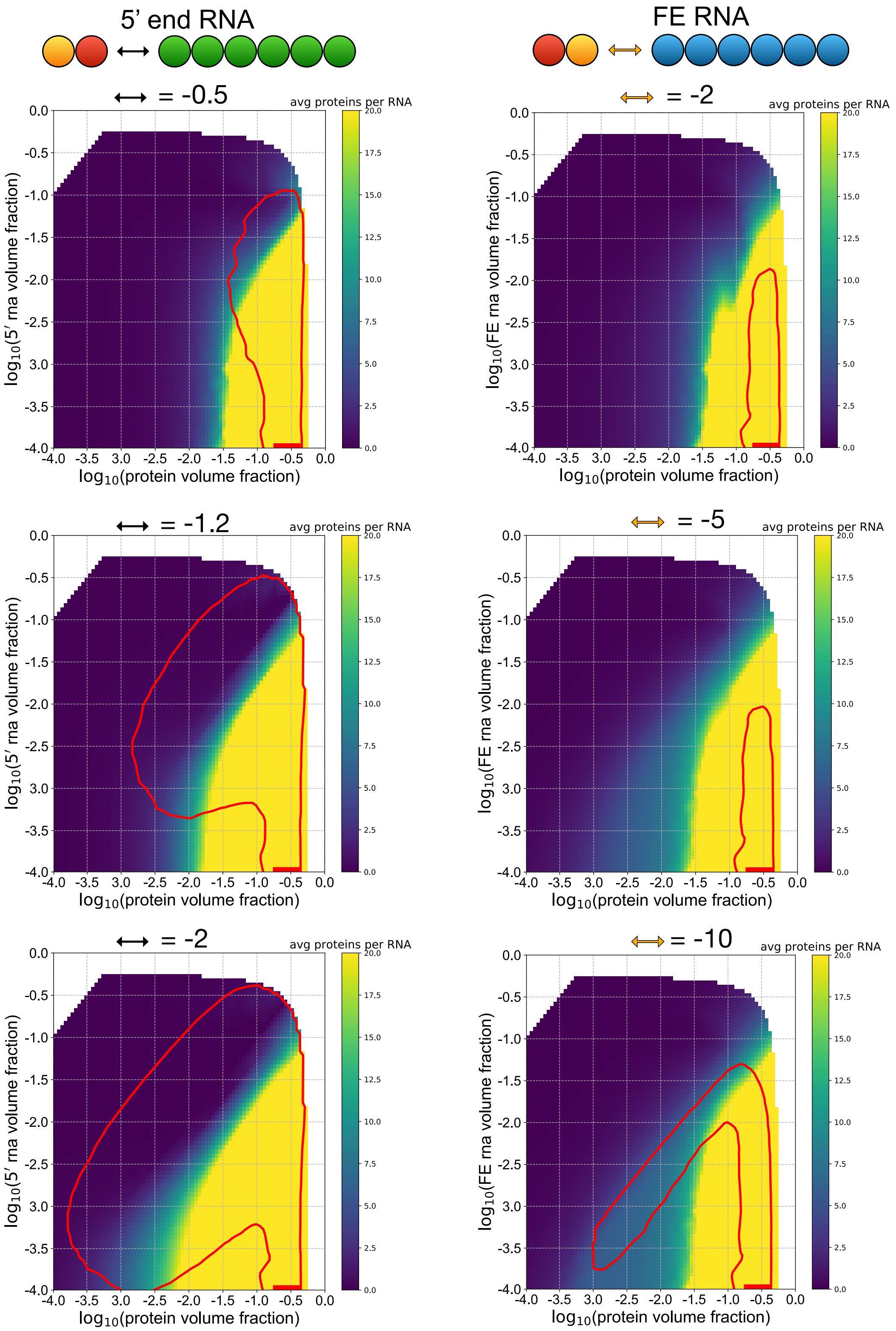


Figure S3.3: Sweeps over interaction energies between N-protein and 5′ end RNA and N-protein and FE RNA. For default N-protein binding energies, the isotropic binding between N-protein and 5′ end RNA is changed along the first column, and the anisotropic binding between N-protein and FE RNA is changed along the second column.

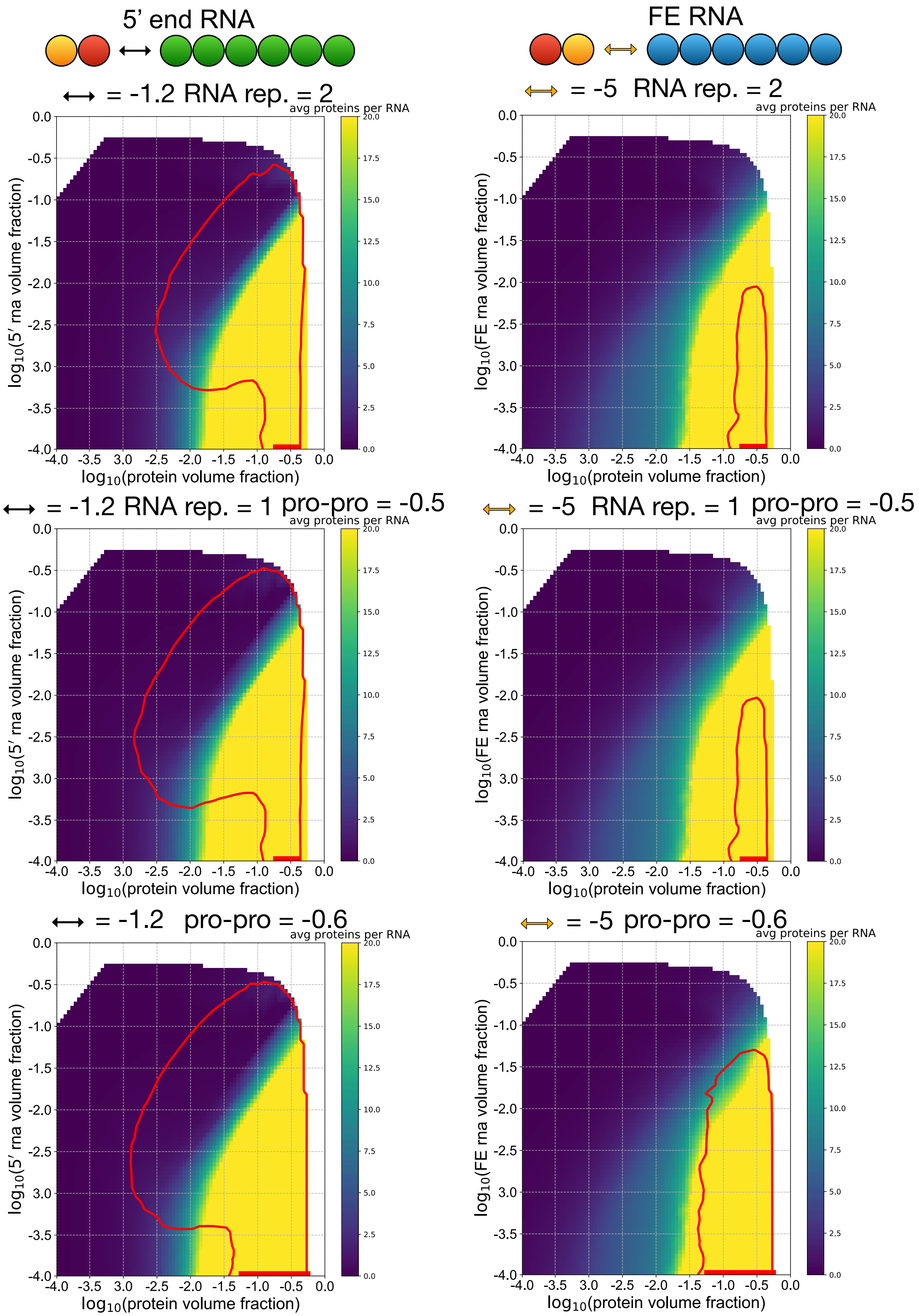


Figure S3.4: Results for altered RNA-RNA repulsion energies and protein-protein isotropic interaction energies. The first row demonstrates the effect of a doubled RNA-RNA isotropic repulsion, with all other parameters set to default values. The second row shows results for default values of all parameters. The third row shows the effect of incorporating N-protein with a higher isotropic self-binding energy.

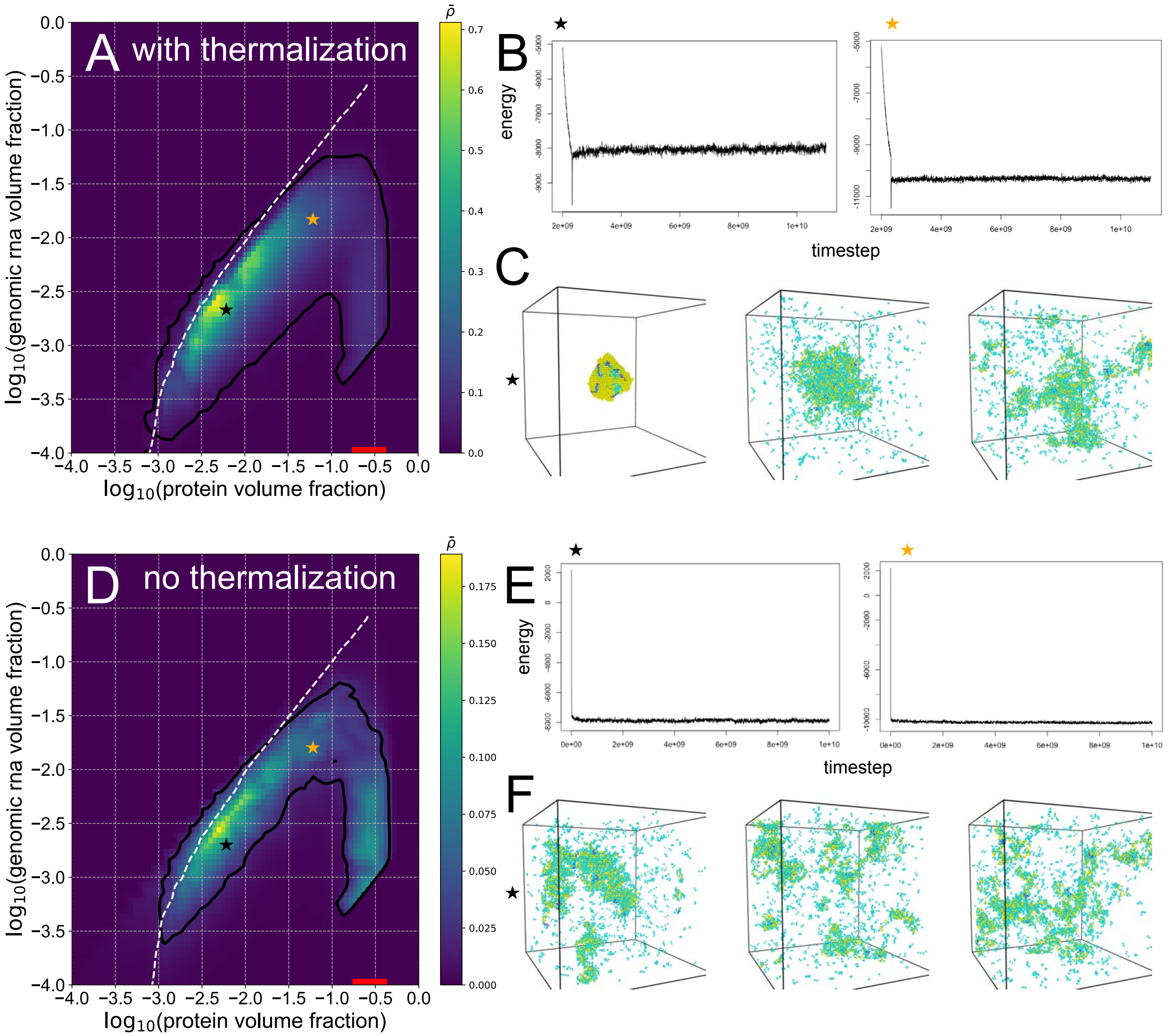


Figure S3.5: Comparison of energies and system conformational states with and without thermalization for the default N-protein-gRNA parameters. (A) Phase diagram for a system run with 1e9 steps of thermalization. (B) Total energy versus timestep plots for two simulations corresponding to black and orange stars in (A). (C) Snapshots of the black star simulation at the beginning during simulation, halfway through, and at the end of the simulation. (D) Phase diagram for the system run with no thermalization. (E) Total energy versus timestep plots for two simulations corresponding to black and orange stars in (D). (F) Snapshots of the black star simulation at the beginning of the simulation, halfway through, and at the end of the simulation.

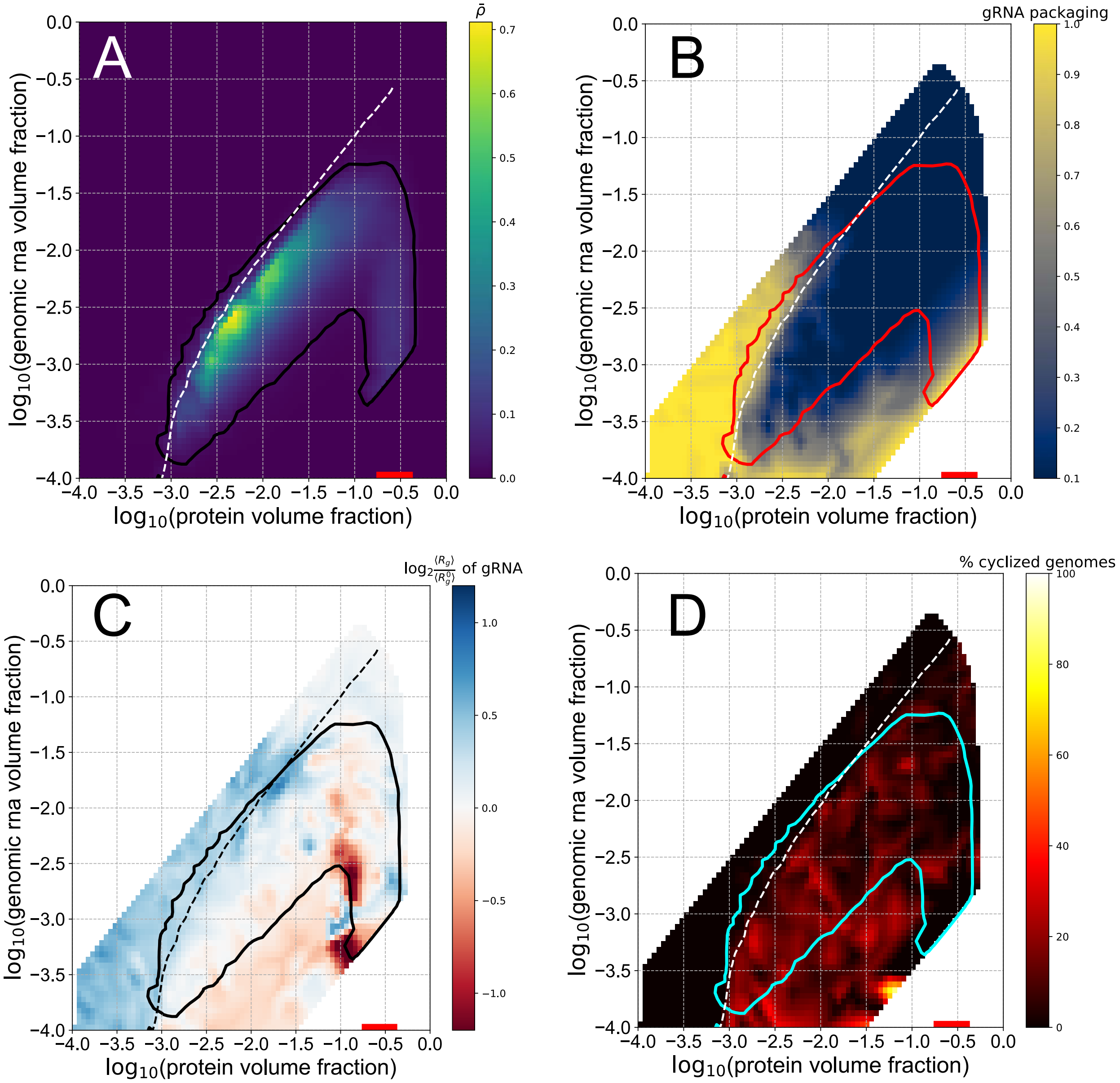


Figure S3.6: Phase behavior of the default N-protein with gRNA system with 1e9 steps thermalization at beginning of simulation. (A) Heatmap shows $\bar{\rho}$ calculated for N-protein – N-protein interactions only, as described in methods. The black line is the contour drawn at $\bar{\rho} = 0.02$, and the white line denotes the bulk volume fractions at which an equal volume fraction of protein is bound to each gRNA molecule, on average. (B) Phase envelope in red overlaid on the single packaging metric. (C) Phase envelope in black overlayed on the fold change in the radius of gyration relative to a system with no interactions. (D) Phase envelope in cyan overlayed on the percentage of cyclized genomes.

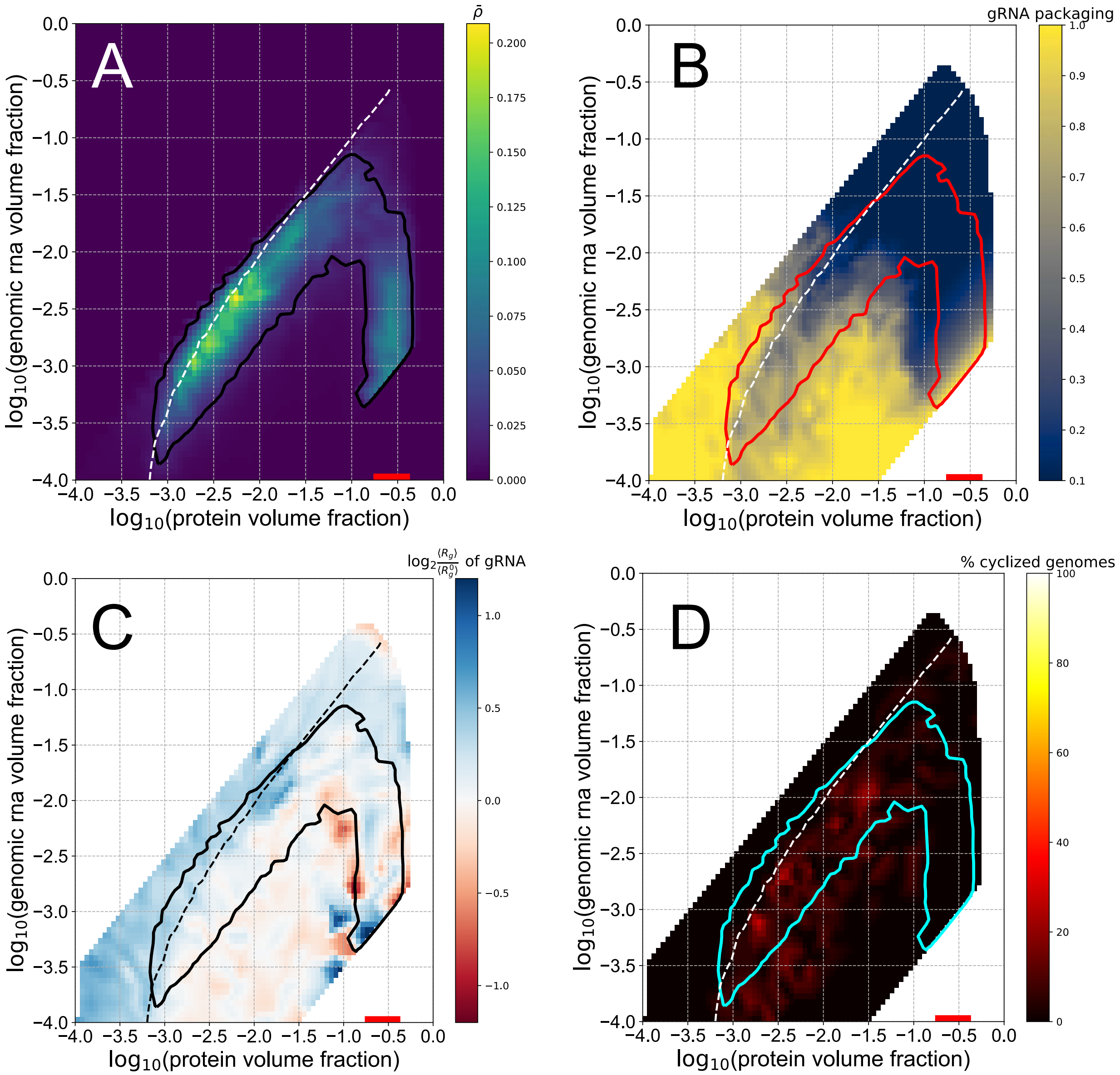


Figure S3.7: Phase behavior of N-protein with gRNA with isotropic binding energy between N-protein and 5′ end RNA in the gRNA increased to -2. (A)-(D) Related to panels in **Fig. S3.6**.

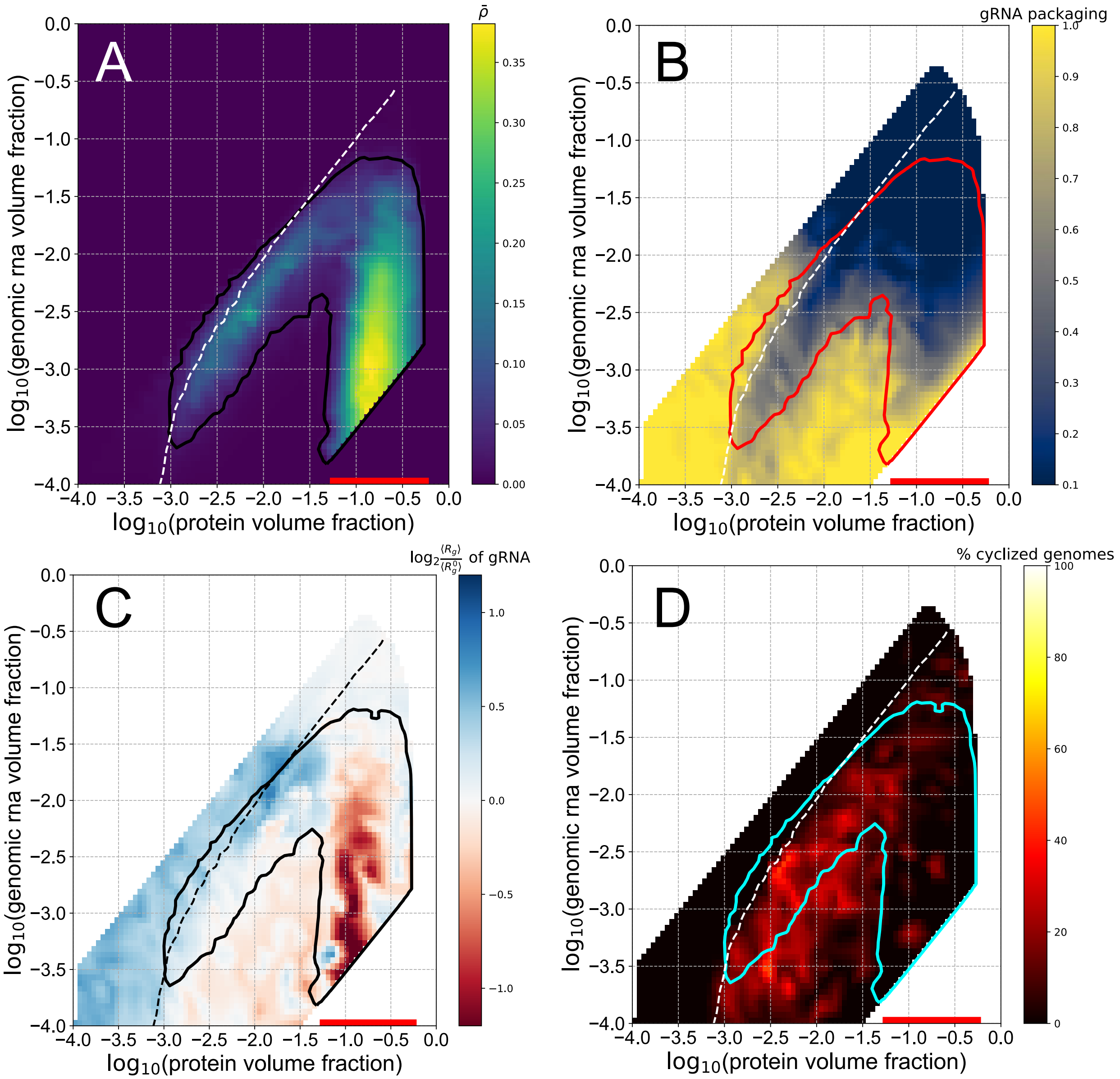


Figure S3.8: Phase behavior of the N-protein with gRNA system with N-protein – N-protein isotropic interactions increased to -0.6. (A)-(D) Related to panels in **Fig. S3.6**.

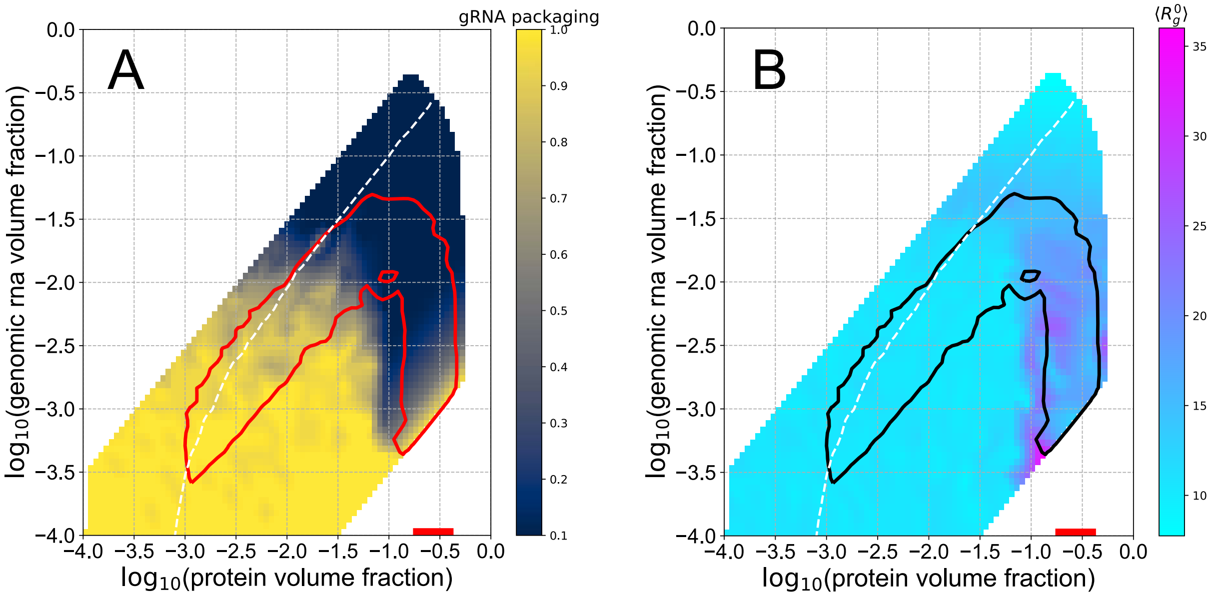


Figure S3.9: gRNA packaging and radius of gyration for simulations with no interaction energies. The only interactions among particles are excluded volume such that only one bead can occupy a lattice site at a given time.

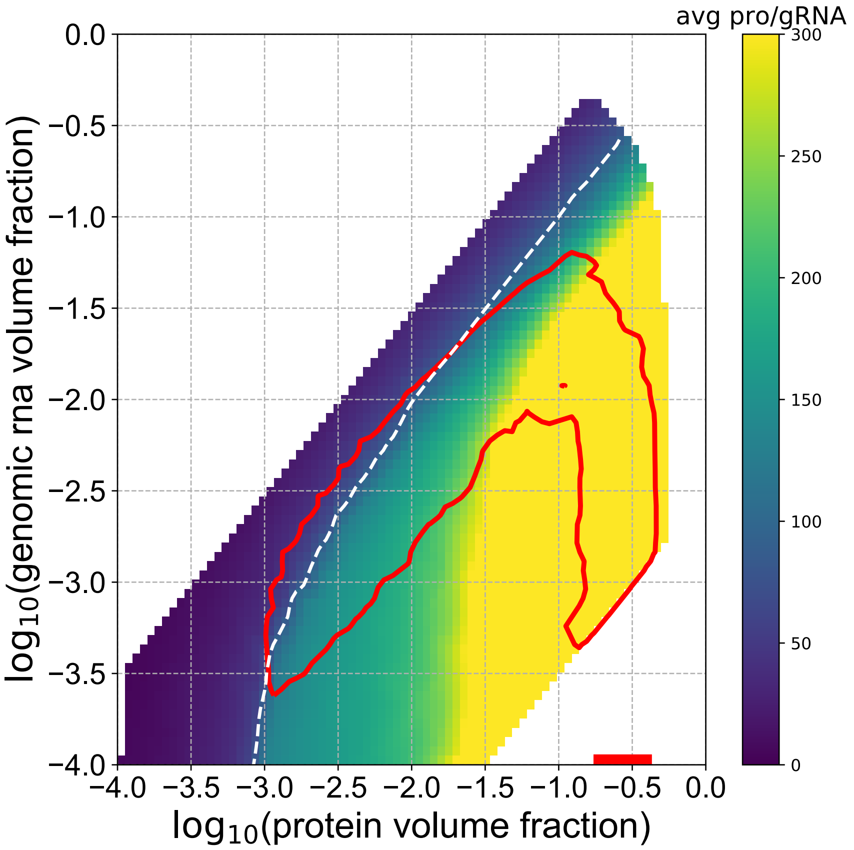


Figure S3.10: Average number of N-proteins bound to each gRNA for the WT system. Phase envelope is drawn in red. The heatmaps indicate the average number of protein chains per gRNA chain for each cluster identified in each simulation. If more than one gRNA chain is in a cluster, the ratio of protein chains to gRNA chains within that cluster is reported.

CHAPTER 4: DISCUSSION

It is clear that phase separation coupled to percolation (PSCP) is a fundamental self-assembly principle for biomolecules. However, biomolecules can have modular architectures with some domains that promote PSCP and others that promote distinct modes of assembly. As changes are made to these domains, the balance between different assembly modes is shifted which can move systems throughout the phase diagram and even across phase boundaries.

In Chapter 2, we showed that a transiently structured helical region within the QRR of Whi3 can be stabilized or destabilized with point mutations designed to enhance or prevent coiled-coil formation, respectively. Stable coiled-coil formation leads to the formation of larger oligomers in the dilute phase which reduces the concentration of protein in the dense phase. On the other hand, prevention of coiled-coil formation reduces the abundance of dilute phase oligomers and increases dense phase concentrations of protein. In this case, the disordered QRR promotes PSCP while the coiled-coil forming domain leads to oligomer formation resulting in clusters that are inert with respect to large-scale network formation as in PSCP. It appears that local order/disorder transitions within the CC domain in Whi3 enable fine tuning of the influence of two opposing yet coexisting modes of self-assembly.

In Chapter 3, we established a theoretical framework for two types of protein-RNA interactions that explains the opposing behaviors of 5′ end RNA and FE RNA with N-protein first observed in (42). Although both types of RNA were modeled as 6-bead chains, their binding to N-protein was sufficiently different to lead to either PSCP or dissolution of N-protein. When each bead in 5′ end RNA interacts with N-protein isotropically and with a low binding affinity, large interaction networks form and PSCP is robust. If instead the beads in the FE RNA interact with N-protein anisotropically, i.e. in a 1-to-1 binding fashion, and with a high binding affinity, stable and inert complexes of N-protein and FE RNA are formed in the dilute phase which minimize the system free energy more effectively than the formation of a macroscopic dense phase. With high enough concentrations of FE RNA, N-protein PSCP can be completely prevented as N-protein is sequestered into small RNP complexes and the available pool of N-protein is diluted below the saturation concentration for N-protein-only PSCP. We next incorporated these opposing RNA elements into 180-bead chains representing the 30kb RNA genome of SARS-CoV-2. Due to the large size difference between the gRNA molecules and N-proteins (90x), we found that it was more appropriate to regard the gRNA molecules as surfaces upon which N-protein PSCP took place rather than only being regarded as another component in multi-component PSCP. There are several implications of this change in viewpoint that will be discussed in a later section of this chapter. The gRNA molecules remain deformable chains, however, and spatial patterning of 5′ end-like and FE-like RNA elements along their contours resulted in alterations to gRNA chain compaction and circularization. Spatial patterning also influenced the extent to which the system free energy was minimized by a dense phase consisting of all gRNA chains or individual complexes of single gRNA chains with adsorbed N-proteins. We found that when PSCP-promoting elements were clustered and these clusters uniformly space throughout the gRNA chain, including the ends, compaction, circularization, and single packaging were optimal.

Both of the studies reviewed above may be understood as explorations of the effects of altered effective valence and effective binding site affinities on PSCP, similar to those discussed and observed in (1, 29, 30). The Whi3 CC-QRR may be understood in the following way. In the Whi3 (CC → CC^-) mutant, in which CC formation is fully prevented using Pro point mutations, the CC-QRR is fully disordered. In the language of stickers and spacers, this region has many dispersed, low-affinity stickers of relatively equal weight. This architecture leads to robust PSCP. However, in the Whi3 (CC → CC^+) mutant, the 21 residues of the CC are stabilized by Leu point mutations which enhance the hydrophobic moment of the resulting coiled-coil forming helix. These residues may be regarded as transforming from the dispersed, low-affinity architecture in the unfolded CC within Whi3 (CC → CC^-) into a single, high-affinity sticker. In this sense, Whi3 (CC → CC^+) has a lower effective valence than Whi3 (CC → CC^-) but gains a single, high-affinity binding site in its CC region. The WT protein is an intermediate between these two extremes. Strikingly, although the CC comprises only a fifth of the ~100 residue QRR within Whi3 which is itself 729 residues, this architectural shift leads to drastically different self-assemblies. The behavior of Whi3 (CC → CC^+) is consistent with the behavior of other sequences modified to have a lower effective valence with higher affinity binding sites as studied in (29). It is interesting to speculate that further enhancement of the CC within Whi3 may result in complete abrogation of PSCP in favor of aggregation, as observed for polymers with sufficiently clustered stickers in (29). Similarly, the clustered stickers within our optimal gRNA designs explored in Chapter 3 are reminiscent of the effects of addition of a single, high-affinity binding site to gRNA chains explored in (30). In both cases, gRNA chains were dynamically arrested as complexes with N-protein containing only single gRNA molecules. If we regard our clustered stickers instead as high-affinity binding sites, these results are consistent.

We have identified a second example of homo-oligomerization opposing phase separation in the N-protein of SARS-CoV-2 (140). N-protein contains a dimerization domain adjacent to or possibly overlapping with its second RNA-binding domain (RBD2). A 32-residue deletion in this domain shifts the very high affinity dimer population to a mostly monomeric one, as measured by mass photometry and shown in Figs. 3H, I of (140). The reduced affinity dimer mutant also undergoes greatly enhanced PSCP, both alone and with RNA, relative to WT N-protein, as shown in Figs. 3D, G in (140). An outstanding question is why does oligomerization oppose PSCP for Whi3 and N-protein, but enhance it for other systems? A particularly puzzling comparison is the CC in Whi3 and the helix within the C-terminal IDR of TDP-43 (74). The ~20 residue helix within TDP-43 mediates dimerization which is enhanced by Ala point mutations. Unlike in Whi3, PSCP of TDP-32 is enhanced when helix formation and dimerization are stabilized (74). The authors report that the binding affinity of the TDP-43 helix-helix dimers is low (~100µM), unlike traditional coiled-coil or leucine zipper interactions which usually have higher affinities. It is therefore possible that low affinity SODs within IDRs can enhance phase separation, as in TDP-32, while high affinity SODs within IDRs can oppose it, as may be the case in Whi3 and N-protein. However, measurements of binding affinities and kinetics of SOD-mediated oligomerization within proteins that undergo PSCP is very rare and will need to be undertaken to test this hypothesis. It may also be possible that the unfolded conformation of the TDP-43 helix-forming region comprises stickers with such low affinities that they are unable to support PSCP. In this context, helix formation may push the sticker affinities over a required affinity threshold such that PSCP is favored. Here, the specific residues involved in this region may encode these features, and the differences among the TDP-43 helix and the Whi3 CC may be explained by sequence composition. Since the Whi3 CC is primarily composed of Glu which clearly drives PSCP in the rest of the Whi3 QRR, the unfolded conformation of the CC may be in the correct affinity regime for promoting PSCP, while stabilization as a helix may push this domain into a high-affinity zone that instead promotes the formation of inert oligomers and opposes PSCP. The suggestion of the existence of a “goldilocks zone” for binding affinities that promote PSCP will need to be theoretically and experimentally validated. Additionally, careful studies of both sequence and structure contributions in SODs and IDRs in proteins will be essential for understanding which self-assembly pathways are favored in a given context.

Recent work from the Gladfelter lab has identified some of the RNA sequence and structure features that give rise to the different modes of association of segments of the SARS-CoV-2 gRNA with N-protein (140). N-protein has two RNA binding domains which are proposed to bind to different RNA elements. RBD1 binds specific sequences and experiences enhanced binding when those sequences are structured, i.e. double-stranded, while RBD2 has no sequence preference and instead prefers double-stranded RNA regions in general (140). These two modes of association presumably lead to different protein-RNA binding affinities, although this has not yet been measured. Additionally, since RBD2 overlaps with the dimerization domain in N-protein, it is possible that a temperature-dependent dimer dissociation process facilitates RBD2-driven PSCP at high temperatures, explaining the lower critical solution temperature (LCST) phase behavior of N-protein with RNA (140). Clearly the complex interplay of RNA-protein binding affinities and SOD-mediated protein interactions can result in complex phase behavior, encoding temperature dependent PSCP and small-scale complex formation.

Identification of the molecular features that encode small-scale assemblies vs macroscopic condensates formed by PSCP is not only important from a fundamental chemical physics perspective but also from a biological one. Unlike the micron-scale condensates commonly observed *in vitro*, many condensates observed within cells appear as diffraction limited puncta by light microscopy, or are not visible at all (8). This observation raises several questions, including whether micron-scale or nano-scale assemblies are relevant for *in vivo* function, and whether condensates observed in cells are formed by PSCP or some other self-assembly pathways either on the way to PSCP or completely distinct from it. A recent study demonstrates that subsaturated solutions of phase separating proteins form a wide distribution of cluster sizes *in vitro* (141). The authors propose that the presence of multiple energy scales of binding among proteins can lead to heterogeneous cluster sizes. This claim is reminiscent of the behaviors seen for Whi3 and N-protein with their RNA binding partners. It may become increasingly clear in the coming years that such subsaturated solutions and/or diffraction-limited condensates formed by processes other than PSCP are functionally relevant *in vivo*, in addition to the large macroscopic condensates formed by PSCP that the field has been primarily focused on thus far.

There are undoubtedly other molecular features that control effective valence and binding affinities of effective binding sites that have yet to be explored. Transiently structured SODs within IDRs, as in Whi3 and TDP-43, may be particularly sensitive to changes in solution parameters like temperature, salt concentrations, and pH. There is evidence for both hot (142) and cold (143) temperature induced condensate formation in *Arabidopsis thaliana*. The molecular mechanisms underlying temperature sensitivity in these systems are not known but may implicate temperature-dependent folding/unfolding transitions of transiently structured SODs. The Whi3 CC contains five His residues within its 21-residue stretch. His has a pI of 7.6 at 25C, rendering its charge state especially sensitive to changes in pH around this value. The high concentration of His within the CC may provide a pH-dependence to its stability and binding affinity. Further studies will need to be undertaken to test this hypothesis.

The role of RNA-RNA oligomer formation and RNA structure is not well-studied in the context of PSCP. RNA molecules can base pair with other RNAs to create RNA oligomers, or with themselves to form complex and dynamic secondary structures (20). Studies of RNA-RNA interactions mediated by nucleotide complementary that influence PSCP are uncommon, but important examples include the role of secondary structure in maintaining distinct condensates formed by Whi3 and several different mRNAs (49), the role of disordered RNAs in forming extensive RNA-RNA contacts that lead to dynamically arrested condensates (90), and the suggestion that stress granules are formed by aggregation of disordered RNAs undergoing extensive RNA-RNA interactions (144). It will be essential to study how the effective valence and the binding affinities of effective binding sites in RNA-RNA interactions may affect PSCP, and how their influence is similar to or distinct from the protein-protein and protein-RNA interactions discussed previously.

An RNA molecule with intramolecular base pairing has a secondary structure that effectively renders it a branched polymer, making it distinct from linear proteins (145). RNAs have been treated with graph theory for many years as a simple way to represent their branched structures (145-149). The Gelbart group has shown that if an RNA secondary structure is known, its representation as a graph allows for accurate prediction of its hydrodynamic radius, which is significantly smaller than that of an unstructured, linear molecule of equal molecular weight (150, 151). Using simulations, they have also predicted that RNAs with more compact graph topologies outcompete less compact RNAs in binding proteins (152). Compact branched structures facilitate more protein-protein interactions once proteins are adsorbed onto RNA molecules, effectively increasing the binding affinities of the highly compact RNAs relative to their less compact counterparts. It thus seems likely that alteration of RNA branch architecture can tune the effective valence and the binding affinities of effective protein binding sites, which could influence PSCP and other self-assembly processes as described for proteins previously. Studies of synthetic polymers have shown that branched polymers have significantly different phase behavior (153) and rheological properties (154) than linear polymers. Although these properties have been studied for non-biological polymers, the role of branched architectures in the formation of biomolecular condensates has not been studied. My future work will explore the influence of RNA branch architectures on PSCP and self-assembly.

A final scenario which modulates self-assembly processes in the context of PSCP is surface adsorption. Generally, surface adsorption catalyzes reactions relative to bulk reactions (155). A recent study in the Gladfelter lab demonstrated that when *CLN3* RNA is tethered to membrane bilayers, Whi3 can co-phase separate with the surface-bound RNA at higher salt concentrations and at Whi3 bulk concentrations orders of magnitude lower than its saturation concentration when alone (8). Surprisingly, membrane-adsorbed RNP condensates reached steady-state within minutes, remaining as a dispersion of condensates hundreds of nanometers in diameter (8). This dynamic arrest was attributed to greatly reduced diffusion by large complexes tethered to membranes such that coalescence is suppressed (8). Membranes represent an important class of surfaces that may alter PSCP, but they are not the only relevant surfaces in the cell. If nucleic acid polymers or cytoskeletal filaments are orders of magnitude larger than a phase-separating protein, they may be more accurately regarded as pseudo-one-dimensional surfaces onto which proteins can adsorb. The transcription factor FoxA1 has been shown to form condensates on DNA molecules, generating pulling forces that sequester nearby DNA into condensates which increases tension on the whole DNA polymer (156). The microtubule-associated protein TPX2 has been shown to adsorb to microtubules, break into droplets due to the Rayleigh-Plateau instability, and nucleate microtubule branching from the resulting droplets (9). In all of these studies, condensates remained small and localized to specific locations due to surface interactions which may represent a key strategy utilized in cells for both size control and spatial targeting of biomolecular condensates.

The gRNA molecules studied in Chapter 3 are sufficiently large relative to N-protein that they may also be regarded as surfaces. As shown in Chapter 3, N-protein phase separates on gRNA molecules at concentrations far below its saturation concentration when alone. N-protein condensates on gRNA molecules remained small and localized to regions of PSCP-promoting sites on the gRNA molecule. However, since the gRNA molecules are deformable polymers in bulk solution, they can also participate in intra- and inter-molecular interactions, which complicates their interpretation as surfaces analogous to membranes, for example. Instead, they occupy some intermediate regime, having properties of both surfaces and additional components in a multi-component phase separation system with N-protein. The length scale of the large polymer which delineates multi-component PSCP from surface adsorption is not known. For systems that appear to fall in the middle of these two extremes, it is unclear which physical framework is most appropriate in their interpretation. Are single complexes of gRNA with ~100 N-proteins dense phases formed by PSCP, adsorption on a deformable surface saturated with its adsorbate, cluster formation in subsaturated conditions, or something else? Theoretical progress is needed to fully characterize these biologically common scenarios.

The biomolecular condensate field has traditionally understood that biomolecular condensate formation is due to LLPS. However, it is becoming increasingly clear that macroscopic dense phases are not simple liquids but are instead viscoelastic fluids. A more accurate characterization of their self-assembly process is therefore PSCP. Additionally, the importance of other types of smaller assemblies commonly associated with PSCP is gaining appreciation. These off-pathway or subsaturated clusters may be of functional relevance *in vivo* in addition to their better-studied macroscopic counterparts. The effective valence and binding affinities of effective binding sites appear to drive molecules along these myriad assembly pathways. Stability of SODs, sequence composition and sticker patterning within IDRs, RNA-RNA interactions, RNA branch architectures, and surface adsorption likely all play crucial roles in determining the valence and affinities of molecules. The works in this thesis contribute to the understanding of the molecular features relevant for the formation of biomolecular condensates.

APPENDIX A: BIOPHYSICAL PROPERTIES GOVERNING SEPTIN ASSEMBLY[5]

Septins are cytoskeletal proteins that are found in all eukaryotes except plants. In *Saccharomyces cerevisiae* there are 4 septin proteins that oligomerize in palindromic fashion to create rod-like octamers which act as the basic subunit of the septin cytoskeleton. Octamers bind to and diffuse on membranes where they can polymerize end-on to create filaments that can further associate to create complex, higher-order structures. Although polymerization of other cytoskeletal proteins like actin and microtubules is well studied and has been shown to be cooperative, the mechanisms underlying septin membrane adsorption and polymerization on membranes are not characterized biophysically. We studied the polymerization of purified septins on lipid bilayers *in vitro* using TIRF microscopy. At steady-state, filament length distributions are exponentially distributed, suggesting that polymerization is isodesmic, not cooperative. Additionally, time-lapses of septin membrane adsorption do not have lag phases and are always fit well by single exponential functions, suggesting that cooperativity in adsorption is not relevant for the experimental parameter regimes investigated.

To further understand our experimental data, we developed a plugin to the molecular dynamics package HOOMD-blue which allows for simulations of septin polymerization on membranes with prescribed reaction rates, persistence lengths, and excluded volume interactions. Under the assumption that septin polymerization is isodesmic, we show that while exponential steady-state length distributions are expected, membrane adsorption kinetics can either show features of cooperativity or not depending on the relationships between the reaction rates and the diffusion coefficient. We experimentally determined membrane binding kinetics, filament

---

[5] This work is currently under review at *eLife*, but due to major revisions it is in a state of flux and is not suitable for inclusion as a full chapter. My contributions (will) include development of the theories and models and analysis of their results, *in vitro* septin experiments and their analysis, and writing the revised draft.

fragmentation rates, and octamer diffusion of septins on membranes. By using these measurements to parameterize our model, we are able to fit the unknown filament annealing rate by matching experimentally measured steady-state filament length distributions across a range of concentrations, supporting our model of isodesmic septin polymerization and explaining the non-cooperative form of the membrane adsorption kinetics we observe experimentally. We also explore the effects of the reaction rates, concentration, and filament persistence length on local filament alignment and compare to experiments.

Finally, we used yeast extracts to explore the effects of possible septin regulators by observing the behavior of septins in knockout mutants. We used our parameterized model to hypothesize about how regulators may affect certain reactions rates and physical parameters and explain observed mutant phenotypes. We found that both lowering filament persistence length and fragmentation rate promotes assembly of septin rings, while increasing persistence length increases local alignment at sufficiently high concentrations. This work reports biophysical measurements of key septin reaction rates and establishes a physical framework for septin polymerization and membrane adsorption.

APPENDIX B: CONTRIBUTIONS TO OTHER WORKS

The following is a list of my contributions to other works that have not been discussed in Chapters 2 or 3, or Appendix A.

W.T. Snead, A.P. Jalihal, T.M. Gerbich, I. Seim, Z. Hu, A.S. Gladfelter. Membrane surfaces regulate assembly of a ribonucleoprotein condensate. *Nature Cell Biology*, in press, 2022. I designed mean-field models that were used in earlier drafts of the work, and I provided intellectual input throughout the project including suggestions of specific data analyses.

C.R. Roden, Y. Dai, I. Seim, M. Lee, R. Sealfon, G. A. McLaughlin, M. A. Boerneke, C. Iserman, S.A. Wey, J.L. Ekena, O.G. Troyanskaya, K. Weeks, L. You, A. Chilkoti, A. Gladfelter. Double-stranded RNA drives SARS-CoV-2 nucleocapsid protein to undergo phase separation at specific temperatures. In review at *Nucleic Acids Research*. I designed, implemented, and analyzed simulations that were used in earlier drafts of the work. I helped to perform mass photometry experiments on N-protein and the RBD2 del N-protein mutant (Figs. 3H, I). I wrote and edited sections of the manuscript.

K.S. Cannon, J.M. Vargas-Muniz, N. Billington, I. Seim, J. Ekena, J. Sellers, P. Philippsen, A.S. Gladfelter. A gene duplication of a septin provides a developmentally-regulated filament length control mechanism. to be submitted to the *Journal of Cell Biology*, 2022. bioRxiv link: https://doi.org/10.1101/2021.04.23.441157. I developed software that allows for fitting of left-truncated exponential distributions and which plots histograms with models fits that was used in this work.

S.T. Kudlacek, S. Metz, D. Thiono, A.M. Payne, T.T.N. Phan, S. Tian. L.J. Forsberg, J. Maguire, I. Seim, S. Zhang, A. Tripathy, J. Harrison, N.I. Nicely, L. Premkumar, A.M. de Silva, B. Kuhlman. Designed, highly expressing, thermostable dengue virus 2 envelope protein dimers elicit quaternary epitope antibodies. *Science Advances*, 7(42), 2021. During my rotation in the Kuhlman lab, I used the Rosetta simulation software to design stabilizing mutants of the Dengue virus envelope protein (E-protein) dimer. I developed several protocols that allowed for exhaustive mutagenesis and analysis of resulting free energy structures. The data I produced was later used to experimentally test E-protein mutants in this work.

Y. Ling, I. Seim, M. Lysy, J.M. Newby, D.B. Hill, J. Cribb, M.G. Forest. Measurement error correction in particle tracking microrheology. *Annals of Applied Statistics*, in press, 2022. I observed non-physical effects of different camera settings when analyzing microbead rheology data by down-sampling position time series. This observation led to the models developed in this work. I analyzed experimental data with models developed in this work, provided intellectual input throughout the project, and helped to write and edit the paper.

APPENDIX C: NOVEL TOOLS AND SOFTWARE

The following is a list of theoretical tools and software that I have developed during graduate school that have each proved useful in several projects and which I anticipate using in future projects.

Quantitative speckle microscopy

Quantitative speckle microscopy is the term I use to describe the technique in which a certain quantity of a labeled molecule is doped in with an unlabeled version of the same molecule to obtain a desired labeled percentage and total concentration. This labeled percentage can be used to directly quantify the number of particles in a particle tracking experiment, or to estimate concentrations in lower resolution microscopy images when compared with dye calibration curves. The following quantities must be measured or known: $[u]$, the concentration of the unlabeled molecule; $[l]$, the concentration of the labeled molecule; l_L, the labeled fraction of the labeled molecule; V, the final reaction volume; C, the desired final concentration of the molecule; and L, the desired final labeled fraction of the molecule. Using these variables, we will solve for u, the volume of unlabeled protein, and l, the volume of labeled protein needed to satisfy both C and L. First, we can write down an expression for the desired final labeled fraction:

$$\frac{l_L[l]l}{[u]u + [l]l} = L$$

Solving for l:

$$l = \frac{L[u]}{(l_L - L)[l]}u = \lambda u$$

where $\lambda = \frac{L[u]}{(l_L - L)[l]}$.

We define two convenient variables for later use: $\phi_u = \frac{u}{u+l}$ and $\phi_l = 1 - \phi_u$. Since $l = \lambda u$, we have $\phi_u = \frac{u}{u+\lambda u}$ → $\phi_u = \frac{(l_L - L)[l]}{(l_L - L)[l] + L[u]}$. Also, $\phi_l = 1 - \phi_u = \frac{L[u]}{(l_L - L)[l] + L[u]}$.

We can write a second expression which relates volumes and concentrations:

$$\frac{u + l}{V} = \frac{C}{\phi_u[u] + \phi_l[l]}$$

Solving for u: $\frac{\frac{1}{\phi_u}u}{V} = \frac{C}{\phi_u[u]+\phi_l[l]}$ → $u[u] + u\frac{\phi_l}{\phi_u}[l] = VC$ → $u[u] + u\frac{L[u]}{(l_L - L)[l]}[l] = VC$

→ $u = \frac{VC}{[u]+\frac{L[u]}{(l_L - L)[l]}[l]}$. Substituting in λ, we obtain the desired formulas:

$$u = \frac{VC}{[u] + \lambda[l]}$$

$$l = \lambda u$$

Dybond plugin for HOOMD-blue

I created a C++ plugin for the molecular dynamics software, HOOMD-blue, which allows for the runtime creation and deletion of interparticle bonds with prescribed valence, on/off kinetics, and an optional angular dependence of binding, and which allows for the runtime creation and deletion of angle potentials between triplets of bonded beads with prescribed energies that allows for encoding of persistence length. This plugin is compatible with all existing bond and angle potentials within HOOMD, and since it is written at the C++ level, it can access the internal neighborlist such that its efficiency is $O(N)$. The code is available on my github page: https://github.com/iseim

Fitting left-truncated exponentially distributed data

When observing exponentially distributed septin filaments lengths by TIRF microscopy, the lengths of filaments shorter than the diffraction limit of light (~200nm) are not resolvable. The resulting length distributions are left-truncated exponentials, in which the data below a cutoff value (here the diffraction limit of light) is missing. The following describes properties of these distributions and the resulting scheme that can be for fitting them:

Let X be an exponentially distributed random variable, $X \sim \text{Exp}(\lambda)$. X has the following PDF:

$$f(x) = \lambda \exp(-\lambda x)$$

with expected value $\frac{1}{\lambda}$. X has the property that it is *memoryless*, which means that:

$$P(X > x + a | X > a) = P(X > x)$$

for some $a > 0$. This has implications for left-truncated exponential distributions, which have the following PDF:

$$f_{trunc}(x) = \frac{g(x)}{\exp(-\lambda a)}$$

where

$$g(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x \geq a \\ 0 & \text{otherwise} \end{cases}$$

This distribution can be shifted back to 0 by subtracting the lower bound, resulting in the following:

$$\begin{aligned} f_{trunc}(x + a) &= (\exp(-\lambda a))^{-1} \lambda \exp\big(-\lambda(x + a)\big) \\ &= (\exp(-\lambda a))^{-1} (\exp(-\lambda a))\lambda \exp(-\lambda(x)) = f(x) \end{aligned}$$

Therefore, shifting the truncated exponential back to 0 results in the same PDF as the original distribution. When dealing with a left-truncated exponentially distributed dataset, the lower bound can be subtracted from every observation and the resulting data can be fit with a simple exponential distribution, $f(x)$, to obtain the desired parameter λ. To overlay the truncated dataset with the fit exponential, the fit exponential PDF needs to be scaled by the factor $\exp(\lambda a)$ to account for the missing probability density to the left of a in the dataset.

REFERENCES

1. J.-M. Choi, A. S. Holehouse, R. V. Pappu, Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics* **49**, 107-133 (2020).

2. A. W. Fritsch *et al.*, Local thermodynamics govern formation and dissolution of *Caenorhabditis elegans* P granule condensates. *Proceedings of the National Academy of Sciences* **118**, e2102772118 (2021).

3. C. A. Weber, D. Zwicker, F. Jülicher, C. F. Lee, Physics of active emulsions. *Reports on Progress in Physics* **82**, 064601 (2019).

4. J. Berry, C. P. Brangwynne, M. Haataja, Physical principles of intracellular organization via active and passive phase transitions. *Reports on Progress in Physics* **81**, 046601 (2018).

5. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology* **30**, 39-58 (2014).

6. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434 (2019).

7. D. Zwicker, A. A. Hyman, F. Jülicher, Suppression of Ostwald ripening in active emulsions. *Physical Review E* **92**, 012317 (2015).

8. W. T. Snead, T. M. Gerbich, I. Seim, Z. Hu, A. S. Gladfelter, Membrane surfaces regulate assembly of a ribonucleoprotein condensate. *bioRxiv* 10.1101/2021.04.24.441251, 2021.2004.2024.441251 (2021).

9. S. U. Setru *et al.*, A hydrodynamic instability drives protein droplet formation on microtubules to nucleate branches. *Nature Physics* **17**, 493-498 (2021).

10. A. W. Folkmann, A. Putnam, C. F. Lee, G. Seydoux, Regulation of biomolecular condensates by interfacial protein clusters. *Science* **373**, 1218-1224 (2021).

11. M. Rubinstein, R. H. Colby, "Chapter 6: Random Branching and Gelation" in Polymer Physics. (Oxford University Press, 2003).

12. M. Rubinstein, R. H. Colby, "Chapter 7: Networks and Gels" in Polymer Physics. (Oxford University Press, 2003).

13. L. Jawerth *et al.*, Protein condensates as aging Maxwell fluids. *Science* **370**, 1317-1323 (2020).

14. I. Alshareedah, M. M. Moosa, M. Pham, D. A. Potoyan, P. R. Banerjee, Programmable viscoelasticity in protein-RNA condensates with disordered sticker-spacer polypeptides. *Nature Communications* **12**, 6620 (2021).

15. S. L. Currie, M. K. Rosen, Using quantitative reconstitution to investigate multicomponent condensates. *RNA* **28**, 27-35 (2022).

16. W. M. Jacobs, D. Frenkel, Phase Transitions in Biological Systems with Many Components. *Biophysical Journal* **112**, 683-691 (2017).

17. W. M. Jacobs, Self-Assembly of Biomolecular Condensates with Shared Components. *Physical Review Letters* **126**, 258101 (2021).

18. K. Shrinivas, M. P. Brenner, Phase separation in fluids with many interacting components. *Proceedings of the National Academy of Sciences* **118**, e2108551118 (2021).

19. S. F. Banani *et al.*, Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663 (2016).

20. C. Roden, A. S. Gladfelter, RNA contributions to the form and function of biomolecular condensates. *Nature Reviews Molecular Cell Biology* **22**, 183-195 (2021).

21. D. Bracha *et al.*, Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175**, 1467-1480.e1413 (2018).

22. M. Garcia-Jove Navarro *et al.*, RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nature Communications* **10**, 3230 (2019).

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

24. J. R. Espinosa *et al.*, Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. *Proc Natl Acad Sci U S A* **117**, 13238-13247 (2020).

25. E. B. Stukalin, L.-H. Cai, N. A. Kumar, L. Leibler, M. Rubinstein, Self-Healing of Unentangled Polymer Networks with Reversible Bonds. *Macromolecules* **46**, 7525-7541 (2013).

26. J. Wang *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **174**, 688-699. e616 (2018).

27. E. W. Martin *et al.*, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694-699 (2020).

28. R. K. Das, R. V. Pappu, Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences* **110**, 13392-13397 (2013).

29. U. Rana, C. P. Brangwynne, A. Z. Panagiotopoulos, Phase separation vs aggregation behavior for model disordered proteins. *J Chem Phys* **155**, 125101 (2021).

30. J. Cubuk *et al.*, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nature Communications* **12**, 1936 (2021).

31. D. W. Sanders *et al.*, Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **181**, 306-324.e328 (2020).

32. P. Yang *et al.*, G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325-345 e328 (2020).

33. K. B. Shpargel, J. K. Ospina, K. E. Tucker, A. G. Matera, M. D. Hebert, Control of Cajal body number is mediated by the coilin C-terminus. *Journal of Cell Science* **116**, 303 (2003).

34. S. K. Powers *et al.*, Nucleo-cytoplasmic Partitioning of ARF Proteins Controls Auxin Responses in Arabidopsis thaliana. *Molecular Cell* **76**, 177-190.e175 (2019).

35. Y. Shin *et al.*, Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-171.e114 (2017).

36. S. Roberts *et al.*, Injectable tissue integrating networks from recombinant polypeptides with tunable order. *Nature Materials* **17**, 1154-1163 (2018).

37. A. Clery, F. H.-T. Allain, "From Structure to Function of RNA Binding Domains" in RNA Binding Proteins*,* L. Zdravko, Ed. (Landes Bioscience, 2011).

38. P. A. Chong, R. M. Vernon, J. D. Forman-Kay, RGG/RG Motif Regions in RNA Binding and Phase Separation. *J Mol Biol* **430**, 4650-4665 (2018).

39. J. A. Riback *et al.*, Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209-214 (2020).

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

41. S. Maharana *et al.*, RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918-921 (2018).

42. C. Iserman *et al.*, Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Molecular Cell* **80**, 1078-1091.e1076 (2020).

43. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* **18**, 285-298 (2017).

44. J. Guillen-Boixet *et al.*, RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **181**, 346-361 e317 (2020).

45. C. J. Decker, R. Parker, P-Bodies and Stress Granules: Possible Roles in the Control of Translation and mRNA Degradation. *Cold Spring Harbor Perspectives in Biology* **4** (2012).

46. W. Xing, D. Muhlrad, R. Parker, M. K. Rosen, A quantitative inventory of yeast P body proteins reveals principles of composition and specificity. *eLife* **9**, e56525 (2020).

47. J. A. Greig *et al.*, Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Molecular Cell* **77**, 1237-1250.e1234 (2020).

48. C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences USA* **108**, 4334-4339 (2011).

49. E. M. Langdon *et al.*, mRNA structure determines specificity of a polyQ-driven phase separation. *Science* **360**, 922 (2018).

50. S. Boeynaems *et al.*, Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proceedings of the National Academy of Sciences USA* **116**, 7889-7898 (2019).

51. I. Alshareedah, M. M. Moosa, M. Raju, D. A. Potoyan, P. R. Banerjee, Phase transition of RNA−protein complexes into ordered hollow condensates. *Proceedings of the National Academy of Sciences* **117**, 15650 (2020).

52. M. C. Cohan, R. V. Pappu, Making the Case for Disordered Proteins and Biomolecular Condensates in Bacteria. *Trends in Biochemical Sciences* **45**, 668-680 (2020).

53. P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340 (2012).

54. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).

55. J.-M. Choi, A. A. Hyman, R. V. Pappu, Generalized models for bond percolation transitions of associative polymers. *Physical Review E* **102**, 042403 (2020).

56. T. S. Harmon, A. S. Holehouse, R. V. Pappu, Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. *New Journal of Physics* **20**, 045002 (2018).

57. Timothy J. Nott *et al.*, Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell* **57**, 936-947 (2015).

58. R. M. Vernon *et al.*, Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7**, e31486 (2018).

59. J. P. Brady *et al.*, Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc Natl Acad Sci U S A* **114**, E8194-E8203 (2017).

60. T. H. Kim *et al.*, Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* **365**, 825 (2019).

61. T. Kaur *et al.*, Sequence-encoded and composition-dependent protein-RNA interactions control multiphasic condensate morphologies. *Nature Communications* **12**, 872 (2021).

62. X. Su *et al.*, Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595-599 (2016).

63. P. St George-Hyslop *et al.*, The physiological and pathological biophysics of phase separation and gelation of RNA binding proteins in amyotrophic lateral sclerosis and fronto-temporal lobar degeneration. *Brain Research* **1693**, 11-23 (2018).

64. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nature Physics* **11**, 899-904 (2015).

65. H. Zhang *et al.*, RNA controls PolyQ protein phase transitions. *Molecular Cell* **60**, 220-230 (2015).

66. A. Toto *et al.*, Binding induced folding: Lessons from the kinetics of interaction between NTAIL and XD. *Archives of Biochemistry and Biophysics* **671**, 255-261 (2019).

67. K. Bugge *et al.*, αα-Hub domains and intrinsically disordered proteins: A decisive combo. *Journal of Biological Chemistry* **296**, 100226 (2021).

68. H. J. Dyson, P. E. Wright, Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300*. *Journal of Biological Chemistry* **291**, 6714-6722 (2016).

69. S. L. Shammas, M. D. Crabtree, L. Dahal, B. I. M. Wicky, J. Clarke, Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies*. *Journal of Biological Chemistry* **291**, 6689-6695 (2016).

70. R. Vancraenenbroeck, Y. S. Harel, W. Zheng, H. Hofmann, Polymer effects modulate binding affinities in disordered proteins. *Proceedings of the National Academy of Sciences* **116**, 19506 (2019).

71. T. R. Sosnick, S. Jackson, R. R. Wilk, S. W. Englander, W. F. DeGrado, The role of helix formation in the folding of a fully α-helical coiled coil. *Proteins: Structure, Function, and Bioinformatics* **24**, 427-432 (1996).

72. E. Dürr, H. R. Bosshard, Folding of a three–stranded coiled coil. *Protein Science* **9**, 1410-1415 (2000).

73. J. K. Myers, T. G. Oas, Reinterpretation of GCN4-p1 folding kinetics: partial helix formation precedes dimerization in coiled coil folding11Edited by P. E. Wright. *Journal of Molecular Biology* **289**, 205-209 (1999).

74. A. E. Conicella *et al.*, TDP-43 α-helical structure tunes liquid–liquid phase separation and function. *Proceedings of the National Academy of Sciences* **117**, 5883 (2020).

75. C. Lee, P. Occhipinti, A. S. Gladfelter, PolyQ-dependent RNA–protein assemblies control symmetry breaking. *Journal of Cell Biology* **208**, 533-544 (2015).

76. A. N. Lupas, M. Gruber, "The Structure of α-Helical Coiled Coils" in Advances in Protein Chemistry. (Academic Press, 2005), vol. 70, pp. 37-38.

77. M. Gruber, J. Söding, A. N. Lupas, Comparative analysis of coiled-coil prediction methods. *Journal of Structural Biology* **155**, 140-145 (2006).

78. Y. Cheng *et al.*, Mining α-Helix-Forming Molecular Recognition Features with Cross Species Sequence Alignments. *Biochemistry* **46**, 13468-13477 (2007).

79. C. J. Oldfield *et al.*, Coupled Folding and Binding with α-Helix-Forming Molecular Recognition Elements. *Biochemistry* **44**, 12454-12470 (2005).

80. R. K. Das, S. L. Crick, R. V. Pappu, N-Terminal Segments Modulate the α-Helical Propensities of the Intrinsically Disordered Basic Regions of bZIP Proteins. *Journal of Molecular Biology* **416**, 287-299 (2012).

81. T. S. Harmon *et al.*, GADIS: Algorithm for designing sequences to achieve target secondary structure profiles of intrinsically disordered proteins. *Protein Engineering, Design and Selection* **29**, 339-346 (2016).

82. M. A. Kelly *et al.*, Host−Guest Study of Left-Handed Polyproline II Helix Formation. *Biochemistry* **40**, 14376-14383 (2001).

83. J. M. Scholtz *et al.*, Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. *Proceedings of the National Academy of Sciences* **88**, 2854 (1991).

84. A. J. Miles, S. G. Ramalli, B. A. Wallace, DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. *Protein Science* **n/a**.

85. T. Ishida, K. Kinoshita, PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research* **35**, W460-W464 (2007).

86. D. W. Urry, T. H. Ji, Distortions in circular dichroism patterns of particulate (or membranous) systems. *Archives of Biochemistry and Biophysics* **128**, 802-807 (1968).

87. D. W. Urry, J. Krivacic, Differential Scatter of Left and Right Circularlay Polarized Light by Optically Active Particulate Systems. *Proceedings of the National Academy of Sciences* **65**, 845 (1970).

88. L. N. M. Duyens, The flattening of the absorption spectrum of suspensions, as compared to that of solutions. *Biochimica et Biophysica Acta* **19**, 1-12 (1956).

89. N. E. Zhou, C. M. Kay, R. S. Hodges, Synthetic model proteins: the relative contribution of leucine residues at the nonequivalent positions of the 3-4 hydrophobic repeat to the stability of the two-stranded .alpha.-helical coiled-coil. *Biochemistry* **31**, 5739-5746 (1992).

90. W. Ma, G. Zheng, W. Xie, C. Mayr, In vivo reconstitution finds multivalent RNA–RNA interactions as drivers of mesh-like condensates. *eLife* **10**, e64252 (2021).

91. A. Z. Akcasu, I. C. Sanchez, Foundations and Extensions of the Cahn-Hilliard-Cook Theory. *Molecular Crystals and Liquid Crystals Incorporating Nonlinear Optics* **180**, 147-153 (1990).

92. X. Zeng, A. S. Holehouse, A. Chilkoti, T. Mittag, R. V. Pappu, Connecting Coil-to-Globule Transitions to Full Phase Diagrams for Intrinsically Disordered Proteins. *Biophysical Journal* **119**, 402-418 (2020).

93. F. Sciortino, R. Bansil, H. E. Stanley, P. Alstrom, Interference of Phase-Separation and Gelation - a Zeroth-Order Kinetic-Model. *Physical Review E* **47**, 4615-4618 (1993).

94. E. Zaccarelli, P. J. Lu, F. Ciulla, D. A. Weitz, F. Sciortino, Gelation as arrested phase separation in short-ranged attractive colloid-polymer mixtures. *J Phys-Condens Mat* **20** (2008).

95. K. Rhine *et al.*, ALS/FTLD-Linked Mutations in FUS Glycine Residues Cause Accelerated Gelation and Reduced Interactions with Wild-Type FUS. *Molecular Cell* **80**, 666-681.e668 (2020).

96. I. M. Lifshitz, V. V. Slyozov, The kinetics of precipitation from supersaturated solid solutions. *Journal of Physics and Chemistry of Solids* **19**, 35-50 (1961).

97. E. S. Freeman Rosenzweig *et al.*, The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell* **171**, 148-162.e119 (2017).

98. M. Ramazzotti, E. Monsellier, C. Kamoun, D. Degl'Innocenti, R. Melki, Polyglutamine Repeats Are Associated to Specific Sequence Biases That Are Conserved among Eukaryotes. *PLOS ONE* **7**, e30824 (2012).

99. A. Urbanek *et al.*, Flanking Regions Determine the Structure of the Poly-Glutamine in Huntingtin through Mechanisms Common among Glutamine-Rich Human Proteins. *Structure* **28**, 733-746.e735 (2020).

100. B. Kokona, Z. P. Rosenthal, R. Fairman, Role of the Coiled-Coil Structural Motif in Polyglutamine Aggregation. *Biochemistry* **53**, 6738-6746 (2014).

101. L. K. Ford, L. Fioriti, Coiled-Coil Motifs of RNA-Binding Proteins: Dynamicity in RNA Regulation. *Frontiers in Cell and Developmental Biology* **8** (2020).

102. M. C. Ferrolino, D. M. Mitrea, J. R. Michael, R. W. Kriwacki, Compositional adaptability in NPM1-SURF6 scaffolding networks enabled by dynamic switching of phase separation mechanisms. *Nature Communications* **9**, 5064 (2018).

103. B. Xu *et al.*, Rigidity enhances a magic-number effect in polymer phase separation. *Nature Communications* **11**, 1561 (2020).

104. B. Bolognesi *et al.*, The mutational landscape of a prion-like domain. *Nature Communications* **10**, 4162 (2019).

105. L. A. Compton, W. C. Johnson, Jr., Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem* **155**, 155-167 (1986).

106. N. Sreerama, S. Y. Venyaminov, R. W. Woody, Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem* **287**, 243-251 (2000).

107. J. G. Lees, A. J. Miles, F. Wien, B. A. Wallace, A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **22**, 1955-1962 (2006).

108. N. Sreerama, R. W. Woody, Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* **287**, 252-260 (2000).

109. N. Sreerama, S. Y. Venyaminov, R. W. Woody, Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci* **8**, 370-380 (1999).

110. J. C. Yen, F. J. Chang, S. Chang, A new criterion for automatic multilevel thresholding. *IEEE Trans Image Process* **4**, 370-378 (1995).

111. P. Soille, L. Vincent, *Determining watersheds in digital pictures via flooding simulations*, Visual Communications and Image Processing '90 (SPIE, 1990), vol. 1360.

112. L. A. Ruiz-Taylor *et al.*, Monolayers of derivatized poly(L-lysine)-grafted poly(ethylene glycol) on metal oxides as a class of biomolecular interfaces. *Proc Natl Acad Sci U S A* **98**, 852-857 (2001).

113. F. Aguet, C. N. Antonescu, M. Mettlen, S. L. Schmid, G. Danuser, Advances in analysis of low signal-to-noise images link dynamin and AP2 to the functions of an endocytic checkpoint. *Dev Cell* **26**, 279-291 (2013).

114. A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering* **18**, 015012 (2009).

115. P. J. Flory, Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics* **10**, 51-61 (1942).

116. F. W. Altena, C. A. Smolders, Calculation of liquid-liquid phase separation in a ternary system of a polymer in a mixture of a solvent and a nonsolvent. *Macromolecules* **15**, 1491-1497 (1982).

117. J. Zhu, L. Q. Chen, J. Shen, V. Tikare, Coarsening kinetics from a variable-mobility Cahn-Hilliard equation: application of a semi-implicit Fourier spectral method. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **60**, 3564-3572 (1999).

118. S. G. Johnson, Notes on FFT-based differentiation.

119. S. Boeynaems *et al.*, Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**, 420-435 (2018).

120. S. Brocca, R. Grandori, S. Longhi, V. Uversky, Liquid-Liquid Phase Separation by Intrinsically Disordered Protein Regions of Viruses: Roles in Viral Life Cycle and Control of Virus-Host Interactions. *Int J Mol Sci* **21** (2020).

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

122. V. Rincheval *et al.*, Functional organization of cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus. *Nat Commun* **8**, 563 (2017).

123. J. Nikolic *et al.*, Negri bodies are viral factories with properties of liquid organelles. *Nat Commun* **8**, 58 (2017).

124. S. Guseva *et al.*, Measles virus nucleo- and phosphoproteins form liquid-like phase-separated compartments that promote nucleocapsid assembly. *Sci Adv* **6**, eaaz7095 (2020).

125. A. Monette *et al.*, Pan-retroviral Nucleocapsid-Mediated Phase Separation Regulates Genomic RNA Positioning and Trafficking. *Cell Rep* **31**, 107520 (2020).

126. S. Klein *et al.*, SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. *Nat Commun* **11**, 5885 (2020).

127. P. K. Hsieh *et al.*, Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent. *J Virol* **79**, 13848-13855 (2005).

128. T. M. Perdikari *et al.*, SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J* **39**, e106478 (2020).

129. A. Savastano, A. Ibanez de Opakua, M. Rankovic, M. Zweckstetter, Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat Commun* **11**, 6041 (2020).

130. H. Chen *et al.*, Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res* **30**, 1143-1145 (2020).

131. J. Towns *et al.*, XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **16**, 62-74 (2014).

132. D. Yang, J. L. Leibowitz, The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res* **206**, 120-133 (2015).

133. O. Ziv *et al.*, The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol Cell* **80**, 1067-1077 e1065 (2020).

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

135. M. Fixman, Radius of Gyration of Polymer Chains. *The Journal of Chemical Physics* **36**, 306-310 (1962).

136. O. Ziv *et al.*, COMRADES determines in vivo RNA structures and interactions. *Nat Methods* **15**, 785-788 (2018).

137. R. D. Cadena-Nava *et al.*, Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. *J Virol* **86**, 3318-3326 (2012).

138. T. Yamazaki *et al.*, Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol Cell* **70**, 1038-1053 e1037 (2018).

139. T. Hirose, T. Yamazaki, S. Nakagawa, Molecular anatomy of the architectural NEAT1 noncoding RNA: The domains, interactors, and biogenesis pathway required to build phase-separated nuclear paraspeckles. *Wiley Interdiscip Rev RNA* **10**, e1545 (2019).

140. C. A. Roden *et al.*, Double-stranded RNA drives SARS-CoV-2 nucleocapsid protein to undergo phase separation at specific temperatures. *bioRxiv* 10.1101/2021.06.14.448452, 2021.2006.2014.448452 (2021).

141. M. Kar *et al.*, Phase separating RNA binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *bioRxiv* 10.1101/2022.02.03.478969, 2022.2002.2003.478969 (2022).

142. J.-H. Jung *et al.*, A prion-like domain in ELF3 functions as a thermosensor in Arabidopsis. *Nature* **585**, 256-260 (2020).

143. P. Zhu, C. Lister, C. Dean, Cold-induced Arabidopsis FRIGIDA nuclear condensates for FLC repression. *Nature* **599**, 657-661 (2021).

144. N. Ripin, R. Parker, Are stress granules the RNA analogs of misfolded protein aggregates? *RNA* **28**, 67-75 (2022).

145. T. Schlick, Adventures with RNA graphs. *Methods* **143**, 16-33 (2018).

146. R. Nussinov, G. Pieczenik, J. R. Griggs, D. J. Kleitman, Algorithms for Loop Matchings. *SIAM Journal on Applied Mathematics* **35**, 68-82 (1978).

147. M. S. Waterman, T. F. Smith, RNA secondary structure: a complete mathematical analysis. *Mathematical Biosciences* **42**, 257-266 (1978).

148. B. A. Shapiro, K. Zhang, Comparing multiple RNA secondary structures using tree comparisons. *Bioinformatics* **6**, 309-318 (1990).

149. H. H. Gan *et al.*, RAG: RNA-As-Graphs database--concepts, analysis, and features. *Bioinformatics* **20**, 1285-1291 (2004).

150. A. Gopal, Z. H. Zhou, C. M. Knobler, W. M. Gelbart, Visualizing large RNA molecules in solution. *RNA* **18**, 284-299 (2012).

151. A. Borodavka *et al.*, Sizes of Long RNA Molecules Are Determined by the Branching Patterns of Their Secondary Structures. *Biophysical Journal* **111**, 2077-2085 (2016).

152. S. W. Singaram, R. F. Garmann, C. M. Knobler, W. M. Gelbart, A. Ben-Shaul, Role of RNA Branchedness in the Competition for Viral Capsid Proteins. *The Journal of Physical Chemistry B* **119**, 13991-14002 (2015).

153. M. Daoud, P. Pincus, W. H. Stockmayer, T. Witten, Phase separation in branched polymer solutions. *Macromolecules* **16**, 1833-1839 (1983).

154. R. G. Larson, Combinatorial Rheology of Branched Polymer Melts. *Macromolecules* **34**, 4556-4571 (2001).

155. K. J. Laidler, *Chemical Kinetics* (Prentice Hall, ed. 3rd Revised Edition, 1987).

156. T. Quail *et al.*, Force generation by protein–DNA co-condensation. *Nature Physics* **17**, 1007-1012 (2021).