Thermodynamics and kinetics of phase separation of protein–RNA mixtures by a minimal model

Jerelle A. Joseph\textsuperscript{1}, Jorge R. Espinosa\textsuperscript{2}, Ignacio Sanchez-Burgos\textsuperscript{3}, Adiran Garaizar\textsuperscript{4}, Daan Frenkel\textsuperscript{5}, and Rosana Collepardo-Guevara\textsuperscript{6,\*}  
\textsuperscript{1,2,3,4,6}Maxwell Centre, Cavendish Laboratory, Department of Physics, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0HE  
\textsuperscript{2,3,4,5,6}Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW  
\textsuperscript{1,2,3,4,6}Department of Genetics, University of Cambridge, Cambridge, CB2 3EH  
\*Correspondence: rc597@cam.ac.uk

ABSTRACT Intracellular liquid–liquid phase separation (LLPS) enables the formation of biomolecular condensates, such as ribonucleoprotein granules, which play a crucial role in the spatiotemporal organization of biomolecules (e.g., proteins and RNAs). Here, we introduce a patchy-particle–polymer model to investigate LLPS of protein–RNA mixtures. We demonstrate that, at low to moderate concentrations, RNA enhances the stability of RNA-binding protein (RBP) condensates because it increases the molecular connectivity of the condensed-liquid phase. Importantly, we find that RNA can also accelerate the nucleation stage of phase separation. Additionally, we assess how the capacity of RNA to increase the stability of condensates is modulated by the relative protein–protein/protein–RNA binding strengths. We find that phase separation and multiphase organization of multicomponent condensates is favored when the RNA binds with higher affinity to the lower valency proteins in the mixture, than to the cognate higher valency proteins. Collectively, our results shed light on the roles of RNA in ribonucleoprotein granule formation and the internal structuring of stress granules.

SIGNIFICANCE The interior of cells contains several membraneless compartments that are composed of proteins and RNA. These compartments are formed and sustained by LLPS. Here, we introduce a minimal coarse-grained model to study LLPS of protein–RNA mixtures. We find that RNA can increase the stability of phase-separated compartments by enhancing the molecular connectivity of proteins. Additionally, our results show that RNA actively recruits proteins—accelerating the nucleation and fusion stages of LLPS. Interestingly, we find that spatial segregation within protein–RNA compartments is controlled by fine-tuning the interaction strengths and stoichiometries of components. Our model, therefore, provides a useful tool for building a comprehensive mechanistic and thermodynamic view of protein–RNA LLPS.

1 INTRODUCTION

In recent years, it has become clear that liquid–liquid phase separation (LLPS) is responsible for the formation of membraneless organelles, including P granules and nuclear bodies (NBs) (1–4). These cellular bodies, often referred to as biomolecular condensates, display liquid-like properties, such as the ability to flow, coalesce, and drip (5–9), and are thought to self-assemble via condensation of proteins and other macromolecules in the cytoplasm and nucleoplasm. Specifically, multidomain proteins (5, 8, 10) and those comprising intrinsically disordered regions (IDRs) (11–14) have been shown to undergo LLPS, both in vitro and in cells. LLPS is mainly driven by multivalent protein–protein interactions (15, 16); for example, several phase-separating proteins possess low complexity domains (LCDs) that foster protein–protein condensation (2, 17, 18). Interactions with RNA (19–23) have been shown to strongly mediate LLPS; indeed, the vast majority of membraneless organelles [including nucleoli (19, 24, 25), stress granules (26, 27), P granules (13, 20, 28), and processing bodies (29, 30)] are ribonucleoprotein (RNP) granules (31–35), consisting of RNA and RNA-binding proteins (RBPs) (36). A central goal of the present work is to further elucidate one of the potential mechanisms by which RNA can enhance (or inhibit) LLPS and obtain predictive rules governing the composition of protein–RNA condensates.

In vitro experiments reveal that, although condensates are multicomponent systems, only a small subset of components may be required for LLPS (24, 37). For example, P granules are made up of several RNA and protein molecules; however, LAF-1 (a protein found in P granules) can self-associate into droplets that resemble P granules in vitro (13). In some cases, macromolecules that undergo LLPS may bind to and
recruit other molecules to phase-separated droplets. Landmark work by Banani and collaborators (38) showed that polySUMO and polySIM proteins assembled into droplets when mixed, and subsequently recruited fluorescently labeled SIM and SUMO monomers, respectively, to the condensates. Components, such as polySUMO and polySIM, that drive LLPS are often classified as ‘scaffolds’, and molecules that partition into droplets formed by scaffolds (e.g., SIM and SUMO monomers) are termed ‘clients’ (38). Scaffold and client stoichiometric ratios, valencies, and binding affinities have been postulated as crucial in compositional control of membraneless organelles (38).

Importantly, RNA molecules can be selectively recruited to condensates, and this recruitment is tuned by several factors including RNA length, flexibility, and shape (12, 22, 39, 40). For instance, some RNA-binding proteins contain structured RNA-binding regions—including certain RNA recognition motifs (RRMs) and zinc fingers—that bind to specific RNA sequences and drive phase separation (22, 41). In other cases, RBPs recruit RNA in a nonspecific manner (e.g., via IDR and arginine rich regions) (20, 36, 42); thereby, partitioning RNA molecules into phase-separated droplets. Still several RBPs may contain sites for both specific and nonspecific RNA-binding (36, 41, 43, 44). Accordingly, RNA polymers tend to mimic LCDs—exhibiting multivalence character and high propensity for adopting numerous conformations that promote condensation (23). In general, RNA valence increases with length, and longer RNAs have been found to drive stress granule formation (26). Additionally, the RBP-to-RNA ratio is also very important; with the high RNA ratios suppressing LLPS and lower RNA concentrations promoting LLPS of several proteins (21, 36, 41, 43, 45). In the latter case, RNA molecules may essentially act as scaffolds for LLPS. To better understand the interplay between these and other factors, we therefore, require biophysical models that can investigate the phase behavior of RBP–RNA mixtures.

Polymer physics provides key rules for predicting phase behavior of polymeric systems (46). For a given homopolymer–solvent mixture, the system produces two phases (i.e., a dense, polymer-rich phase and a dilute, solvent-rich phase) when the enthalpy of mixing exceeds the entropy of mixing. Flory–Huggins Theory quantifies the entropic and enthalpic terms of such systems and estimates the critical condition for phase separation (47, 48). Subsequent to these works, analytical (14, 49–56) and mean-field theoretical approaches (57–60) have been developed to study phase behavior of charged polymers and IDPs. Additionally, sequence-dependent continuum models (61–65) have been designed for probing biomolecular phase behavior. These models have proved extremely useful in identifying correlations between protein sequence and LLPS. Minimal continuum models (66–74) (including patchy-particle models) (68–73) and lattice-based approaches (16, 75–79) (notably “stickers-and-spacers” representations) (16, 75, 77) have significantly complemented and augmented these studies—revealing concentration-dependent features of multicomponent protein systems, aiding in the design of experimental studies on phase behavior, and accessing complete phase diagrams. The study of phase diagrams of biopolymer systems provides valuable insight in the various factors that influence phase separation. However, measuring complete phase diagrams is time consuming and the study of biomolecular phase behavior at atomic resolution is computationally expensive. Hence, approaches that simultaneously preserve important molecular and physicochemical details of LLPS and capture observable phase behavior in an efficient manner are appealing; since, they can be implemented at moderate computational cost.

Here we develop a simple coarse-grained approach that approximates key features of phase-separating RBPs and their RNA counterparts. The model is simple enough to permit simulation of phase transitions in mixtures containing thousands of proteins at low computational cost; it can be conveniently implemented, and it does not require extensive optimization of parameter sets. Using this minimal model, we investigate the effect of adding an RNA-like polymer to a pure RBP system. We find that the RNA polymer enhances the connectivity of RBPs and increases the critical temperature for phase separation. These results are consistent with experiments, where RNA was found to decrease the critical Concentration for LLPS in RBP–RNA mixtures (21, 36, 41, 43). Additionally, we demonstrate that RNA-like polymers accelerate the nucleation stage of protein condensate formation. We then study competition and cooperative effects in multi-component protein–RNA mixtures, and demonstrate how the droplet composition is tuned by the valencies, stoichiometries, and relative interaction strengths of the molecular components. Taken together, our work demonstrates the usefulness of minimal coarse-grained models in obtaining general rules governing RNA-driven LLPS and suggests possible molecular mechanisms involved in intracellular phase separation.

2 METHODS

Minimal coarse-grained model

We have developed a minimal coarse-grained model for RBPs that captures their multivalency (80) and RNA-binding ability—the two essential characteristics for LLPS of RBP–RNA mixtures—and can probe condensed-matter properties of biomolecular phase separation in an efficient manner. A key advantage of our approach is that all the potentials in our minimal model are continuous, and can be conveniently implemented in parallelised molecular dynamics (MD) software, which allows us to investigate systems of up to ten thousand proteins. Additionally, the design of the model eliminates the need for optimizing and storing extensive parameter sets, which ensures its utility and portability.

Our minimal model represents RBPs as hard spheres (of molecular diameter $\sigma = 3.405$ Å) decorated with attractive patches or “stickers” on their surface (Fig. 1A). We model
Figure 1: Obtaining phase diagrams of mixtures of RBPs and RNA. (A) Patchy-particle–polymer model for simulating interactions between proteins and RNAs. RNA is modeled as a (hard-sphere) self-avoiding polymer. RBPs are represented by hard spheres decorated with attractive patches. (B) Phase diagrams computed in terms of inverse inter-protein interaction strength ($E_{\text{RBP}}^c / E_{\text{prot–prot}}^c$; $E_{\text{RBP}}^c$ is the critical inverse protein–protein interaction strength of the pure RBP) and volume fraction of the RBP (i.e., $\phi_{\text{RBP}}$). In each system, 1000 RBPs were used, along with $n (n = 0, 1, 2, 3)$ chains of a 40-mer RNA. The densities of the dilute and condense phases are estimated from the density profiles, as described in the main text and SI Appendix Fig. S1. Horizontal error bars represent the standard deviations in the coexisting volume fractions. (C) Snapshots of direct coexistence simulations of RBP (top) and RBP+RNA (3x40-mer) mixture (bottom) at $E_{\text{RBP}}^c / E_{\text{prot–prot}}^c ≈ 1.05$. (D) Residue-level coarse-grained representations of PolyU RNA and RBP FUS. In the model, each protein residue or nucleic acid is represented as a single bead (see text and SI Appendix for details). (E) Phase diagrams computed in terms of temperature ($T / T_{\text{FUS}}^c$; $T_{\text{FUS}}^c$ is the critical temperature of pure FUS) and density of FUS ($\rho_{\text{FUS}}$). In each system, 24 chains of FUS were used, along with $n (n = 0, 1, 2, 3)$ chains of PolyU 175 nucleotides (nt) long. The densities of the dilute and condense phases are estimated from the density profiles, as described in the SI Appendix. Horizontal error bars represent standard deviations in the coexisting densities. (F) Snapshots of direct coexistence simulations of pure FUS (top) and FUS+PolyU (3x175 nt) mixture (bottom) at $T / T_{\text{FUS}}^c ≈ 1.00$. In phase diagrams, individual critical values of the inverse $E_{\text{prot–prot}}^c$ mixture (bottom) are estimated by fitting the differences in $\phi_{\text{RBP}}$ or $\rho_{\text{FUS}}$ (as described in ref. (80)), and corresponding critical $\phi_{\text{RBP}}$ or $\rho_{\text{FUS}}$ are derived by assuming that the law of rectilinear diameters and critical exponents (81) hold in the vicinity of $E^c$ or $T^c$. Uncertainties in the estimation of critical values (error bars) are obtained by performing the fitting procedure on three independent data sets.
two types of RBPs: (1) those that can drive phase separation via homotypic RBP–RBP interactions and, hence, act as scaffolds, and (2) those whose homotypic interactions are insufficient to drive phase separation and, hence, act as clients that are recruited into condensates via their interactions with the scaffolds. Based on our previous work exploring the role of valency in protein LLPS (80), we set three attractive LLPS-binding sites per scaffold RBP (3-valency RBPs), and two per client RBP (2-valency RBPs); this simple distinction allows us to capture the essential difference in the phase behavior between scaffolds and clients. Although many proteins, including RBPs, that undergo LLPS possess intrinsically disordered regions (11–14, 16, 82), our model captures the effects of protein multivalency and, therefore, approximates the way in which intrinsically disordered proteins interact with each other (72) and with RNA.

RNA molecules contain negatively charged sugar-phosphate backbones, and therefore tend to behave as self-avoiding polymers. Hence, as an extension to our previous work (80), we represent single-stranded unstructured RNAs (i.e., A- or U-rich with negligible base-pairing probability) as flexible, self-avoiding polymers (i.e., chains of hard spheres (83)) that interact with the minimal RBPs via attractive interactions (Fig. 1A). The weak attractive interactions between patch “RBP” particles and polymers typify RBP–RNA interactions observed in biomolecular condensates (20, 21, 36, 41–43).

**Direct coexistence simulations**

We employ direct coexistence simulations (84–86) to compute the phase diagrams of the protein–RNA mixtures; i.e., both liquid phases are simulated via MD in the same simulation box (80). Specifically, we measure the volume fraction of the protein (Fig. 1B) in each phase as a function of the protein–protein interaction strength. The volume fraction of the RBP (\( \phi_{RBP} \)) in a phase is defined as \( \phi_{RBP} = \frac{N_{RBP}}{V \times C_{RBP}} \), where \( N \) is the total number of RBPs, \( V_{RBP} \) is the molecular volume of an RBP [i.e., \( 4/3 \pi (\sigma/2)^3 \)], \( V \) is the total system volume, and \( C_{RBP} \) is the concentration). LLPS in cells occur over narrow temperature ranges, therefore, we assess the phase behavior of our mixtures by varying the inter-protein interaction strengths (\( E_{prot-prot} \) at a fixed temperature.

At a given value of the inter-protein interaction strength, the system is simulated until the potential energy and the density profile along the box long axis have converged (SI Appendix Fig. S1A). LLPS is then marked by the presence of sharp interfaces (SI Appendix Fig. S1B). The volume fractions of the protein-rich and protein-poor phase are then computed by averaging the volume fractions (i.e., from the density profile) for the respective phases, excluding regions near the interfaces (SI Appendix Fig. S1B).

Hence, for a particular protein, the compositions of the coexisting protein-rich and protein-poor phases define the range of volume fractions (or concentrations) for which LLPS takes place (Fig. 1B). Having established this reference phase diagram, we then assess the effect of adding an RNA-like polymer or different proteins on the location of the phase boundaries (Fig. 1B and C). This approach allows us to obtain general rules for how proteins and RNA molecules may partition into two-liquid phases.

Further details of our patchy-particle model (80) and the simulation parameters used are found in SI Appendix, Sections I and II.

**3 RESULTS AND DISCUSSION**

**Minimal model validation**

Many proteins that undergo homotypic LLPS have been shown to participate in multivalent heterotypic interactions with RNA that also facilitate LLPS (36, 41, 43). For example, Molliex and coworkers (43) showed that, while RBP hnRNPA1 was able to phase-separate on its own, LLPS of hnRNPA1 was significantly enhanced in the presence of RNA. This effect was marked by a dramatic decrease in the concentration of hnRNPA1 required for phase separation (43). Similar results were also reported by Lin and colleagues (36). RNA has also been shown to drive droplet formation of RBP Whi3 (22, 41) and PGL-3 protein (20) at physiological protein concentrations.

We first employ our minimal model to probe the effects of adding RNA at low to moderate concentrations (i.e., \( \phi_{RNA}/\phi_{RBP} < 15 \% \)) on the phase boundary of a single type of RNA binding protein (i.e., a trivalent RBP). The protein can self-associate via three binding sites on its surface and, as such, serves as a prototype for proteins that exhibit homotypically-driven LLPS. At different inter-protein interaction strengths (i.e., \( 11.5 \leq E_{prot-prot} \leq 14 \ k_BT \)), we compute the phase boundaries of the pure protein system (Fig. 1B). We then calculate the coexistence curve and the minimal interaction strength needed for phase separation (i.e., the critical point; at \( E_{prot-prot} \sim 10.786 \ k_BT \)). Above the critical point the pure protein system forms a single well-mixed phase (Fig. 1C, top panel). At larger interaction strengths (i.e., below the critical point), the system separates into a protein-rich and protein-poor phase (Fig. 1B).

We then add a polymer that mimics an intrinsically disordered RNA chain (i.e., a self-avoiding (83) fully flexible polymer made up of 40 monomeric units; 1x40-mer) that interacts with the RBPs via moderately weak attractive interactions (\( E_{prot-RNA} \sim 3 \ k_BT \)). We choose this polymer length, since it allows us to achieve a low RNA-to-protein concentration (i.e., \( \sim 4–12 \% \) RNA) that is consistent with experiments (20, 36, 43), while obtaining marked trends for how RNA affects protein LLPS. We then simulate the new RBP–RNA mixture above the critical point of the pure system (Fig. 1B), where the RBP can no longer sustain LLPS on its own. Note that to assess directly how effectively the RBP condenses in the presence of the RNA-like polymer, we compute our phase.
diagrams in terms of the volume fraction of the RBP, as opposed to assessing the combined volume fraction of the RBP and RNA. Consistent with experimental studies of RBP–RNA mixtures (20, 22, 36, 41, 43), the RNA-like polymer promotes phase separation of the trivalent RBP; i.e., a higher critical inverse interaction strength and a broader coexistence region is obtained for the mixture versus the pure protein system (red curve in Fig. 1B).

As in the case of the Whi3–RNA mixture (22, 41), the trivalent RBPs interact with the RNA polymer via different sites than those used for protein–protein association. This suggests that the RNA polymer acts as a high-valency molecule that effectively enhances the connectivity of the protein liquid network. Indeed the average number of protein–protein bonds in the RBP condensate increases fourfold when three 40-mer RNA-like polymers are added to the RBP system at an inter–protein interaction strength of 11 $k_BT$ (Fig. 2). Experiments and simulations have recently demonstrated that increasing the protein valency raises the critical temperature for phase separation (16, 72). These findings are consistent with earlier predictions by Bianchi et al. (87). Li et al. (5) also obtained a strong correlation between the phase boundary and the valency of interacting molecules for engineered SH3-$\text{PRM}_n$ protein mixtures. Importantly, Rao and Parker (88) demonstrated that mRNA enhanced P-body assembly by providing multiple interaction sites for certain P-body proteins. Moreover, mutations that hindered RNA-binding suppressed P-body formation (88). Ries et al. (89) recently showed that poly methylated mRNAs can act as multivalent scaffolds for binding certain proteins and driving LLPS.

We also test whether multiple RNA-like polymers can also enhance LLPS. Thus, in addition to the original RBP–RNA mixture, we study the phase behavior for systems containing two (2x40-mer) and three (3x40-mer) RNA-like chains. Within our model, we obtain a monotonic increase in the critical point for LLPS and a widening of the coexistence region Fig. 1B. Thus, in all cases tested, the RNA-like polymer enhances LLPS of our patchy RBPs. We note that, by construction, our model is only valid at low to moderate RNA concentrations, where our approximation that RNA behaves as a self-avoiding polymer (rather than exhibiting long-range repulsion) and the assumption that RNA does not compete for RBP–RBP binding sites are reasonable. In other words, the monotonic LLPS enhancement we observe is not expected to hold at high RNA concentrations.

To verify the preceding predictions obtained via the minimal model, we investigated the effects of PolyU RNA on the phase behavior of the ALS-associated RBP fused in sarcoma (FUS). FUS is an ideal example of an RBP that can both undergo LLPS on its own via homotypic interactions (17, 90, 91) and exhibit phase separation enhancement (i.e., undergo LLPS at low protein concentrations) in the presence of a low to moderate concentrations of RNA (92–94). Using a reparameterization (65) of the residue-level coarse-grained model of Dignon et al. (61) (Fig. 1D), we compute the phase diagram of the pure FUS system (Fig. 1E), and then probe the impact of adding moderate concentrations of PolyU [modeled using the parameters proposed in ref. (95)]. For the high-resolution simulations, each residue/nucleic acid is represented by a single bead (with associated hydrophobicity, mass, and charge); therefore, we are able to delineate which protein/RNA regions drive LLPS.

Consistent with experiments (17, 90, 91), we observe that FUS phase separates due to homotypic interactions—mainly via hydrophobic PLD–PPLD (PPLD = prion-like domain) interactions and cation–π interactions between Tyr residues in the PLD and Arg residues in RGG (RGG = Arg-Gly-Gly rich regions) domains (see SI Appendix Fig. S2B). Importantly, when we add increasing amounts of PolyU, we observe that the critical temperature of the mixture also exhibits a monotonic increase, by 0.8–2.4% for 1–3 chains of 175 nt (Fig. 1E), which implies that addition of RNA permits phase separation of FUS at higher temperatures, where the phase separation of FUS alone is unfavorable. Since temperature in the coarse-grained model directly impacts the relative strength of protein–protein interactions, this result also suggests that addition of RNA would permit the formation of FUS condensates under other unfavorable LLPS conditions (e.g., salt, pH, etc.). Unlike our minimal model, the residue-level coarse-grained model considers explicitly the electrostatic repulsion among RNA chains and the competition of RNA for specific scaffold–scaffold binding sites—e.g., it accounts for arginine-rich regions being utilized for FUS–FUS interactions, as well as for FUS–RNA interactions. Therefore, it can be used to investigate the role of higher concentrations of RNA in FUS condensates. These analyses reveal that higher RNA concentrations dissolve the FUS condensates (SI Appendix Fig. S3), as observed experimentally (21).

If we compare the phase diagrams computed with both models at low to moderate concentrations of RNA, we observe that both predict an increase in the size of the coexistence region upon insertion of RNA (Fig. 1B and E). However, the condensed branch of the residue-level coarse-grained phase diagrams (Fig. 1E) do not exhibit the marked increase in density upon insertion of the PolyU chains predicted by our minimal model (Fig. 1B). Although both models predict an enhancement in the connectivity of FUS proteins due to addition of PolyU (see Fig. 2B), our minimal model overestimates such enhancement. We note that the residue-level model likely describes the experimental phase behavior of FUS–RNA mixtures better because it accounts more adequately for the actual size, shape, flexibility, and chemical identity of the different molecular species involved in LLPS.

However, while our minimal model was designed to study general features of RBPs–RNA LLPS, it can be conveniently parameterized to study specific protein–RNA systems (such as FUS–RNA mixtures) more intimately. For example, whereas in the current study the RNA-like polymer interacts with RBPs in a nonspecific manner, selective binding can be introduced in the model via simple pairwise definitions. Despite the
Figure 2: RNA enhances the average valency of RNA-binding proteins. Systems in (A) and (B) are identical, in terms of composition, to those in Fig. 1B and E. (A) Average protein valency (bonds/$\sigma^3$) as a function of number of RNA chains ($n$), computed via the patchy-particle–polymer model at an inter-protein interaction strength (i.e., $E_{prot-prot}$) of 11 $k_B T$. The density of each system was first equilibrated to the coexisting density of the condensed-liquid branch of their respective phase diagram (i.e., without interfaces). The systems were then simulated in the $NVT$ ensemble to determine the average protein valency in the droplet. (Right panel) Zoomed-in snapshots of some RBPs within the simulation box showing differences in inter-protein connectivity. For clarity, the RNA-like polymers not depicted in the uppermost snapshot. (B) Inter-protein connectivity (bonds/atom$^3$); bonds correspond to inter-protein or protein–RNA contacts, as described in the SI Appendix) versus number of PolyU chains ($n$), computed via the reparameterized residue-level coarse-grained model (61, 65, 95) at 393 K. The systems were prepared as in (A); i.e., isotropic $NpT$ simulations, followed by $NVT$ ensemble computations to measure the droplet inter-protein connectivity. (Right panel) Zoomed-in snapshots of some FUS proteins within the simulation box showing differences in inter-protein connectivity. For clarity, the PolyU chains are not depicted in the uppermost snapshot. In (A) and (B) error bars (i.e., standard error) are of the same size or smaller than the symbols.

Limitations of a minimal approximation, the preceding residue-level study demonstrates the usefulness of a minimal model in capturing qualitative trends of RNA-driven LLPS at low to moderate RNA concentrations. Therefore, in what follows, we exploit the advantage of the minimal model to access the sufficiently large system sizes and long timescales needed to elucidate the thermodynamic and kinetic mechanism of RNA-driven RBP LLPS. We then present a comprehensive look at the regulation of condensate stability by RNA, and how this is modulated by the addition of different types of proteins to the RBP–RNA condensates.

Moderate RNA concentrations accelerate the nucleation and growth of RBPs condensates

Beyond increasing the size of the coexistence region by connectivity-enhancement, the question is: how is the thermodynamic and kinetic mechanism of RBP LLPS impacted by RNA? To investigate this, we define a “protein cluster size” order parameter $Q$, that measures the average number of proteins in a given protein cluster (i.e., $Q$ loosely describes the size of the emerging protein condensates) throughout our simulations. We first determine whether a protein is in a diluted or condensed region by employing a nearest neighbor criterion (96); wherein, proteins are assigned to condensed regions if, within a cutoff distance of 1.26 $\sigma$, they have at least three neighbors. Once all proteins are assigned, two proteins are considered to be part of the same condensed region (i.e., protein cluster) if they are separated by less than 1.26 $\sigma$ from each other. The size of the largest protein cluster ($Q_1$), therefore, probes the nucleation potential of the system. We first computed $Q_1$ at different inter-protein interaction strengths for the pure protein system and in the presence of one–three 40-mer RNA-like polymers (Fig. 3A). The critical protein–protein interaction strength (10.786 $k_B T$) for the pure protein system is indicated on the figure (i.e., vertical dashed line). Below this threshold (i.e., left of the vertical line), the largest cluster in the pure RBP system contains about 50 proteins. At a given inter-protein interaction strength, the size of the largest protein cluster increases monotonically as more RNA-like polymers are added. Strikingly, there is a 10-fold increase in the size of the largest protein cluster (ca. 600 proteins) upon adding three 40-mer RNA-like polymers at inter-protein interaction strengths where the pure RBP does not phase separate (i.e., below 10.786 $k_B T$). Additionally, when RNA is added, we obtain a gain (of approximately 1 $k_B T$) in the stability of protein clusters. It follows that RNA effectively recruits numerous protein molecules and increases the thermodynamic stability of protein clusters.

We then probe the evolution of the two largest protein clusters (denoted $Q_1 + Q_2$) as a function of time (Fig. 3B). In the early stages of condensation, protein clusters have a tendency to fuse and segregate often, and so the number of proteins in the two largest clusters is a more robust condensation
parameter than just $Q_1$. To obtain a homogeneous distribution of proteins in the simulation box, we initially zeroed all attractive interactions in each system. Then we activate all attractive interactions; setting the protein–protein interaction strength to a value where both systems undergo LLPS (i.e., $12 \, k_B T$). Therefore, $Q_1 + Q_2$ versus time directly measures the speed of condensate formation. We find that the two largest condensates form approximately four times faster in the presence of the polymer (i.e., 3x40-mer RNAs) than in the pure system. Furthermore, whereas the condensate grows in a roughly monotonic fashion in the RBP–RNA mixture (blue curve in Fig. 3B), condensate sizes fluctuate to a greater extent (black curve in Fig. 3B) in the absence of the RNA.

Our simulations explain the origin of these important differences in the kinetic mechanism of condensate formation in the absence and presence of RNA. In the absence of RNA, RBPs initially form many small protein-rich nuclei. Growth of these small nuclei is dependent on their capacity to outcompete one another for binding to the free RBPs that remain in the diluted phase, and/or are undergoing fusion among themselves. This competition, along with time needed for fusion of many clusters, effectively slows down the condensation process, and results in the initial lag-time we observe, before the exponential growth rate begins (black curve in Fig. 3B).

In contrast, when RNA is present, we observe formation of only a few dominant nuclei (i.e., as many clusters as there are RNA strands), which reduces competition and facilitates their growth and dominance. This results in RNA considerably accelerating the formation rate of the equilibrium condensate, and a negligible lag time (blue curve in Fig. 3B).

Therefore, in addition to increasing the stability of condensates, RNA promotes LLPS by accelerating the nucleation and growth of condensates. In the literature, Falahati et al. (25) proposed a seeding mechanism for the formation of the nucleolus. Specifically, they reported that rRNA transcription precedes nucleolus assembly (25), which subsequently recruits and localizes nucleolar proteins—fostering protein cross-linking and eventual condensation. Our results are consistent with these findings; demonstrating that RNA may drive RBP condensation by facilitating the formation of the first condensate nuclei (even in conditions where RBPs alone cannot nucleate), and also by accelerating the growth of such nuclei once formed (with respect to the rate of growth of RBPs alone).

### Ability of RNA to increase condensate stability depends on the composition of the system

Banani and coworkers (38) provided an initial framework to explain compositional control of membraneless organelles. Specifically, they demonstrated how low valency molecules (termed clients) can be recruited to condensates by binding to scaffolds (molecules that can phase separate on their own). They also highlighted that the valencies and relative concentrations of scaffolds and clients played significant roles in dictating the final droplet composition. In another study, Saha et al. (20) found that competition among certain proteins for RNA-binding sites altered P granule-like droplet composition.

Here, we investigate how RNA may affect condensate
stability in systems containing different types of proteins. In particular, we examine cooperation and competition effects in multicomponent mixtures comprising scaffolds, clients, and low concentrations of RNA. First, we consider the case of an intracellular mixture containing scaffolds and clients. Clients are defined as low valency proteins that cannot phase separate on their own (i.e., via homotypic client–client interactions); importantly, clients can be trafficked into the condensates by binding with high-affinity at the same sites used for homotypic scaffold–scaffold interactions. The protein composition of this mixture is 64% scaffolds (trivalent proteins; Fig. 4A) and 36% clients (divalent proteins; Fig. 4A)—i.e., the mixture contains a scaffold-to-client ratio of 1.78—and all inter-protein interaction strengths are equivalent. Notably, in comparison to the pure RBP (in Fig. 1B), this mixture shows reduced propensity to phase separate (i.e., a drop in the critical inverse protein–protein interaction strength by ~7%; see SI Appendix, Fig. S3). This result agrees well with the original scaffold–client model (38) predictions, and is further supported by previous computational (68) and experimental work (69).

For example, Nguema and Zhou (68) demonstrated that regulators of LLPS (analogous to clients in this case) may lead to suppression of driver (i.e., scaffolds) phase separation if regulator–driver binding affinities are comparable to driver–driver ones. In this scenario, regulators/clients displace drivers/scaffolds from the condensate at the expense of weakening the strength and connectivity of the percolating network (68, 72).

We then add RNA to the reference scaffold–client mixture, and investigate whether RNA enhances LLPS of this system. In particular, we consider the case where, in addition to scaffolds, clients bind to RNA with high affinity. To probe the effects of RNA, we simulate the mixture very close to the critical inverse protein–protein interaction strength of the reference client–scaffold mixture (i.e., 11.75 k_BT; see SI Appendix, Fig. S4 and Fig. 4B).

Upon equilibration, no LLPS is observed (Mixture I in Table 1; Fig. 4C); i.e., absence of sharp and well-defined interfaces. Interestingly, the RNA-like polymer is almost entirely coated by scaffolds (~84%), despite the binding strength of both client and scaffold proteins with the RNA being equal (Table 1).

This disparity in the types of proteins coating the polymer strongly suggests that the dissociation constants for the higher valency scaffolds (that can form more bonds with surrounding molecules) are smaller. Thus, higher valency RBPs have a greater probability of remaining bound to RNA (and bound in the percolating network).

Consequently, in a mixture containing both RNA-binding scaffolds and RNA-binding clients, RNA shows a preference for interacting with the scaffolds (even though in our model both scaffold–RNA and client–RNA bonds have the same interaction strength by construction). On similar grounds, scaffolds preferentially associate with adjacent scaffolds, creating a higher concentration of clients in the remaining mixture. We quantify the scaffold-to-client ratio in the remaining mixture (i.e., the effective ratio) by considering all proteins in the system, except those that are directly coating the RNA polymer. This quantity allows us to estimate the amount of scaffolds that are available for binding to (and recruiting) clients. In Mixture I, we find that the effective scaffold-to-client ratio decreases by ~7% with respect to the reference mixture. Our recent work (72) demonstrates that the addition of low-valency clients that are strong competitors for scaffold–scaffold interactions (like the ones in Mixture I) decreases the stability of condensates by diminishing the connectivity of the liquid network, and that this effect is amplified as the scaffold-to-client ratio decreases. Therefore, in the presence of strong-competing clients and with a limited amount of scaffolds, RNA indirectly decreases the connectivity of the condensed liquid by essentially reducing the effective scaffold-to-client ratio. Hence, the net effect is that droplet formation is not enhanced in the presence of RNA; indeed, the size of the largest protein cluster (~Q_1) decreases by ~3% when the RNA-like polymer is added to the mixture. Conversely, we find that when clients do not compete for scaffold binding sites a reduction in the effective scaffold-to-client ratio does not inhibit LLPS (SI Appendix, Fig. S5); since clients cannot replace LLPS-stabilizing scaffold–scaffold connections with LLPS-inhibiting scaffold–client ones.

Next, we test the effect of the RNA-like polymer on the phase behavior of a mixture of RNA-binding scaffolds and non-RNA-binding clients (i.e., the interaction strength between the client proteins and the RNA is set to zero; Mixture II in Table 1 and Fig. 4D). Clients compete with the RNA polymer (and with the scaffolds themselves) for binding to the scaffold proteins. However, clients are now unable to bind directly to the RNA polymer; hence, upon equilibration, ~99% of the molecules coating the RNA polymer are scaffolds. The interaction of scaffolds with RNA and adjacent scaffolds effectively decreases the ratio of available scaffolds to clients by ~13% with respect to the reference mixing ratio. There are effectively more clients in the remaining mixture that compete strongly with scaffolds for scaffold–scaffold binding sites; hence, the more LLPS-stabilizing scaffold–scaffold connections are replaced by connectivity-diminishing scaffold–client ones. Thus, the size of the largest protein cluster diminishes (i.e., ~Q_1 ~ -7%) and droplet formation is not enhanced (Mixture II in Table 1). In fact, the reduction in the size of the largest protein clusters (i.e., nucleation is suppressed) in Mixtures I and II suggests that the RNA inhibits phase separation in these systems.

Finally, we assess the effects of RNA on the phase boundaries of a mixture of non-RNA-binding scaffolds and RNA-binding clients (i.e., we switch off the attractive interaction between the scaffolds and the RNA; Mixture III in Table 1).

Addition of the RNA-like polymer increases the size of the largest protein cluster by ~7%. RNA enhances LLPS of this system (Fig. 4E) because it exclusively recruits clients and effectively sequesters them away from scaffolds. Accordingly, RNA decreases the availability of clients that would otherwise compete with scaffold self assembly. Hence, droplet formation...
Table 1: We start with a reference mixture containing 64% (768 trivalent proteins) scaffolds and 36% (432 divalent proteins) clients (i.e., a ref. scaffold-to-client ratio of 1.78) that can undergo LLPS. We then add a 40-mer RNA at an inter-protein interaction strength of $11.75 k_B T$ (i.e., close to the critical protein–protein interaction strength of the original scaffold–client mixture), modulate the cross interactions between the mixture components, and analyze whether the RNA polymer enhances LLPS reference of the mixture. To assess the ratio of scaffolds to clients remaining after coating the RNA, we consider all scaffolds and clients that are not directly bound to the RNA polymer. The change in this ratio is computed with respect to the 1.78 reference value. We also assess the change in $Q_1$ (number of scaffolds and clients in the largest cluster) to quantify the enhancement or inhibition of LLPS of the mixture upon adding RNA. Snapshots of observed phase behaviors are given in Fig 4.

<table>
<thead>
<tr>
<th>Mixture</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>client–scaffold interaction?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>client–RNA interaction?</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>scaffold–RNA interaction?</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>proteins coating RNA</td>
<td>84±0.5% scaffolds</td>
<td>99±0.5% scaffolds</td>
<td>1±0.5% scaffolds</td>
</tr>
<tr>
<td>scaffold-to-client ratio after RNA coating</td>
<td>1.65±0.01</td>
<td>1.55±0.01</td>
<td>2.27±0.01</td>
</tr>
<tr>
<td>change in scaffold-to-client ratio</td>
<td>-7±0.5%</td>
<td>-13±0.5%</td>
<td>+28±0.5%</td>
</tr>
<tr>
<td>variation in size of largest protein cluster ($\Delta Q_1$)</td>
<td>-3±1%</td>
<td>-7±1%</td>
<td>+7±1%</td>
</tr>
<tr>
<td>RNA enhances LLPS?</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

CONCLUSIONS

In this work, we have introduced a patchy-particle–polymer model capable of probing LLPS in protein–RNA mixtures from a mechanistic and thermodynamic point of view. Specifically, we have studied the effects of RNA: (i) on phase separation of model RBP s, (ii) on the thermodynamics and kinetics of condensate formation, and (iii) in the stability of multicomponent phase-separated protein–RNA mixtures. From our simulations, fundamental rules/features relating to LLPS in these systems emerge, at low to moderate RNA concentrations: (1) RNAs can act as high-valency molecules that promote phase separation by increasing the effective valency (i.e., connectivity) of RBPs (scaffolds); (2) RNA molecules increase the stability of condensates and accelerate the nucleation process; (3) addition of RNA to phase-separated scaffold–client (low-valency proteins) mixtures can lead to suppression of LLPS due to competition between RNA and clients for scaffold binding sites; (4) spatial segregation of components within phase-separated protein–RNA droplets is controlled by fine-tuning the effective ratio of scaffold to client proteins and interaction strengths of components. In particular, our work suggests that inhomogeneous cellular bodies may form at low RNA concentrations, when the RNA polymer exhibits a higher affinity for client proteins than for scaffolds.

Our findings provide an account of how low to moderate concentrations of RNA can significantly enhance phase separation of RNA binding proteins—enabling LLPS even under unfavorable conditions (e.g., super critical temperatures, core-shell) may arise. Hence, the present model may prove useful in advancing our understanding of LLPS into multiphase biomolecular condensates.
Figure 4: Composition of biomolecular condensates is regulated by the relative interaction strengths of molecular components. (A) Depiction of clients (divalent proteins), scaffolds (trivalent proteins) and a 40-mer RNA. (B) Reference client–scaffold mixture composed of 36% divalent proteins (clients) and 64% trivalent proteins (scaffolds) at an inter-protein interaction strength of 11.75 $k_BT$. (C)–(E) Phase behavior of client–scaffold–RNA mixtures (i.e., Mixtures I–III) described in Table 1.

By an EPSRC studentship (EP/N09620/1). R.C.-G. is an Advanced Research Fellow from the Winton Programme for the Physics of Sustainability. This work was performed using resources provided by the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service (http://www.hpc.cam.ac.uk/) funded by EPSRC Tier-2 capital grant EP/P020259/1.

REFERENCES


75. Tyler S. Harmon et al. “Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins”. eLife 6 (2017), e30294.


