

## Surfactants or scaffolds? RNAs of different lengths exhibit heterogeneous distributions and play diverse roles in RNA-protein condensates

Ignacio Sanchez-Burgos<sup>‡,1</sup>, Lara Herriott<sup>‡,1</sup>, Rosana Colleparado-Guevara<sup>2,3,\*</sup> and Jorge R. Espinosa<sup>1,\*</sup>

[1] *Maxwell Centre, Cavendish Laboratory, Department of Physics, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom.*

[2] *Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

[3] *Department of Genetics, University of Cambridge, Cambridge, CB2 3EH*

‡ = *These authors contributed equally to this work,* \* = *To whom correspondence should be sent.*

(Dated: November 9, 2022)

Biomolecular condensates, thought to form via liquid–liquid phase separation of intracellular mixtures, are multicomponent systems that can include diverse types of proteins and RNAs. RNA is a critical modulator of RNA-protein condensate stability, as it induces an RNA-concentration dependent reentrant phase transition—increasing stability at low RNA concentrations and decreasing it at high concentrations. Beyond concentration, RNAs inside condensates can be heterogeneous in length, sequence, and structure. Here, we use multiscale simulations to understanding how different RNA parameters interact with one another to modulate the properties of RNA-protein condensates. To do so, we perform residue/nucleotide-resolution coarse-grained Molecular Dynamics simulations of multicomponent RNA-protein condensates containing RNAs of different lengths and concentrations, and either FUS or PR<sub>25</sub> proteins. Our simulations reveal that RNA length regulates the reentrant phase behaviour of RNA-protein condensates: increasing RNA length sensitively rises the maximum value that the critical temperature of the mixture reaches, and the maximum concentration of RNA that the condensate can incorporate before beginning to become unstable. Strikingly, RNA of different lengths are organised heterogeneously inside condensates, which allows them to enhance condensate stability via two distinct mechanisms: shorter RNA chains accumulate at the condensate’s surface acting as natural biomolecular surfactants, whilst longer RNA chains concentrate inside the core to saturate their bonds and enhance the density of molecular connections in the condensate. Using a patchy particle model, we demonstrate that the combined impact of RNA length and concentration on condensate properties is dictated by the valency, binding affinity, and polymer length of the various biomolecules involved. Our results postulate that diversity on RNA parameters within condensates allows RNAs to increase condensate stability by fulfilling two different criteria: maximizing enthalpic gain and minimizing interfacial free energy; hence, RNA diversity should be considered when assessing the impact of RNA on biomolecular condensates regulation.

### I. INTRODUCTION

Intracellular organisation represents a fundamental aspect of regulation with respect to both structure and function. Whilst membrane-bound organelles are responsible for forming large, often permanent compartments within the cell, a more dynamic compartmentalisation can be also achieved through membraneless organelles [1]. Membraneless organelles, also referred to as biomolecular condensates, possess the two key properties of intracellular compartments: the existence of a defined boundary between the compartment and its surroundings, and the ability of components to diffuse freely within the compartment [2, 3]. Biomolecular condensates are thought to form via liquid-liquid phase separation (LLPS) of intracellular mixtures (e.g. proteins, RNA, DNA, and chromatin). Thus, the boundary of condensates is not a traditional lipid membrane, but rather the liquid–liquid interface separating a condensed liquid from its surrounding cytoplasm or nucleoplasm. Since the discovery of P-granules condensates in 2009 [4], important examples of biomolecular condensates including the nucleolus [5, 6], Cajal bodies [7, 8], paraspeckles [9, 10], stress granules [11, 12] and

chromatin [13, 14] (a finding which drove the paradigm shift away from the once prominent theory of the 30nm fibre [15–17]) have been exhaustively investigated.

RNA-binding proteins are common components of intracellular biomolecular condensates [18–22]. Several features of RNA-binding proteins underpin their ability to form condensates that are sensitively regulated by RNA. For instance, RNA-binding proteins are multidomain multivalent molecules, many of which can establish sufficiently strong homotypic interactions to act as scaffolds in biomolecular condensates—e.g., the heterogeneous nuclear ribonucleoprotein 1 (hnRNP1) [12, 23], fused in sarcoma (FUS) [24–26], the GTPase-activating protein SH3 domain-binding protein 1 (G3BP1) [27–30] and the Trans-activation response DNA-binding protein 43 (TDP-43) [31–33]. In addition, RNA-binding proteins can bind to RNA both specifically and promiscuously via their RNA-recognition motifs (RRMs), positively charged domains (PCDs), and intrinsically disordered regions (IDRs) with low-complexity amino acid sequences [34–36]. Both *in vitro* and, more recently, *in silico* experiments have demonstrated the role of RNA as a critical regulator of

RNA-protein condensates [37–44]. The RNA-binding protein Whi3 has been shown to partition into different condensates depending on the secondary structure of the RNA to which it is bound [45]. Proteins such as FUS remain soluble in the nucleus, where RNA concentration is high, but form aggregates in the cytoplasm where RNA concentration is lower [46]. Such impact of RNA concentration on protein aggregation may be relevant to rationalize the presence of pathological FUS aggregates (characteristic of Amyotrophic Lateral Sclerosis (ALS)) in the cytoplasm of post-mortem tissues [47] *versus* their absence from the nuclei [48]. The pattern of low levels of RNA promoting condensation *versus* higher concentrations promoting dissolution is described as RNA-driven reentrant phase behaviour [37, 43, 46] and is particularly important when considering the role of RNA in complex coacervation [44, 49–51]. Complex coacervation frequently enables the phase separation of so-called ‘cognate’ proteins which, unlike FUS, cannot sustain LLPS through protein-protein interactions alone, but instead rely on interactions with a partner biomolecule such as RNA [52]. The 25-repeat proline-arginine peptide (PR<sub>25</sub>) is a representative example of a protein which undergoes complex coacervation at physiological conditions driven mostly by electrostatic interactions with RNA [44, 51].

RNA length has been shown to mediate condensate reentrant phase behaviour [53]. Specifically, transcriptional condensate formation and dissolution was shown to be regulated by both RNA length and concentration: short, nascent RNAs present at transcription initiation stimulate condensation, whilst longer nucleic acids resulting from transcriptional bursts promote dissolution [53]. The nature of transcriptional bursting, with the total number of RNA molecules as well as their lengths increasing [54], means it is unclear whether the reentrant phase behaviour of transcriptional condensates is a function of RNA length, concentration, or a combination of both. In vitro experimentation has proved valuable in demonstrating the various ways in which RNA regulates LLPS of RNA-binding proteins [18–22, 37–41, 46]. Complementary, computational modelling and simulations can provide mechanistic insight into the experimental observations, and molecular detail regarding the condensate’s thermodynamic, kinetic and structural properties [55]. Computer simulations can also elucidate condensate properties such as droplet surface tension, protein molecular contact maps, or protein/RNA/DNA conformational ensembles [56–61]. Moreover, key features of LLPS, such as valency [62, 63], topology [64, 65] or binding affinity [66–68] can be precisely controlled. In that sense, Molecular Dynamics (MD) simulations have proved useful for the study of biomolecular condensates at various levels of resolution: ranging from atomistic force fields to lattice-based physical models [16, 51, 69–72].

Here, we use MD simulations, taking advantage of the benefits of computational modelling [56–61], to investigate in molecular detail the role of RNA length and concentration in the regulation of biomolecular condensates. Specifically, we aim to determine how RNA concentration and RNA length cooperate or compete to affect the RNA-dependent reentrant phase behaviour of RNA-binding proteins [18–20, 37, 38, 44, 46]. We use our sequence-dependent Mpipi model [73], which predicts protein phase diagrams in quantitative agreement with experiments, to study the reentrant phase behaviour of two archetypal proteins, FUS and PR<sub>25</sub>, known to undergo LLPS either by homotypic interactions or complex coacervation respectively [74]. Further, we investigate the effect of RNA length on condensate organisation by simulating proteins with mixtures of RNA of different lengths. In particular, we aim to identify whether patterns similar to those previously described with colloidal scaffold-surfactant models [63] can also exist in RNA-protein systems. Finally, by means of a minimal coarse-grained (CG) model [64], we prove that the observed RNA-length-and-concentration-dependent phase behaviour of RNA-protein condensates can also be observed in colloidal particles, and this is strongly regulated by molecular valency, binding affinity and polymer length.

## II. MODELS AND SIMULATION METHODS

Since the formation of phase-separated condensates entails the collective interactions among thousands of different proteins and other biomolecules, the study of LLPS has benefited from the development and application of coarse-grained approaches, including mean field simulations [75–79], lattice-based models [80–83], minimal models [65, 66], and residue-resolution simulations [57, 58, 67, 84–88]. In this article, we employ two protein/RNA coarse-grained models of distinct resolutions previously developed by us: (1) the residue/nucleotide-resolution Mpipi force field for proteins and RNA [73]; and (2) the MD-Patchy model in which whole proteins are represented as patchy particles, and RNA as self-avoiding flexible polymers [44, 64, 89].

Within the Mpipi force field [73], amino acids and RNA bases are represented by single beads with unique chemical identities (Fig 1(a)) in which hydrophobic,  $\pi$ - $\pi$ , and cation- $\pi$  interactions are modelled through a Wang-Frenkel (mid-range) potential [90], and electrostatic interactions via Yukawa/Debye-Hückel (long-range) potentials [84]. Bonded interactions between consecutive residues within the same protein (or nucleotides within the same RNA strand) are described with a harmonic potential. Furthermore, within this model, intrinsically disordered regions of proteins and RNA strands are treated as fully flexible polymers, whilst globular

domains are described as rigid bodies based on their corresponding crystal structures taken from the Protein Data Bank (PDB) and adapted to the model resolution. The interactions between ‘buried’ amino acids within globular domains are scaled down by 70% as proposed in Refs. [73, 74]. Finally, the physiological concentration of monovalent ions in solution (i.e.,  $\sim 150$  mM NaCl) is approximated by the screening length of the Yukawa/Debye-Hückel implicit-solvent model [84]. Further details on the force field parameters and simulation setups are provided in the Supplementary Material (SM).

In addition to the residue-resolution Mpipi model, we employ a minimal coarse-grained patchy model (MD-Patchy [64]) to elucidate whether the role of RNA length and concentration in RNA-protein condensates is determined by general molecular features such as valency, binding affinity or the relative RNA/protein length. Within our patchy particle simulations, proteins are described as pseudo hard-sphere (PHS) [91] particles decorated with sticky patches which account for the protein binding sites (modelled through square-well-like potentials [92]), and RNA strands are modelled as fully flexible self-avoiding polymers that can interact attractively with RNA-binding proteins via mid-range non-specific interactions [89]. Specifically, RNA self-avoiding polymers are described by a PHS potential in combination with a Yukawa/Debye-Hückel screened potential (for further details on the model potential and parameters see Section III C, and SM). With our minimal model, each RNA bead accounts for several nucleotides and has the same size as those of the proteins [89]. Moreover, as in the residue-resolution coarse-grained model, an implicit solvent is used. Accordingly, the diluted phase (i.e., the protein-poor liquid phase) and the condensed phase (i.e., the protein-rich liquid phase) are effectively a vapour and a liquid phase, respectively.

To determine the stability of RNA-protein condensates, we evaluate the phase diagrams (in the temperature–density plane) of the different systems by means of Direct Coexistence simulations [93, 94]. Within the Direct Coexistence approach (Fig. 1(a)), the two coexisting phases of a given system are placed in the same simulation box, which is a rectangular box with an elongated side perpendicular to the interfaces—long enough to capture the bulk density of each phase—while the parallel sides are chosen such that proteins and RNA cannot interact with themselves across the periodic boundary conditions [43]. We employ the canonical ensemble (constant number of molecules ( $N$ ), system volume ( $V$ ) and temperature ( $T$ ), or  $NVT$ ). Once Direct Coexistence simulations reach equilibrium, we measure the coexisting densities of both the diluted and condensed phases along the long axis of the box, excluding the fluctuations at the interfaces and keeping the center of mass of the system fixed. By repeating this procedure at different temperatures—until we reach supercritical

temperatures, i.e., where no phase separation is observed any longer—we can evaluate phase diagrams (Fig. 1(b)). Finally, to avoid finite system-size effects close to the critical point, we estimate the critical temperature ( $T_c$ ) and density ( $\rho_c$ ) using the law of critical exponents and rectilinear diameters [95] (as shown in Refs. [64, 89]). Fig 1(a) shows a Direct Coexistence simulation with a system composed by PR<sub>25</sub> and poly-Uridine (polyU) RNA strands of 400 nucleotides (nt) and FUS with polyU strands of the same length at conditions in which both systems undergo LLPS.

Furthermore, to evaluate condensate densities while controlling accurately the concentrations of individual components inside them, we perform additional simulations in the isothermal-isobaric ensemble ( $NpT$ ). In contrast, fixing the desired composition of a multicomponent condensate, e.g. the protein/RNA proportion, within the condensed phase in Direct Coexistence simulations is far from trivial. As shown in Fig. 1(b),  $NpT$  simulations provide a similar representation of the condensed phase obtained by Direct Coexistence simulations (although avoiding the effects of interfaces). By fixing the system’s pressure to zero, and enabling the volume of the simulation box to isotropically fluctuate, we allow the condensed phase to equilibrate. Stable phase-separated condensates exhibit an equivalent coexisting density as that obtained by Direct Coexistence simulations, whereas unstable systems cannot sustain the condensed phase and tend towards infinitely dilute densities. Within the  $NpT$  ensemble, condensates are only stable at temperatures where the interactions between biomolecules are sufficiently strong to overcome the entropic cost of forming a percolating liquid network without the need for pressure to be exerted on the simulation box. Finally, once the systems’ densities are equilibrated, the critical temperature of each condensate can be estimated within the interval between the highest temperature at which the system is stable and the lowest temperature at which it is not. A comparison between the condensate coexistence densities and critical temperatures obtained via  $NpT$  vs. Direct Coexistence simulations for mixtures of polyU-PR<sub>25</sub> with different strand lengths is shown in Fig. 1(b).

### III. RESULTS AND DISCUSSION

#### A. RNA length modulates the RNA concentration-dependent reentrant phase behaviour of RNA-protein condensates

We first investigate how the RNA-concentration-dependent reentrant phase behaviour of RNA-protein condensates is influenced by the length of single-stranded RNA. For this, we perform DC and bulk  $NpT$  simulations using the residue/nucleotide-resolution Mpipi

force field [73], which has been shown to achieve quantitative agreement with experimental phase diagrams of RNA-binding proteins. We compare the behaviour of two phase-separating RNA-binding proteins, FUS (526 residues, sequence in Supplementary Material) and PR<sub>25</sub>, for which phase behaviour is modulated differently by RNA [44]. FUS can form single-component condensates via homotypic interactions, which increase in stability at moderate RNA concentrations [37, 43, 46]. PR<sub>25</sub> is an arginine-rich peptide, which requires RNA to phase separate via heterotypic RNA–protein interactions at physiological conditions [44, 51]. For each case, we simulate solutions containing tens to hundreds of individual proteins in the presence of disordered single-stranded polyU RNA molecules with five different lengths: 20, 50, 100, 200, and 400 nucleotides. For all RNA lengths, we test different polyU concentrations defined through the U/protein mass ratio, which allows us to quantify the total number of U nucleotides in the mixtures, regardless of whether they are assembled in longer or shorter polyU chains. Importantly, it has been shown that combining RNA-binding proteins and RNA at ratios resulting in electroneutral mixtures enhances the stability of RNA-protein condensates [42, 43]. Thus, here we explore a range of U/protein mass ratios that lie around the electroneutral point. We focus on single-stranded polyU RNA for simplicity and to follow previous works on RNA-protein phase separation [18, 37, 74].

For the polyU-PR<sub>25</sub> mixtures, the electroneutral point lies at the 1.21 U/protein mass ratio. Thus, we perform simulations for polyU-PR<sub>25</sub> mixtures spanning the range between 0.78 and 2.24 U/PR<sub>25</sub> mass ratios. For the electroneutral polyU-PR<sub>25</sub> system, we first demonstrate that  $NpT$  simulations quantitatively reproduce the condensed phase coexistence densities found in phase diagrams constructed using standard  $NVT$  Direct Coexistence simulations [64, 84]. From a set of Direct Coexistence simulations at varying temperatures, we extract the phase diagrams for two polyU-PR<sub>25</sub> mixtures containing RNA strands of varying lengths (20 and 400 nucleotides) but keeping a constant U concentration (mass ratio of 1.21) regardless of RNA length (Fig. 1(b)). Then, we simulate these systems in the  $NpT$  ensemble at  $p=0$  (using a cubic box). As shown in Fig. 1(b), this approach provides consistent condensate densities to those obtained via Direct Coexistence simulations (as long as the density of the dilute phase is very low). As our Direct Coexistence simulations indicate, this assumption is reasonable for most of the temperatures with densities of the dilute phase being of the order  $1 \times 10^{-3}$  g/cm<sup>3</sup> (please note that the solvent is implicitly considered within the force field). Since with  $NpT$  simulations the density of the dilute phase cannot be measured, the value for the critical temperature cannot be calculated using the law of critical exponents and rectilinear diameters [95], as is the case in the Direct Coexistence simulations. However, an interval at which

the critical temperature falls can be estimated. Figure 1(b) shows that the values of the critical temperatures estimated from bulk  $NpT$  simulations (as the mid temperature of the computed interval; described in Section II) lie within the uncertainty of the critical temperatures evaluated from Direct Coexistence simulations and the law of critical exponents and rectilinear diameters.

Having established that  $NpT$  simulations can provide robust estimates of the critical temperature, we move forward with the  $NpT$  ensemble to investigate the interplay between RNA concentration and the effects of RNA length on condensate stability. To consider the effect of RNA length independently, we keep the total amount of U nucleotides in the mixtures fixed, and assemble them in polyU chains of different lengths. When the U/PR<sub>25</sub> mass ratio is kept constant, we observe a monotonic increase in the critical temperature of the condensates as the RNA length increases. For example, for polyU-PR<sub>25</sub> mixtures at ratios satisfying the electroneutral point (i.e., 1.21 polyU/PR<sub>25</sub> mass ratio), there is a 20% enhancement in the critical temperature when the RNA chain length increases from 20 to 400 nucleotides (Figure 1(c)). When we next fix the RNA chain length, and investigate the impact of RNA concentration on the stability of polyU-PR<sub>25</sub> condensates, for all the RNA lengths we study, we confirm that PR<sub>25</sub> exhibits the well-known RNA-concentration dependent reentrant behaviour of RNA-binding proteins discovered experimentally [22, 37, 46]. That is, the stability of polyU-PR<sub>25</sub> condensates—quantified by the values of the critical temperature—gradually increases as the RNA concentration goes from low to moderate (up to approximately the electroneutral point, at 1.21 mass ratio), then it reaches a maximum value, and finally decreases as the RNA concentration increases even further. Remarkably, looking at the combined effects of RNA length and concentration reveals that RNA length significantly modulates such reentrant behaviour. Specifically, increasing RNA length sensitively raises the maximum value that the critical temperature of the mixture reaches (i.e. how much the condensate stability can be boosted by RNA), and the maximum concentration of U nucleotides that the condensate can incorporate before beginning to become unstable (i.e., when the RNA chains are longer, more nucleotides can form part of the condensate before it begins to dissolve). Thus, the most stable polyU-PR<sub>25</sub> condensates are formed by the longest RNA chains we study, and, unexpectedly, contain U/PR<sub>25</sub> concentrations above the electroneutral point (Fig. 1(c)).

We next investigate whether this behaviour also holds for proteins that are able to undergo phase separation on their own, i.e., via homotypic protein–protein interactions [65, 96]. For this, we focus on the protein FUS, and test the impact of adding polyU of varying lengths and at different concentrations. Our simulations contain 48

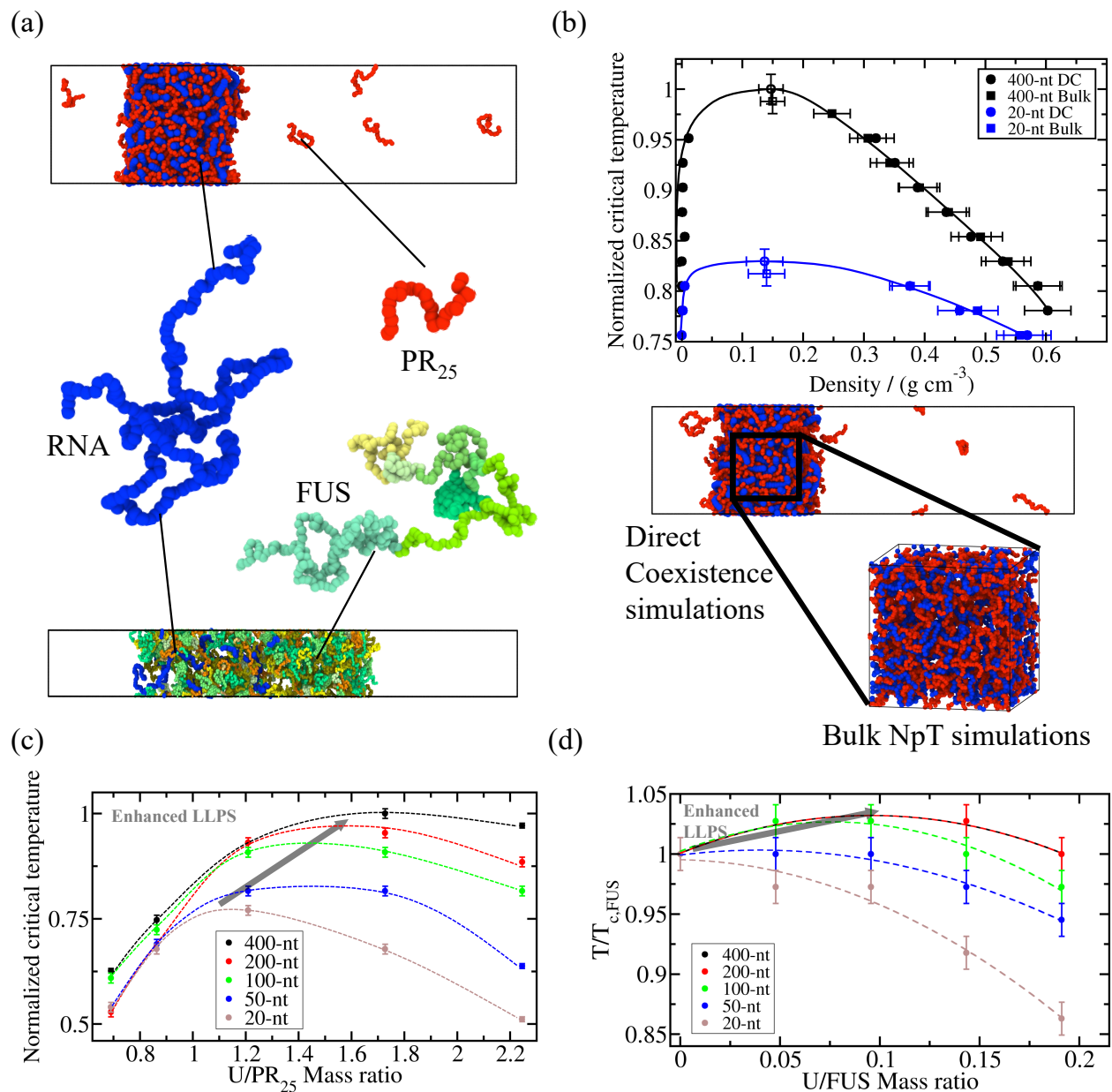


FIG. 1: Reentrant phase behaviour driven by RNA is regulated by both concentration and length. (a) Residue-resolution coarse-grained simulations with the Mpipi model [73] to investigate phase-separation of RNA-protein mixtures. Coarse-grained representation of (full-sequence) FUS, PR<sub>25</sub>, and a 400-nt polyU RNA strand using the Mpipi model [73] in which each amino acid or nucleotide is represented by a single bead. Please note that the size of the beads depicted in this panel has been conveniently rescaled for visualization purposes. In FUS protein, beads of different colours indicate distinct protein domains. Direct Coexistence simulations of polyU-PR<sub>25</sub> (Top) and FUS-polyU (Bottom) are also included. (b) Comparison of the predicted condensate densities as a function of temperature (renormalized by the highest critical temperature) for polyU-PR<sub>25</sub> mixtures composed by polyU strands of 400-nt (black symbols) and 20-nt (blue symbols) using Direct Coexistence simulations (filled circles) and Bulk *NpT* simulations (filled squares). The estimated critical temperature of each system by both ensembles is depicted by empty symbols of the corresponding shape and colour. Snapshots of a Direct Coexistence simulation and a Bulk *NpT* simulation are included to illustrate the analogy between both ensembles when describing the system condensed phase. (c) Normalized critical temperature of polyU-PR<sub>25</sub> mixtures as a function of the U/PR<sub>25</sub> mass ratio for different polyU strand lengths as indicated in the legend. (d) Normalized critical temperature of FUS-polyU mixtures as a function of the U/FUS mass ratio for different polyU strand lengths as indicated in the legend. While in Panel (c) all temperatures have been normalized by the highest *T* at which phase separation was observed (*T* = 435K), in Panel (d) all temperatures have been normalized by the critical temperature of pure FUS (*T*<sub>c,FUS</sub> = 365K). Please note that higher critical temperatures in our model correspond to higher driving forces to undergo LLPS (i.e., lower saturation concentration).

copies of FUS (full-sequence; see SM) and polyU chains of either 20-nt, 50-nt, 100-nt, 200-nt or 400-nt in length. For all the different polyU lengths, we prepare mixtures at concentrations spanning the range of U/FUS mass ratio from 0 to 0.19. First, when we look at mixtures containing RNA at the concentration corresponding to the electroneutral point (mass ratio of 0.049), we confirm that performing  $NpT$  bulk simulations reproduces the length-dependent increase in critical temperature identified previously via Direct Coexistence simulations [44]. Specifically, when we mix FUS with RNA of either 100-nt, 200-nt or 400-nt at the electroneutral ratio (0.049 U/FUS mass ratio), we see a marginal increase in critical temperature—3% with respect to the value for pure FUS system (Figure 1(d))—in agreement with our previous Direct Coexistence results [44]. Adding 20-nt polyU to FUS, the shortest polyU molecules we study, hinders phase separation, with the effect being amplified at higher polyU concentrations. This occurs because 20-nt polyU is too short to bridge FUS molecules and enhance the connectivity of the condensed liquid, as shown previously [43, 44]. For RNA lengths of 50-nt, 100-nt, 200-nt and 400-nt, the RNA concentration-dependent reentrant phase behaviour of FUS is observed. That is, gradually increasing the concentration of polyU up to a given threshold (in our case a U/FUS mass ratio of around  $\sim 0.1$ ) increases the critical temperature, whilst adding polyU at concentrations beyond such threshold reduces the critical temperature. The observation of peak critical temperatures for FUS condensates containing 50-nt polyU at polyU concentrations surpassing the electroneutral point (i.e.,  $\sim 0.1$ ), is consistent with *in vitro* studies [20]. As in PR<sub>25</sub>/polyU mixtures, we find that the maximum enhancement in phase separation occurs at higher polyU/protein mass ratios for longer chain polyU systems. Overall, when polyU length is increased from 20 to 400 nucleotides, the mass ratio at which the system displays the highest critical temperature increases from 0 (no RNA) to 0.12 (Figure 1(d)). Nevertheless, we observe that the RNA length-dependent effects for FUS/polyU mixtures are significantly smaller than for polyU-PR<sub>25</sub> mixtures (see Figs. 1(c) and 1(d)).

Our results for FUS are consistent with a number of *in vitro* studies which identified similar patterns of RNA concentration-dependent reentrant phase behaviour [41, 46]. However, that low levels of RNA promote phase separation, whilst higher RNA levels promote dissolution contrasts with the observation that some condensates are able to form in regions of the cell with very high levels of RNA (up to RNA/protein mass ratios of  $\sim 40$ ) [97]. Moreover, despite RNA being at high concentrations, it has been shown to play a key role in initiating phase separation of FUS by nucleating protein condensates [20, 46, 97]. For example, RNA plays a prominent role in triggering P-bodies formation, which occurs as a result of a strong increase in

mRNA concentration [97, 98]. Also, FUS-containing paraspeckles in the nucleus are thought to be nucleated by the long non-coding RNA Neat1, despite the high background RNA concentration [20, 46]. Even though *in vitro* studies have not explicitly revealed how the length of RNA defines the maximum RNA concentration that a condensate rich in RNA-binding proteins can take before its stability begins to decrease, there is evidence which indirectly supports this theory: a number of *in vitro* and *in silico* studies have identified that minimum RNA lengths are needed to promote phase separation of RNA-binding proteins [43, 99] and others have found that shorter RNA molecules are more potent promoters of condensate dissolution [46], in agreement with our results from Figure 1(d). Here, we show directly that increasing the length of the nucleic acid chain increases the capacity of the RNA to promote phase separation and to be incorporated into the condensate at higher RNA concentrations than their shorter counterparts. Therefore, previous findings which have shown that high concentrations of RNA promote condensate dissolution ought to be contextualised with information on the length of the studied nucleic acids. As demonstrated here, it is possible for phase separation to be favoured at high RNA concentrations if the RNA chains are long enough to bind multiple proteins simultaneously. Assembling nucleotides in longer RNA chains has the additional advantage of decreasing the electrostatic repulsion among phosphates of separate nucleotides, due to the presence of more covalent bonds.

Since our results support the hypothesis that longer RNA molecules are more powerful enhancers of phase separation due to their ability to increase the connectivity of the condensed liquid [66, 96], we next investigate this behaviour by examining the stability of condensates containing both long and short polyU chains (Fig. 2(a)). To do so, we simulate mixtures of PR<sub>25</sub> in the presence of polyU strands of two different lengths, 50-nt and 400-nt, such that each length represents half of the total polyU mass ratio concentration in each system. For these mixed RNA length systems, we explore polyU/PR<sub>25</sub> mass ratios ranging from 0.76 to 2.1. We find that mixed RNA length systems display critical temperatures in between those of the pure short and long RNA chain systems, and with the maximum enhancement of critical temperature also occurring at intermediate polyU/PR<sub>25</sub> mass ratios between that of the pure short and pure long RNA systems (Figure 2(b)). We also study FUS-polyU mixtures of 50-nt and 400-nt polyU strands (such that each length represents half of the total polyU mass ratio concentration) spanning mass ratios from 0 to 0.196. Similar to the results for PR<sub>25</sub>, the mixed length systems in FUS present intermediate critical temperatures compared to the pure long-chain and short-chain RNA systems, as well as patterns of reentrant phase behaviour in which the maximum critical temperature peaks at an intermediate polyU/FUS mass ratio between that of the

pure long-chain and short-chain RNA systems (Figure 2(c)).

Our findings reveal that mixed length polyU systems are generally more stable than short-chain polyU systems; further emphasising the ability of long polyU strands to show a stabilising effect on condensates up to higher concentrations, given that we observe this behaviour even when shorter polyU molecules are also present. Our observations are consistent with *in vitro* FUS/RNA phase separation assays, reporting that the lncRNA Neat1 (short isoform, length=3.7kb) was able to drive reappearance of FUS droplets which had previously been solubilised with tRNA (70-100 nucleotides [100]) [46].

### B. RNA length and concentration determine the internal organisation of molecules in RNA-protein condensates and the properties of their interfaces

To rationalise the interaction of RNA length and concentration in modulating the stability of RNA-protein condensates, we now characterise the organisation of the different molecules inside the RNA-protein condensates by quantifying the densities of polyU *versus* proteins across the condensate. These density profiles reveal that there are considerable differences in the distribution of proteins *versus* polyU depending on the concentration of RNA. Protein-rich systems (i.e., low polyU/PR<sub>25</sub> mass ratios) form condensates with a surface coated by PR<sub>25</sub> peptides (Fig. 3(a-b)), whereas the surfaces of polyU-rich droplets (i.e., high polyU/PR<sub>25</sub> mass ratios) are mostly coated by polyU chains (Fig. 3(c-d)).

In polyU-PR<sub>25</sub> condensates combining RNAs of different lengths, we find that the RNAs exhibit a striking spatially heterogenous distribution. The longer 400-nt polyU molecules accumulate in the condensate core, while the shorter 50-nt polyU molecules concentrate preferentially towards the condensate surface (Figs. 3(a) and 3(c)). Simulations approximating proteins as patchy colloids have revealed the same heterogenous organisation of high and low valency proteins within multicomponent condensates. By looking at the problem from a condensed matter perspective, the simulations revealed that burying high valency proteins in the centre, and exposing low valency species to the interface, maximises the enthalpic gain for condensate formation (most bonds of the higher valency molecules are satisfied at the condensate core) and reduces the interfacial free energy at the droplet interface because the lower valency molecules act as surfactants [63, 66, 101]. In Figure 3(e), we also show that polyU-FUS condensates of mixed RNA length (with equal concentration of 50-nt and 400-nt strands) display multilayered RNA organisation, preferentially locating short polyU chains at the interface and long RNA strands in the core. Such

structural organisation maximizes at the same time the condensate liquid network connectivity and minimizes the interfacial penalty.

Since distinct molecules at the condensate interface can translate into significantly different interfacial properties, such as interfacial free energies [22, 43], coalescence fusion rates [102], size-conservation [6, 63], or uneven molecule exchange rates [66], we now calculate the interfacial free energy ( $\gamma$ ; see SM for further details on these calculations). We focus on PR<sub>25</sub> condensates because converging the value of  $\gamma$  for FUS-based droplets is computationally unfeasible due to the size of FUS (526 amino acids). We start by computing the interfacial free energy for two types of 400-nt polyU-PR<sub>25</sub> condensates: with a PR<sub>25</sub>-rich interface (i.e., low U/PR<sub>25</sub> mass ratio of 0.8) and with a polyU-rich interface (i.e., high U/PR<sub>25</sub> mass ratio of 1.6), as shown in Figure 3(f). To facilitate the comparison, in both cases we simulate the systems at the same temperature ( $T/T_c=0.85$ ). Strikingly, we find that the condensate with a polyU-rich interface presents a much lower interfacial free energy (almost half) than that with a PR<sub>25</sub>-rich interface. Such an unexpected finding explains how a high concentration of polyU, beyond the electroneutral point, can boost the stability of condensates: by incorporating a large concentration of RNA, condensates experience a trade-off between the destabilising effect of a decreased enthalpic gain due to the larger electrostatic repulsion among equally charged nucleotides, and the stabilising effect of the decrease in the condensate energetic penalty of forming an interface when it is coated with polyU. Our results, therefore, reveal that in condensates with an excess of polyU, polyU behaves like a surfactant.

To analyse the impact of RNA length, we now calculate the interfacial free energy of two different condensates containing the shorter 50-nt polyU RNA: a PR<sub>25</sub>, 50-nt and 400-nt polyU condensate, and a PR<sub>25</sub> and 50-nt polyU condensate. We estimate the interfacial free energy at the same temperature used above ( $T/T_c=0.85$ ) and at the lower U/PR<sub>25</sub> mass ratio of 0.8 ( $T/T_c=0.85$ ), since imposing a high RNA concentration with short polyU molecules within the condensed phase in a Direct Coexistence simulation is not feasible. We observe that increasingly adding 50-nt polyU molecules progressively decreases the interfacial free energy of the condensate, with respect to the value of the 400-nt polyU-PR<sub>25</sub> condensate (Fig. 3(f)). This observation reinforces the idea that in RNA-protein condensates that contain RNAs of different lengths, positioning the shorter RNAs species at the interface reduces the surface tension as such shorter RNAs act as better surfactants than longer ones [103]. Hence, the advantage of mixed length RNA condensates exhibiting multilayered organization results from them presenting similar low surface tensions as those only formed by short RNA strands, while showing considerably higher

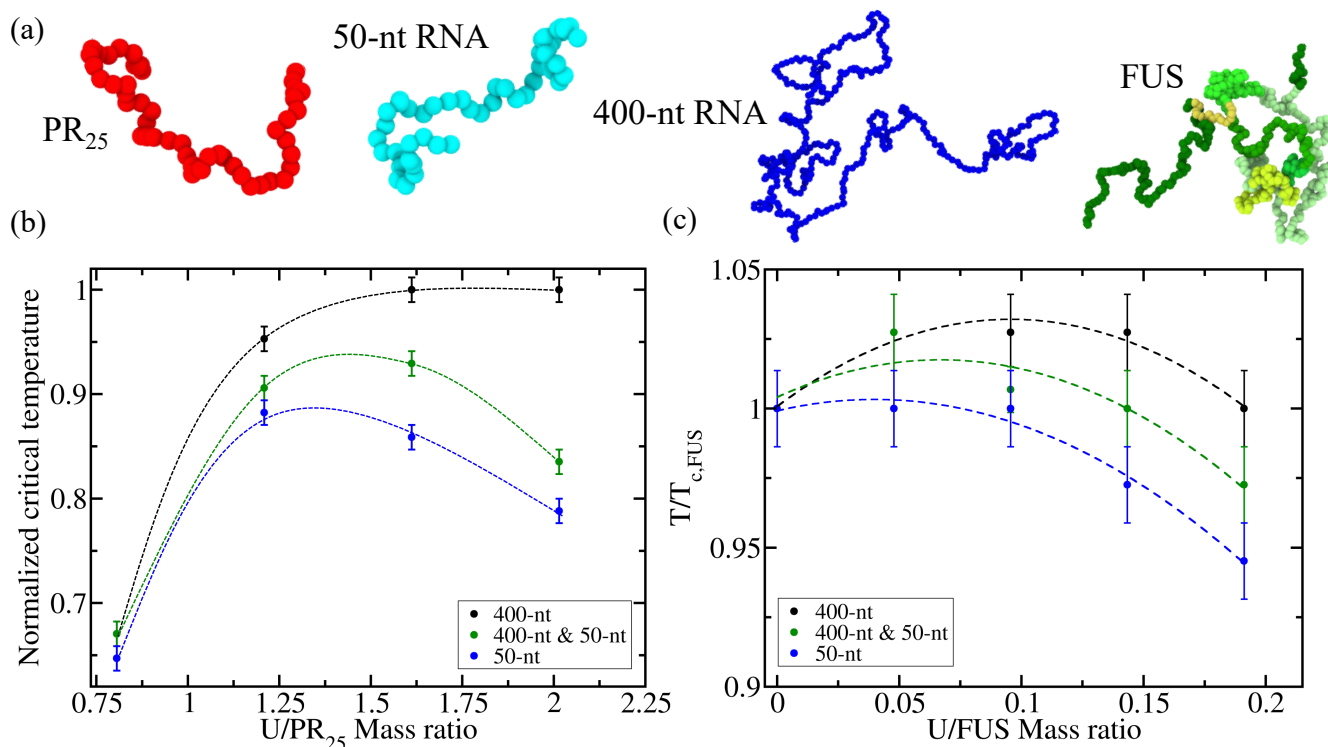


FIG. 2: RNA-driven reentrant phase behaviour of FUS and PR<sub>25</sub> polyU systems including mixtures with different RNA strand lengths. (a) Representation of PR<sub>25</sub>, FUS, and two RNA strands of 50 and 400-nucleotide each following the same colour code and considerations discussed in Fig. 1(a). Please note that the size of the beads depicted in this panel has been conveniently rescaled for visualization purposes. (b) Normalized critical temperature of polyU-PR<sub>25</sub> mixtures as a function of the U/PR<sub>25</sub> mass ratio for different polyU strand lengths as indicated in the legend. (c) Normalized critical temperature of FUS-polyU mixtures as a function of the U/FUS mass ratio for different polyU strand lengths as indicated in the legend. For the systems with mixed polyU lengths, each length represents half of the total polyU concentration. While in Panel (b) all temperatures have been normalized by the highest  $T$  at which phase separation was observed ( $T=425\text{K}$ ), in Panel (c) all temperatures have been normalized by the critical temperature of pure FUS ( $T_{c,FUS}$ ).

stability due to long RNA strands increasing the enthalpic gain for condensate formation by contributing more connections to the liquid network connectivity together with PR<sub>25</sub>.

Multilayered condensates such as those found in Figs. 3(a), (c) and (e) for PR<sub>25</sub> and FUS-polyU mixtures respectively can be found across the cell and include FUS-containing paraspeckles [9], stress granules [30] and the nucleolus [6]. Indeed, condensate structure and organisation has important implications for the behaviour of the various components, with those located in the core exhibiting slower exchange rates compared to molecules in the outer shell [66, 96]. Experiments and simulations reveal that longer RNA strands lead to higher viscosities in RNA-protein condensates [18–20, 43, 104]. Thus, in mixed RNA condensates, a stable core containing long polyU strands is expected to have a higher viscosity, while an outer shell with short polyU strands a lower viscosity [43]. Gelification of RNA-protein condensates

via fibrillation, with potential pathological implications [21, 47, 105], has been shown to be seeded at the interface due to a local increase of protein density at the surface [106]; hence, incorporating short RNAs into RNA-protein condensates might contribute to preventing their maturation because it decreases the probability of high density fluctuations at the interface.

### C. Self-avoiding polymers trigger concentration-dependent reentrant phase behaviour of colloidal patchy-particle condensates modulated by polymer length

To investigate whether the observed patterns of RNA-length-and-concentration-dependent phase behaviour found in RNA-protein condensates rely on general molecular features such as protein valency, binding affinity, or the relative polymer size/length between proteins and RNA, we employ a minimal CG model of colloidal patchy particles with self-avoiding polymer



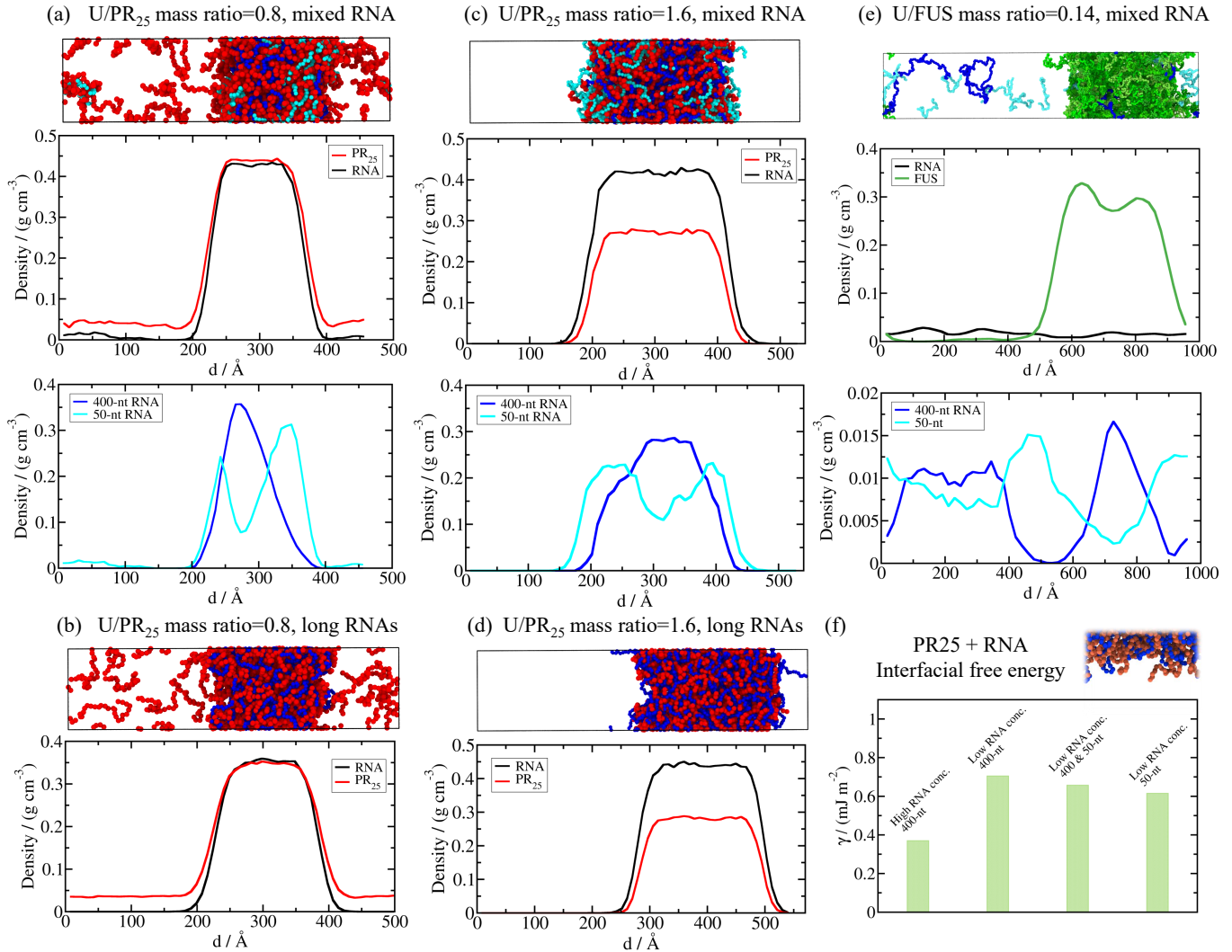


FIG. 3: Structural condensate organization of RNA-binding proteins in the presence of long and short RNA strands. (a) PolyU-PR<sub>25</sub> mixture with a 0.8 U/PR<sub>25</sub> mass ratio where polyU strands are 50-nt and 400-nt long (each length contributing half to the total polyU concentration). Top: Representative snapshot of a Direct Coexistence simulation of the system, where PR<sub>25</sub> molecules are coloured in red and long and short RNAs are depicted in blue and cyan respectively. Middle: Density profile of PR<sub>25</sub> (red) and polyU RNA (black) along the long axis of the simulation box. Bottom: RNA density profile decomposed in 400-nt (blue) and 50-nt (cyan) polyU chains. (b) The same as in Panel (a), but for a polyU-PR<sub>25</sub> mixture with a U/PR<sub>25</sub> mass ratio of 0.8 and polyU strands of 400-nt only. Note that we only show one density profile since here all RNAs are of equal length. (c) The same as in Panel (a), but for a polyU-PR<sub>25</sub> mixture with a U/PR<sub>25</sub> mass ratio of 1.6, where polyU strands are also 50 and 400-nt long. (d) The same as in Panel (b), but for a polyU-PR<sub>25</sub> mixture with a U/PR<sub>25</sub> mass ratio of 1.6 and polyU strands of 400-nt only. The temperature of systems shown in Panels (a)-(d) was 0.9 with respect to their corresponding critical temperature. (e) FUS-RNA mixture at  $T/T_{c,FUS}=0.98$  and with a U/FUS mass ratio of 0.14, where polyU RNA strands are 50-nt and 400-nt long (each length contributing half to the total polyU concentration). Top: Representative snapshot of a Direct Coexistence simulation. Middle: Density profile of FUS (green) and RNA (black) along the long axis of the simulation box. Bottom: RNA density profile decomposed in 400-nt (blue) and 50-nt (cyan) polyU chains. (f) Surface tension for different polyU-PR<sub>25</sub> mixtures, all of them at a temperature of 0.85 with respect to its corresponding critical temperature for the system indicated in the legend. The high RNA concentration corresponds to a U/PR<sub>25</sub> mass ratio of 1.6, while the low RNA concentration corresponds to a U/PR<sub>25</sub> mass ratio of 0.8.

chains to mimic proteins and RNA respectively [89]. Iterations of this minimal coarse-grained model have been previously used to investigate critical factors in LLPS such as surface tension, droplet size conservation, or condensate substructure [44, 66]. The aim of our minimal simulations here is to assess, beyond protein sequence and specific molecular features, the thermodynamic parameters that explain the general differences between the impact of RNA length and concentration on homotypic phase separation *versus* RNA-protein heterotypic complex coacervation.

We start by computing the phase diagrams of two different types of colloidal patchy particles in the presence of different concentrations and lengths of a self-avoiding polymer that mimics RNA. The first type are patchy particles decorated with 3-binding sites in a planar arrangement separated by 120 degrees angles (Fig. 4(a); see SM for further details on the model). Like FUS, these colloidal particles—referred to henceforth as scaffold proteins—are able to phase separate on their own via homotypic interactions below a reduced temperature of  $T^*=0.09$  [89] (see details on reduced units in the SM). On the other hand, the second type of patchy colloids possess 2-binding sites in a polar arrangement, which by construction can only form linear chains and not 3-dimensional percolated networks that sustain phase-separation [62, 64, 65] (Fig. 4(a)). Like PR<sub>25</sub>, 2-binding site colloidal particles—referred to henceforth as cognate proteins—cannot phase separate on their own [44]. We perform Direct Coexistence simulations of both scaffold and cognate proteins for different polymer-bead/protein ratios using polymer chains of 10, 20, 50 and 100 beads such that the polymer-bead/protein ratio is defined in terms of the number of polymer and protein beads, ranging from 0.2 to 1.

First, by focusing on a given polymer-bead/protein ratio (i.e., 0.4), we capture the length-dependent enhancement of LLPS reported in our sequence-dependent CG simulations of PR<sub>25</sub> and FUS (Figure 1) where systems containing longer RNA strands can phase-separate up to higher critical temperatures (Fig. 4(b-c)). For cognate colloidal particles, increasing the ‘RNA’ chain length from 10 to 100 beads at an RNA/protein ratio of 0.4 leads to a 50% increase in critical temperature (Fig. 4(b)). In contrast, the enhancement in critical temperature for the same polymer/protein concentration with scaffold proteins is just 8% (Fig. 4(c)). Clearly, scaffolds show much more subtle dependency on RNA length compared to cognate proteins, in full agreement with our previous sequence-dependent simulations shown in Figs. 1 and 2.

Secondly, by gradually varying the polymer-bead/protein ratio (from 0.2 to 1; Fig. 4(b-c)), patterns of concentration-dependent reentrant phase behaviour for both types of proteins and distinct RNA polymer

lengths can be observed. A moderate increase in the polymer-bead/protein ratio leads to an enhancement in the critical temperature, while any further increase in self-avoiding polymer levels results in LLPS inhibition (i.e., lower critical temperatures). For the cognate protein mixtures, we find a significant shift in the ratio which gives the highest critical temperature as a function of polymer length; consistent with our results for PR<sub>25</sub> with polyU (Fig. 1(c)). On the other hand, for the scaffold protein mixtures, the shift in the maximum critical temperature towards higher polymer/protein ratios with length is more modest; also in qualitative agreement with our polyU-FUS simulations shown in Fig. 1(d). Remarkably, by means of the colloidal patchy particle model, we can explore mixtures with polymer RNA chains much longer than the size of the proteins. We discover that while condensates with RNA lengths of 50 or 100 beads are stable up to higher critical temperatures compared to those of shorter lengths (i.e., 20 beads), for all polymer-bead/protein ratios and for both scaffold and cognate proteins, beyond a certain length at which RNA greatly exceeds the size of the proteins (e.g., 50 times longer than the protein size), the effect of RNA length on LLPS becomes extremely mild (Fig. 4(b-c); red and black curves). These results also support the notion that the longer RNA strands promote condensate stability by increasing the connectivity among proteins within the condensate network [43, 44]. Moreover, since even the shortest polymer length added to the scaffold/cognate mixtures meets this criterion, we do not observe the disruptive influence on LLPS of extremely short RNA chains (i.e., 20-nt) with FUS (526 amino acids) observed in Fig. 1(d).

Our results are consistent with the experimental observation that longer RNA strands present weaker dissociation constants with N-RRM1-2 domains of TDP-43 (which, like PR<sub>25</sub>, cannot phase separate on their own at physiological conditions) than 3-fold shorter RNA strands [33]. It has also been shown that length and charge segregation in the IDR domain of VRN1-like proteins has a critical impact on modulating the DNA-induced VRN1 phase separation, where liquid-like, gel-like or no phase-separation behaviour can be switched depending on the IDR length and the presence of neutral *vs.* highly charged domains [107]. Overall, our patchy particle results presented here, highlight that the RNA length-and-concentration-dependent reentrant phase behaviour observed for both homotypic and heterotypic phase-separating proteins is a general property of soft-matter systems, like biomolecules. Therefore, general features of biomolecules, such as their valency, topology, binding affinity and relative length or size, is what ultimately dictates the intricate phase behaviour of multicomponent biomolecular condensates.

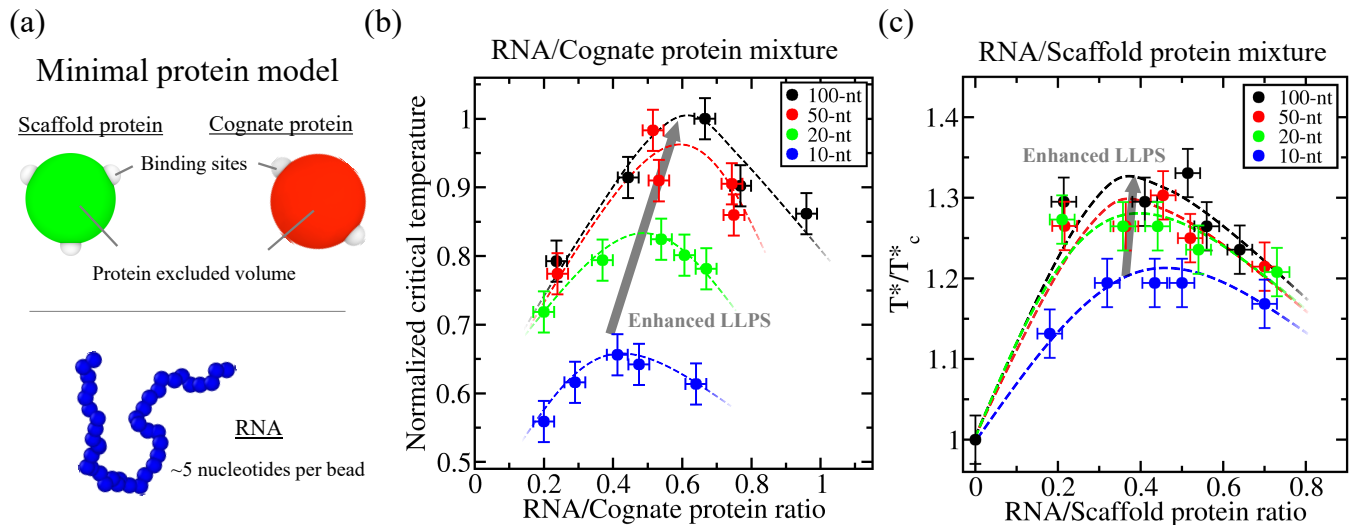


FIG. 4: Minimal coarse-grained model for protein LLPS. (a) Green and red spheres represent the excluded volume of scaffold and cognate proteins respectively, while gray patches represent the binding sites of the proteins. Two different proteins are modeled: scaffold proteins, with 3 binding sites in a planar equidistant arrangement, and cognate proteins, with 2 binding sites in a polar arrangement. Blue spherical beads account for  $\sim 5$  nucleotides each in the RNA model. Please note that for visualization purposes, the size of the RNA beads has been scaled down.

For further technical details on the model, please see the Supplementary Material. (b) Normalized critical temperature of RNA/cognate protein mixtures as a function of the RNA/cognate protein ratio for different RNA strand lengths as indicated in the legend. (c) Normalized critical temperature of RNA/scaffold protein mixtures as a function of the RNA/scaffold protein ratio for different RNA strand lengths as indicated in the legend. While in Panel (b) all temperatures have been normalized by the highest  $T^*$  ( $T^*=0.11$ ) at which phase separation was observed, in Panel (c) all temperatures have been normalized by the critical temperature of the scaffold protein in absence of RNA ( $T_c^*=0.09$ ).

#### IV. CONCLUDING REMARKS

In this study, using a multiscale modelling approach we reveal how RNA length can tune concentration-dependent reentrant phase behaviour of RNA-protein condensates. Drawing together the results from our sequence-dependent model and patchy particle simulations, we demonstrate that long RNA polymers act as enhancers of phase separation. Not only do longer RNA strands enable phase separation up to higher critical temperatures, they also facilitate phase separation up to higher RNA/protein ratios. We show that this is a physical feature of long polymers, since consistent mass ratios and net charges are used for all our RNA/polymer systems with different lengths. Our finding that longer chain RNA molecules increase the capacity of RNA to raise the critical solution temperature to higher values and that the corresponding condensates can incorporate a higher proportion of RNA nucleotides had not been directly demonstrated previously. However, these results are consistent with the wide body of work demonstrating the stabilising role of molecules with an increasing number of repeating chemical building blocks, or higher valencies, on biomolecular condensates (e.g., longer RNAs [44], DNAs [108, 109], or polySUMO/polySIM [96], and

IDPs with more stickers [35, 56, 110]).

Additionally, our finding that long RNA tends to localise to the condensate's core whereas shorter strands accumulate at the droplet surface provides further support for the theory that the stabilising and enabling effects of long RNA chains comes, in part, from their ability to act as nucleators [46] and scaffolds [44] in phase-separated condensates. While long RNA stabilise condensates by increasing the density of molecular connections of the liquid network, short RNAs act as better surfactants that increase condensate stability by reducing their interfacial free energy [63]. Moreover, the influence of RNA length on condensate organisation and viscoelastic properties has important implications on the dynamics of the different components and the mechanisms explaining liquid-to-solid transitions, with potential implications in neurodegenerative disorders [18, 19, 21, 47, 102].

Considering the significant interplay between RNA length and concentration might be relevant to rationalise why some biomolecular condensates, such as stress granules or paraspeckles, can be nucleated by long RNA strands [20, 46, 97] despite being in RNA-rich environments. According to our results, such behaviour

might be facilitated by RNAs serving as scaffolds and surfactants of RNA-protein condensates at high concentration. In particular, while high RNA-concentrations are expected to inhibit phase separation [53], a diverse population of shorter and longer RNAs can yield stable condensates by incorporating the longer RNAs at concentrations beyond the electroneutral point at the condensate core—to enhance condensate connectivity—and the shorter RNAs at the condensate interface—to reduce the interfacial free energy. The observation that RNA length has a similar impact on the concentration-dependent reentrant phase behaviour for both FUS and PR<sub>25</sub> condensates, although quantitatively shifting the maximum critical temperature by substantially different extents, is also significant and further emphasises that this observation results from the physical properties of long RNA chains rather than merely on the type of interaction driving phase separation.

Our results also provide evidence for the hypothesis by Henninger *et al.* [53] that RNA length and concentration may act in concert to regulate the formation of intranuclear condensates. Specifically, the hypothesis that formation and dissolution of transcriptional condensates is regulated by RNA through a negative feedback loop [53] could be explained by RNA-concentration-dependent reentrant phase behaviour alone. However, our findings suggesting that longer chain RNAs, which are produced during a transcriptional burst, stabilise condensates up to higher RNA concentrations may provide robustness to the feedback loop. As higher nucleotide concentrations would need to be reached for condensate dissolution, the transcriptional condensates would remain intact until sufficient levels of longer chain mRNA have been synthesised.

Taken together, our results demonstrate some of the many ways in which RNA act as a key regulator of phase separation in the cell and adds to the growing consensus considering its role as essential for developing a robust understanding of the regulation and dysregulation of biomolecular condensates [21, 111]. Several causes of condensate damage in neurodegenerative disorders, such as disease associated mutations [112, 113], do not alter the fundamental changes in phase behaviour which differentiate liquid-like condensates from solid-like

aggregates [102], but act to increase the likelihood of these changes occurring [105]. In other words, the properties of condensates containing mutated proteins might not be discernibly different from those containing wild type proteins other than in terms of the timescale of their maturation to solid-like aggregates. This is a strong indication that the focus of research into potential treatments for condensate-associated diseases should shift towards more fundamental and general mechanisms of biomolecular condensate regulation [114–117]. In that sense, the growing body of evidence that RNA plays a major role in driving changes of condensate phase behaviour, to which this work contributes, positions RNA-protein liquid-liquid phase separation as an area of research with great potential.

## ACKNOWLEDGEMENTS

This project has received funding from the Oppenheimer Research Fellowship of the University of Cambridge. I. S.-B. acknowledges funding from the Oppenheimer Fellowship, Derek Brewer scholarship of Emmanuel College and EPSRC Doctoral Training Programme studentship, number EP/T517847/1. J. R. E. also acknowledges funding from the Roger Ekins Research Fellowship of Emmanuel College. R.C.-G. acknowledges funding from the European Research Council (ERC) under the European Union’s Seventh Framework Programme (FP7/2007-2013) through the ERC grant PhysProt (agreement no. 337969). This work has been 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.

## SUPPLEMENTARY INFORMATION

Please see the Supplementary Material (SM) for further details on the models, simulation details, and computational methods employed in this work.

- 
- [1] A. A. Hyman, C. A. Weber, and F. Jülicher, Annual Review of Cell and Developmental Biology **30**, 39 (2014).
  - [2] E. Gomes and J. Shorter, Journal of Biological Chemistry **294**, 7115 (2019).
  - [3] S. Alberti, Current Biology **27**, R1097 (2017).
  - [4] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoegge, J. Gharakhani, F. Jülicher, and A. A. Hyman, Science **324**, 1729 (2009).
  - [5] C. C. Correll, J. Bartek, and M. Dunder, Cells **8**, 869 (2019).
  - [6] M. Feric, N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, and C. P. Brangwynne, Cell **165**, 1686 (2016).
  - [7] M. Platani, I. Goldberg, J. R. Swedlow, and A. I. Lamond, The Journal of cell biology **151**, 1561 (2000).
  - [8] K. E. Handwerker, J. A. Cordero, and J. G. Gall, Molecular biology of the cell **16**, 202 (2005).

- [9] J. A. West, M. Mito, S. Kurosaka, T. Takumi, C. Tanegashima, T. Chujo, K. Yanaka, R. E. Kingston, T. Hirose, C. Bond, *et al.*, *Journal of cell biology* **214**, 817 (2016).
- [10] T. Yamazaki, S. Souquere, T. Chujo, S. Kobelke, Y. S. Chong, A. H. Fox, C. S. Bond, S. Nakagawa, G. Pierron, and T. Hirose, *Molecular cell* **70**, 1038 (2018).
- [11] E. Bentmann, M. Neumann, S. Tahirovic, R. Rodde, D. Dormann, and C. Haass, *Journal of Biological Chemistry* **287**, 23079 (2012).
- [12] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, *Cell* **163**, 123 (2015).
- [13] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen, *Nature* **547**, 241 (2017).
- [14] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, *Nature* **547**, 236 (2017).
- [15] G. J. Narlikar, *Journal of Biosciences* **45**, 5 (2020).
- [16] S. E. Farr, E. J. Woods, J. A. Joseph, A. Garaizar, and R. Collepardo-Guevara, *bioRxiv* (2020).
- [17] S. Sanulli, M. Trnka, V. Dharmarajan, R. Tibble, B. Pascal, A. Burlingame, P. Griffin, J. Gross, and G. Narlikar, *Nature* **575**, 390 (2019).
- [18] S. Elbaum-Garfinkle, Y. Kim, K. Szczepaniak, C. C.-H. Chen, C. R. Eckmann, S. Myong, and C. P. Brangwynne, *Proceedings of the National Academy of Sciences* **112**, 7189 (2015).
- [19] M.-T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. C.-H. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, and C. P. Brangwynne, *Nature Chemistry* **9**, 1118 (2017).
- [20] K. A. Burke, A. M. Janke, C. L. Rhine, and N. L. Fawzi, *Molecular cell* **60**, 231 (2015).
- [21] L. Guo and J. Shorter, *Molecular cell* **60**, 189 (2015).
- [22] I. Alshareedah, T. Kaur, J. Ngo, H. Seppala, L.-A. D. Kounatse, W. Wang, M. M. Moosa, and P. R. Banerjee, *Journal of the American Chemical Society* **141**, 14593 (2019).
- [23] X. Gui, F. Luo, Y. Li, H. Zhou, Z. Qin, Z. Liu, J. Gu, M. Xie, K. Zhao, B. Dai, *et al.*, *Nature communications* **10**, 1 (2019).
- [24] S. Qamar, G. Wang, S. J. Randle, F. S. Ruggeri, J. A. Varela, J. Q. Lin, E. C. Phillips, A. Miyashita, D. Williams, F. Ströhl, *et al.*, *Cell* **173**, 720 (2018).
- [25] A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerbe, S. H. Parekh, J. Mittal, and N. L. Fawzi, *Nature structural & molecular biology* **26**, 637 (2019).
- [26] S. N. Rhoads, Z. T. Monahan, D. S. Yee, and F. P. Shewmaker, *International journal of molecular sciences* **19**, 886 (2018).
- [27] P. Yang, C. Mathieu, R.-M. Kolaitis, P. Zhang, J. Messing, U. Yurtsever, Z. Yang, J. Wu, Y. Li, Q. Pan, *et al.*, *Cell* **181**, 325 (2020).
- [28] Y. Gwon, B. A. Maxwell, R.-M. Kolaitis, P. Zhang, H. J. Kim, and J. P. Taylor, *Science* **372**, eabf6548 (2021).
- [29] D. W. Sanders, N. Kedersha, D. S. Lee, A. R. Strom, V. Drake, J. A. Riback, D. Bracha, J. M. Eeftens, A. Iwanicki, A. Wang, *et al.*, *Cell* **181**, 306 (2020).
- [30] J. Guillén-Boixet, A. Kopach, A. S. Holehouse, S. Wittmann, M. Jahnel, R. Schlüssler, K. Kim, I. R. Trussina, J. Wang, D. Mateju, *et al.*, *Cell* **181**, 346 (2020).
- [31] H.-R. Li, W.-C. Chiang, P.-C. Chou, W.-J. Wang, and J.-r. Huang, *Journal of Biological Chemistry* **293**, 6090 (2018).
- [32] L. McGurk, E. Gomes, L. Guo, J. Mojsilovic-Petrovic, V. Tran, R. G. Kalb, J. Shorter, and N. M. Bonini, *Molecular cell* **71**, 703 (2018).
- [33] E. Zacco, R. Graña-Montes, S. R. Martin, N. S. de Groot, C. Alfano, G. G. Tartaglia, and A. Pastore, *Journal of molecular biology* **431**, 1671 (2019).
- [34] J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, *et al.*, *Cell* **174**, 688 (2018).
- [35] A. Bremer, M. Farag, W. M. Borchers, I. Peran, E. W. Martin, R. V. Pappu, and T. Mittag, *bioRxiv* (2021).
- [36] A. Cléry, M. Blatter, and F. H. Allain, *Current opinion in structural biology* **18**, 290 (2008).
- [37] P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, and A. A. Deniz, *Angewandte Chemie* **129**, 11512 (2017).
- [38] S. Matsui and R.-S. Nozawa, *Biomedical Research* **42**, 153 (2021).
- [39] E. R. Kuechler, P. M. Budzyńska, J. P. Bernardini, J. Gsponer, and T. Mayor, *Journal of molecular biology* **432**, 2349 (2020).
- [40] P. A. Chong, R. M. Vernon, and J. D. Forman-Kay, *Journal of molecular biology* **430**, 4650 (2018).
- [41] J. C. Schwartz, X. Wang, E. R. Podell, and T. R. Cech, *Cell reports* **5**, 918 (2013).
- [42] R. M. Regy, G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, *Nucleic Acids Research* **48**, 12593 (2020).
- [43] A. R. Tejedor, A. Garaizar, J. Ramírez, and J. R. Espinosa, *Biophysical Journal* **120**, 5169 (2021).
- [44] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, *PLoS computational biology* **18**, e1009810 (2022).
- [45] E. M. Langdon, Y. Qiu, A. G. Niaki, G. A. McLaughlin, C. A. Weidmann, T. M. Gerbich, J. A. Smith, J. M. Crutchley, C. M. Termini, K. M. Weeks, S. Myong, and A. S. Gladfelter, *Science* **360**, 922 (2018).
- [46] S. Maharana, J. Wang, D. K. Papadopoulos, D. Richter, A. Pozniakovsky, I. Poser, M. Bickle, S. Rizk, J. Guillén-Boixet, T. M. Franzmann, *et al.*, *Science* **360**, 918 (2018).
- [47] B. Portz, B. L. Lee, and J. Shorter, *Trends in Biochemical Sciences* (2021).
- [48] C. Vance, B. Rogelj, T. Hortobágyi, K. J. De Vos, A. L. Nishimura, J. Sreedharan, X. Hu, B. Smith, D. Ruddy, P. Wright, *et al.*, *Science* **323**, 1208 (2009).
- [49] W. M. Aumiller and C. D. Keating, *Nature chemistry* **8**, 129 (2016).
- [50] R. S. Fisher and S. Elbaum-Garfinkle, *Nature communications* **11**, 1 (2020).
- [51] T. J. Welsh, G. Krainer, J. R. Espinosa, J. A. Joseph, A. Sridhar, M. Jahnel, W. E. Arter, K. L. Saar, S. Alberti, R. Collepardo-Guevara, *et al.*, *Nano Letters* **22**, 612 (2022).
- [52] T. Lu and E. Spruijt, *Journal of the American Chemical Society* **142**, 2905 (2020).
- [53] J. E. Henninger, O. Oksuz, K. Shrinivas, I. Sagi, G. LeRoy, M. M. Zheng, J. O. Andrews, A. V. Zamudio, C. Lazaris, N. M. Hannett, *et al.*, *Cell* **184**, 207 (2021).
- [54] E. Tunnacliffe and J. R. Chubb, *Trends in Genetics* **36**, 288 (2020).

- [55] S. Alberti, A. Gladfelter, and T. Mittag, *Cell* **176**, 419 (2019).
- [56] A. Garaizar, I. Sanchez-Burgos, R. Collepardo-Guevara, and J. R. Espinosa, *Molecules* **25**, 4705 (2020).
- [57] A. Garaizar and J. R. Espinosa, *The Journal of Chemical Physics* **155**, 125103 (2021).
- [58] R. M. Regy, J. Thompson, Y. C. Kim, and J. Mittal, *Protein Science* (2021).
- [59] V. Nguemaha and H.-X. Zhou, *Scientific reports* **8**, 1 (2018).
- [60] H.-Y. Chou and A. Aksimentiev, *The Journal of Physical Chemistry Letters* **11**, 4923 (2020).
- [61] S. Ranganathan and E. Shakhnovich, *The Journal of Physical Chemistry B* **125**, 5035 (2021).
- [62] E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino, *Physical review letters* **97**, 168301 (2006).
- [63] I. Sanchez-Burgos, J. A. Joseph, R. Collepardo-Guevara, and J. R. Espinosa, *Scientific Reports* **11**, 15241 (2021).
- [64] J. R. Espinosa, A. Garaizar, C. Vega, D. Frenkel, and R. Collepardo-Guevara, *J. Chem. Phys* **150**, 224510 (2019).
- [65] J. R. Espinosa, J. A. Joseph, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, *Proceedings of the National Academy of Sciences* **117**, 13238 (2020).
- [66] I. Sanchez-Burgos, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, *Biomolecules* **11**, 278 (2021).
- [67] A. Statt, H. Casademunt, C. P. Brangwynne, and A. Z. Panagiotopoulos, *The Journal of Chemical Physics* **152**, 075101 (2020).
- [68] F. Dar and R. Pappu, *Elife* **9**, e59663 (2020).
- [69] A. Garaizar, J. R. Espinosa, J. A. Joseph, G. Krainer, Y. Shen, T. P. Knowles, and R. Collepardo-Guevara, *Proceedings of the National Academy of Sciences* **119**, e2119800119 (2022).
- [70] A. E. Conicella, G. L. Dignon, G. H. Zerze, H. B. Schmidt, M. Alexandra, Y. C. Kim, R. Rohatgi, Y. M. Ayala, J. Mittal, and N. L. Fawzi, *Proceedings of the National Academy of Sciences* **117**, 5883 (2020).
- [71] A. R. Tejedor, I. Sanchez-Burgos, M. Estevez-Espinosa, A. Garaizar, R. Collepardo-Guevara, J. Ramirez, and J. R. Espinosa, *Nature communications* **13**, 5717 (2022).
- [72] W. M. Jacobs, *Physical review letters* **126**, 258101 (2021).
- [73] J. A. Joseph, A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar, and R. Collepardo-Guevara, *Nature Computational Science* **1**, 732 (2021).
- [74] G. Krainer, T. J. Welsh, J. A. Joseph, J. R. Espinosa, S. Wittmann, E. de Csilléry, A. Sridhar, Z. Toprakcioglu, G. Gudíškýté, M. A. Czekalska, *et al.*, *Nature Communications* **12**, 1 (2021).
- [75] J. Lee, Y. O. Popov, and G. H. Fredrickson, *The Journal of chemical physics* **128**, 224908 (2008).
- [76] Y.-H. Lin, J. P. Brady, J. D. Forman-Kay, and H. S. Chan, *New Journal of Physics* **19**, 115003 (2017).
- [77] J. McCarty, K. T. Delaney, S. P. Danielsen, G. H. Fredrickson, and J.-E. Shea, *The journal of physical chemistry letters* **10**, 1644 (2019).
- [78] J.-M. Choi, A. A. Hyman, and R. V. Pappu, *Physical Review E* **102** (2020).
- [79] C. A. Weber, D. Zwicker, F. Jülicher, and C. F. Lee, *Reports on Progress in Physics* **82**, 064601 (2019).
- [80] T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, *elife* **6**, e30294 (2017).
- [81] S. Das, A. Eisen, Y.-H. Lin, and H. S. Chan, *The Journal of Physical Chemistry B* **122**, 5418 (2018).
- [82] T. S. Harmon, A. S. Holehouse, and R. V. Pappu, *New Journal of Physics* **20**, 045002 (2018).
- [83] J.-M. Choi, F. Dar, and R. V. Pappu, *PLOS Computational Biology* **15**, e1007028 (2019).
- [84] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, *Proceedings of the National Academy of Sciences* **115**, 9929 (2018).
- [85] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, *PLoS computational biology* **14**, e1005941 (2018).
- [86] T. M. Perdikari, N. Jovic, G. L. Dignon, Y. C. Kim, N. L. Fawzi, and J. Mittal, *Biophysical Journal* (2021), 10.1016/j.bpj.2021.01.034.
- [87] G. Tesei, T. K. Schulze, R. Crehuet, and K. Lindorff-Larsen, *Proceedings of the National Academy of Sciences* **118**, e2111696118 (2021).
- [88] Y. Qi and B. Zhang, *Nature communications* **12**, 1 (2021).
- [89] J. A. Joseph, J. R. Espinosa, I. Sanchez-Burgos, A. Garaizar, D. Frenkel, and R. Collepardo-Guevara, *Biophysical Journal* **120**, 1219 (2021).
- [90] X. Wang, S. Ramírez-Hinestrosa, J. Dobnikar, and D. Frenkel, *Physical Chemistry Chemical Physics* **22**, 10624 (2020).
- [91] J. Jover, A. J. Haslam, A. Galindo, G. Jackson, and E. A. Müller, *Journal of Chemical Physics* **137** (2012), 10.1063/1.4754275.
- [92] J. Espinosa, C. Vega, and E. Sanz, *The Journal of chemical physics* **141**, 134709 (2014).
- [93] A. Ladd and L. Woodcock, *Chemical Physics Letters* **51**, 155 (1977).
- [94] R. García Fernández, J. L. Abascal, and C. Vega, *The Journal of chemical physics* **124**, 144506 (2006).
- [95] J. S. Rowlinson and B. Widom, *Molecular theory of capillarity* (Courier Corporation, 2013).
- [96] S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, and M. K. Rosen, *Cell* **166**, 651 (2016).
- [97] B. Van Treeck and R. Parker, *Cell* **174**, 791 (2018).
- [98] D. Teixeira, U. Sheth, M. A. Valencia-Sanchez, M. Brengues, and R. Parker, *Rna* **11**, 371 (2005).
- [99] Y.-B. Lee, H.-J. Chen, J. N. Peres, J. Gomez-Deza, J. Attig, M. Štalekar, C. Troakes, A. L. Nishimura, E. L. Scotter, C. Vance, *et al.*, *Cell reports* **5**, 1178 (2013).
- [100] N. Krahn, J. T. Fischer, and D. Söll, *Frontiers in Microbiology* , 2616 (2020).
- [101] I. Palaia and A. Šarić, *The Journal of Chemical Physics* **156**, 194902 (2022).
- [102] A. Garaizar, J. R. Espinosa, J. A. Joseph, and R. Collepardo-Guevara, *Scientific reports* **12**, 4390 (2022).
- [103] A. Cochard, M. G.-J. Navarro, L. Piroška, S. Kashida, M. Kress, D. Weil, and Z. Gueroui, *Biophysical Journal* **121**, 1675 (2022).
- [104] H. Zhang, S. Elbaum-Garfinkle, E. M. Langdon, N. Taylor, P. Occhipinti, A. A. Bridges, C. P. Brangwynne, and A. S. Gladfelter, *Molecular cell* **60**, 220 (2015).
- [105] A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, *et al.*, *Cell* **162**, 1066 (2015).

- [106] Y. Shen, A. Chen, W. Wang, Y. Shen, F. S. Ruggeri, S. Aime, Z. Wang, S. Qamar, J. R. Espinosa, A. Garaizar, P. S. George-Hyslop, R. Collepardo-Guevara, D. A. Weitz, D. Vigolo, and T. P. J. Knowles, *bioRxiv* (2022), 10.1101/2022.08.15.503964.
- [107] Y. Wang, H. Zhou, X. Sun, Q. Huang, S. Li, Z. Liu, C. Zhang, and L. Lai, *Journal of Molecular Biology* **433**, 167269 (2021).
- [108] F. Muzzopappa, M. Hertzog, and F. Erdel, *Biophysical journal* **120**, 1288 (2021).
- [109] S. E. Farr, E. J. Woods, J. A. Joseph, A. Garaizar, and R. Collepardo-Guevara, *Nature communications* **12**, 1 (2021).
- [110] E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag, *Science* **367**, 694 (2020).
- [111] H. Zhu, H. Fu, T. Cui, L. Ning, H. Shao, Y. Guo, Y. Ke, J. Zheng, H. Lin, X. Wu, *et al.*, *Nucleic Acids Research* **50**, D340 (2022).
- [112] S. Boeynaems, E. Bogaert, D. Kovacs, A. Konijnenberg, E. Timmerman, A. Volkov, M. Guharoy, M. De Decker, T. Jaspers, V. H. Ryan, *et al.*, *Molecular cell* **65**, 1044 (2017).
- [113] A. E. Renton, A. Chiò, and B. J. Traynor, *Nature neuroscience* **17**, 17 (2014).
- [114] B. Wang, L. Zhang, T. Dai, Z. Qin, H. Lu, L. Zhang, and F. Zhou, *Signal Transduction and Targeted Therapy* **6**, 1 (2021).
- [115] S. Alberti and D. Dormann, *Annu. Rev. Genet* **53**, 171 (2019).
- [116] R. J. Wheeler, *Emerging Topics in Life Sciences* **4**, 331 (2020).
- [117] V. Verdile, E. De Paola, and M. P. Paronetto, *Frontiers in genetics* **10**, 173 (2019).