

# Sequence dependent co-phase separation of RNA-protein mixtures elucidated using molecular simulations

Roshan Mammen Regy<sup>1</sup>, Gregory L. Dignon<sup>1</sup>, Wenwei Zheng<sup>2</sup>, Young Chan Kim<sup>3</sup>, and Jeetain Mittal<sup>1,\*</sup>

<sup>1</sup> Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, Pennsylvania, 18015, United States

<sup>2</sup> College of Integrative Sciences and Arts, Arizona State University, Mesa, Arizona, 85212, United States

<sup>3</sup> Center for Materials Physics and Technology, Naval Research Laboratory, Washington, DC 20375, United States

\* To whom correspondence should be addressed. Tel: +1 610-758-4791; Fax: +1 610-758-5057; Email: jem309@lehigh.edu

Present Address: Gregory L. Dignon, Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York, 11794, United States

## ABSTRACT

Ribonucleoprotein (RNP) granules are membraneless organelles (MLOs) which majorly consist of RNA and RNA-binding proteins and are formed via liquid-liquid phase separation (LLPS). Experimental studies investigating the drivers of LLPS have shown that intrinsically disordered proteins (IDPs) and nucleic acids like RNA play a key role in modulating protein phase separation. There is currently a dearth of modelling techniques which allow one to delve deeper into how RNA plays its role as a modulator/promoter of LLPS in cells using computational methods. Here we present a coarse-grained RNA model developed to fill this gap, which together with our recently developed HPS model for protein LLPS, allows us to capture the factors driving RNA-protein co-phase separation. We explore the capabilities of the modelling framework with the LAF-1 RGG/RNA system which has been well studied in experiments and also with the HPS model previously. Further taking advantage of the fact that the HPS model maintains sequence specificity we explore the role of charge patterning on controlling RNA incorporation into condensates. With increased charge patterning we observe formation of structured or patterned condensates which suggests the possible roles of RNA in not only shifting the phase boundaries but also introducing microscopic organization in MLOs.

## INTRODUCTION

Membraneless organelles (MLOs) are compartments formed in the cell which consist of a concentrated set of biomolecules without an enclosing membrane separating them from the surrounding cytosol (1–3). Many of these MLOs are assemblies consisting of RNA-binding proteins and RNA are commonly referred to as Ribonucleoprotein (RNP) granules (4). Prominent RNP granules include P bodies, stress granules, germ cell P granules, and neuronal granules which perform diverse functions that are essential for the survival of the cell. P bodies are mRNA and

protein containing cytoplasmic processing bodies associated with RNA metabolism (5). Stress granules are formed when cells respond to stress by selectively including specific mRNA transcripts into granules and regulating/arresting translation (6). Germ cell P granules form during germ cell development (7), whereas neuronal granules transport mRNAs in response to specific exogenous stimuli (8). Despite the highly diverse functions, these RNP granules share a common process through which they localize their constituent proteins and nucleic acids, that is, through liquid-liquid phase separation (LLPS). Therefore, investigating the molecular mechanism underlying LLPS of biomolecules is essential to understanding how RNP granules store, process and control the activities of their constituents.

It is now well acknowledged that these MLOs are formed when a homogeneous mixture of biomolecules undergoes the physical process of LLPS to form coexisting condensed and dilute phases stabilized by a balance between entropic and enthalpic interaction (9, 10). These biomolecular condensates may appear as liquid-like droplets and allow a rapid exchange of components with the environment in a dynamic manner (11). Investigations into the drivers of protein LLPS have also shown that the disordered domains of certain proteins such as the DEAD-box helicase LAF-1 and the RNA-binding protein Fused in Sarcoma (FUS) are able to create phase separated condensates *in vitro* (11, 12), thus suggesting the importance of intrinsically disordered regions (IDRs) in facilitating the formation of RNP granules (13). Incorporation of RNA into condensed phases has been further shown to perturb the physical and chemical properties of these condensates (11, 12, 14–18). For instance, incorporating RNA into the condensed phase of the DEAD-box helicase LAF-1 protein, which is a major component of P bodies, can shift the thermodynamic phase diagram for the LAF-1 protein and its disordered RGG rich domain and change both the saturation concentration and LAF-1 diffusion inside the liquid droplets in *in vitro* experiments (11). In another experimental study varying the concentration of RNA in a co-phase separating mixture of RNA and FUS shows re-entrant protein phase separation behavior where RNA acts as a facilitator of FUS phase separation at low concentration but on increasing the RNA concentration beyond a certain point we see RNA acting as a disruptor of protein LLPS (12, 19). These studies have brought forth how RNA plays an essential role as a modulator of protein phase separation and condensate properties once it is incorporated into the protein-rich phase.

Despite these advances in establishing the role RNA plays in the formation and function of RNP granules, there currently exists a lack of understanding about the fundamental molecular driving forces driving RNA-protein co-phase separation and its more nuanced aspects like the microstructural organization of different components within the condensates formed. Although remarkable progress has been recently made in providing insights into the role of a protein's sequence in controlling its LLPS (20–22), how the protein sequence plays a role in modulating LLPS of protein-RNA mixtures and the material properties of the condensates formed is not well established (19, 23, 24). Part of this stems from the lack of techniques that can provide in-depth information about the RNA-protein interactions driving RNP granule formation and high spatiotemporal resolution on the molecular organization of protein and RNA within the condensate (12, 25, 26). One can attempt to obtain such

in-depth information from *in silico* atomic resolution simulation techniques (20, 27, 28) but studying a macroscopic phenomenon like phase separation would require considerable computational resources making this method quite expensive and prohibitive. This prompts us to look into computational approaches based on coarse-grained (CG) models which can allow investigations into the formation of biomolecular condensates and to provide molecular-level details necessary to develop theories of phase separation making CG models an integral part of the biophysical toolkit to study phase separation (29–32). We have previously developed a CG modeling framework based on the amino acid hydrophathy (HPS model) to study sequence determinants of protein phase separation that does not require input from experimental data (33, 34). The HPS model is based on defining nonbonded interactions between amino acid pairs using the hydrophathy values of the naturally occurring twenty amino acids (33, 35). This model was recently extended to post-translationally modified amino acids as well (36).

In the present work, we first define parameters for RNA nucleotides within the HPS modeling framework using a one bead per nucleotide CG representation of RNA. We then test the new model to study the co-phase separation of the N-terminal disordered RGG domain of the LAF-1 protein (hereafter referred to as LAF-1 RGG) and RNA molecules with the most recent experimental data available (11). The model provides reasonable agreement with the experiment regarding the effects of RNA on modulating LAF-1 RGG phase separation. We then utilize this new CG model to study the co-phase separation LAF-1 RGG protein variants with the same sequence composition but different arrangement of charged amino acids used in our previous work (37). Consistent with the expectations from previous work (31, 37), we find that the phase behavior can be significantly perturbed by changes in the protein sequence, and this change in behavior also alters co-phase separation. Most importantly, we observe a significant change in terms of the co-localization of protein and RNA molecules within the condensate. For the charge segregated variants of the LAF-1 RGG protein, RNA adsorbs on the protein condensate surface rather than mixing evenly throughout as in condensates of WT LAF-1 RGG and RNA mixtures. We provide interpretation for the observed behavior based on the potential of mean force between pairs of protein and RNA molecules. Our work here attempts to provide evidence linking competing intermolecular interactions with the sub-compartmentalization inside MLOs allowing us to better understand what determines the morphology of MLOs.

## MATERIAL AND METHODS

### *HPS model for proteins*

Our CG modeling approach for proteins uses a one bead per amino acid level of resolution, and the twenty naturally occurring amino acids are identified in the model by the following characteristics: the mass, charge, diameter ( $\sigma$ ) (38), and hydrophathy ( $\lambda$ ) of the amino acid. The  $\lambda$  value is derived from the partial charges of atoms belonging to an amino acid in an all-atom force field and scaled to range from 0 (Arginine) to 1 (Phenylalanine), which is based on the approach proposed by Kapcha and Rosky (35). In the CG energy function, we have three types of interactions: bonded, electrostatic, and short-range pairwise interactions. The bonds between consecutive amino acids in the protein sequence are modeled using a harmonic potential with a spring constant of 10 kJ/Å<sup>2</sup> and a bond

length of 3.8 Å. The electrostatic interactions are modeled using the Debye-Hückel (DH) electrostatic screening term (39),

$$E_{i,j} = \frac{q_i q_j}{4\pi D r} \exp\left(-\frac{r}{\kappa}\right)$$

(1)

in which  $\kappa$  is the Debye screening length, and  $D = 80$  is the dielectric constant of the solvent. We set  $\kappa = 10$  Å, corresponding to solution conditions at 100 mM salt. The short-range pairwise interactions are modeled by the Ashbaugh-Hatch functional form (40) as

$$\Phi(r) = \begin{cases} \Phi_{LJ} + (1 - \lambda_{ij})\epsilon, & \text{if } r \leq 2\frac{1}{6}\sigma_{ij} \\ \lambda_{ij}\Phi_{LJ}, & \text{otherwise} \end{cases}$$

(2)

where  $\Phi_{LJ}$  is the standard 12-6 Lennard-Jones potential given by

$$\Phi_{LJ} = 4\epsilon \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - \left( \frac{\sigma_{ij}}{r} \right)^6 \right]$$

(3)

Both the hydrophathy ( $\lambda_{ij}$ ) and the diameter ( $\sigma_{ij}$ ) are set to be the arithmetic average of the values for the interacting pair of amino acids ( $i, j$ ). A value of  $\lambda_{ij}$  close to 1 represents strong attractive interactions whereas a value of 0 represents weak repulsive interactions between the two amino acids. The interaction parameter ( $\epsilon$ ) in eq. 3 is a free parameter and was adjusted to reproduce the experimentally measured radius of gyration of IDPs as 0.2 kcal/mol (33).

### HPS model for RNA

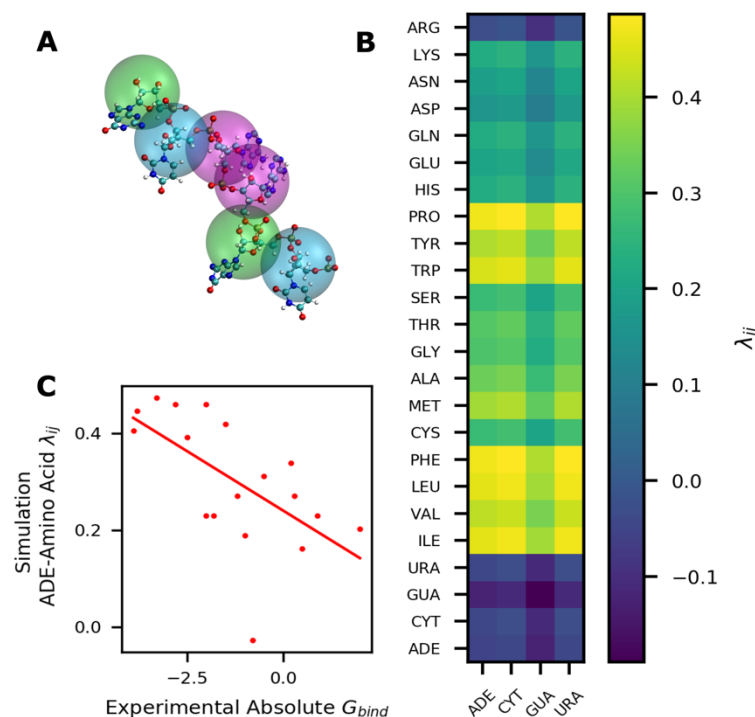
Our purpose for building a new RNA model, as opposed to using an existing model (41–43), is to capture the qualitative features of the RNA-protein co-phase separation behavior with the simplest possible CG representation that is also consistent with our protein model. Hence, we decided to model RNA with each nucleotide represented as a single particle (Fig. 1A). The potential energy also consists of the bonded, electrostatic and short-range pairwise interaction terms, the same as that of the disordered protein but with different parameters (Table 1). A spring constant of 10 kJ/Å<sup>2</sup> and a bond length of 5 Å is used for the bonded term, in which the bond length is derived using the average backbone-backbone distance for single- and double-stranded DNA (44).

Table 1. RNA Model parameters for CG beads, each representing a single nucleotide.

Nucleotide	Mass (amu)	Charge (1)	Diameter, $\sigma$ (Å)	Hydrophathy, $\lambda$ (1, 35)
Adenosine	329.2	-1.0	8.44	-0.054
Cytidine	305.2	-1.0	8.22	-0.027
Guanosine	345.2	-1.0	8.51	-0.189
Uridine	306.2	-1.0	8.17	-0.027

The primary mode of interactions between RNA and proteins is expected to be via electrostatics (24) due to the negatively charged backbone which is captured in our model by placing a negative charge

on each of the nucleotide beads. In addition, nucleotides can interact with all other components via non-electrostatic interactions that are captured by the short-range potential (eq. 2). The hydrophathy value of each nucleotide is assigned in a similar fashion as the amino acids using the atomic scale partial charges from the OPLS all-atom RNA force field (45). The RNA nucleotide hydrophathy values on a normalized scale (based on amino acids) are found to be negative ( $\lambda < 0$ ), which reflects repulsive short-range interactions. The arithmetic average between a pair of CG particles (i.e. twenty amino acids or four RNA nucleotides) is used for characterizing the short-range pairwise interactions. As shown in Fig. 1B, most of the protein-RNA short range interactions are weakly attractive with the most significant attraction between hydrophobic amino acids and nucleotides. Our hydrophobic interaction strengths were in acceptable agreement with reported absolute binding free energy data between nucleotides and amino acids (Fig. 1C) (46).

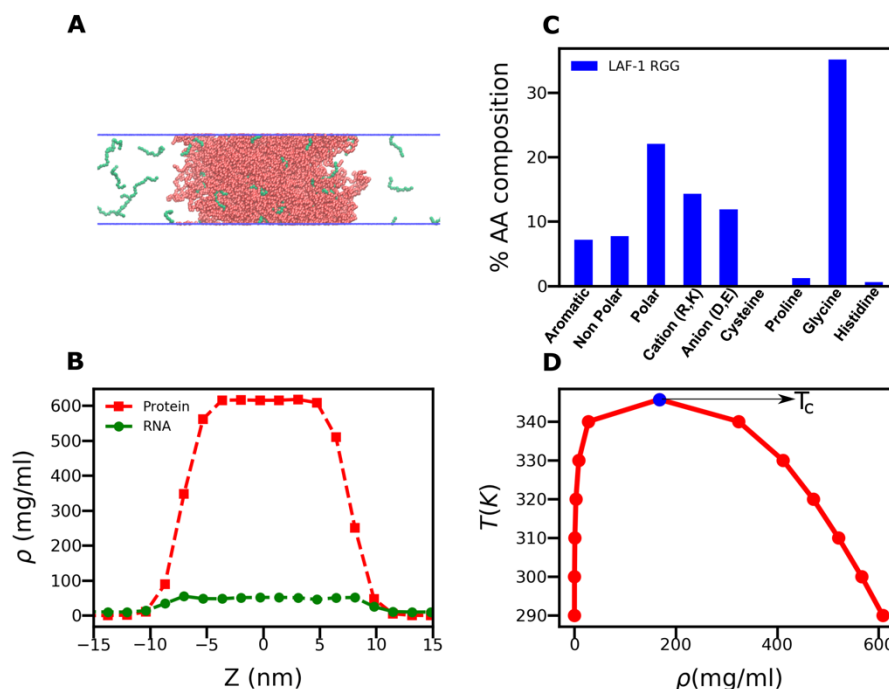


**Figure 1: RNA Coarse Grained model details and comparison:** (A) One bead per nucleotide CG representation of RNA; (B) Average hydrophathy values for pairs of twenty amino acids and four RNA base pairs in the model; (C) Comparison of RNA-protein hydrophobic interaction strengths between the proposed RNA model and existing absolute binding free energy (kJ/mol) data from all-atom umbrella sampling simulations (46).

### Simulation strategy for sampling protein-RNA multicomponent phase separation

Even with the current advances in simulation methodology and the CG nature of the models being used, it is nearly impossible to apply standard free-energy based techniques to sample the phase behavior of long-chain off-lattice polymers (20, 28–31). We and others have been using co-existence simulation methodology (51) to sample the phase behavior of proteins undergoing LLPS successfully and efficiently (20, 32). Here, we use the same strategy, as shown in Fig. 2A, wherein a cuboid shaped periodic box or a 'slab' is simulated with the shorter x and y dimensions of equal length while the z dimension is extended to create a low-density phase (Fig. 2A). The use of planar interfaces in a slab configuration as opposed to a spherical droplet geometry is effective in reducing the system-size

effects in calculating the densities of co-existing phases (30, 50, 51). By conducting such simulations for many temperatures, we can use the co-existing densities (Fig. 2B) to map the binodal (limits of thermodynamic stability) and the critical temperature as well as density values for the phase separating component (Fig. 2C). As in previous work, we conduct an initial equilibration of our system in a cubic simulation box consisting of all components in the NPT ensemble to reach a box size of about 15 nm in each dimension. We then extend the box z-dimension to 280 nm, and conduct coexistence simulations in the NVT ensemble for 5  $\mu$ s. The temperature is controlled using the Langevin thermostat in the low friction limit. All the simulations are conducted with the HOOMD-Blue package to take advantage of its capabilities to speed-up calculations using the graphics processing units (GPUs) (52). For more details about the methodology and the analysis, we refer the readers to our previous work (33, 53).



**Figure 2: Slab simulation methodology:** (A) Slab configuration for CG simulations; (B) Density profiles of LAF-1 RGG and RNA sampled over the course of a slab simulation centered on the largest cluster formed; (C) Amino acid composition of the LAF-1 RGG WT sequence. (D) Binodal phase diagram showing temperature regime over which phase separation is possible.

## RESULTS AND DISCUSSION

### **RNA perturbs LAF-1 RGG phase separation in a concentration-dependent manner**

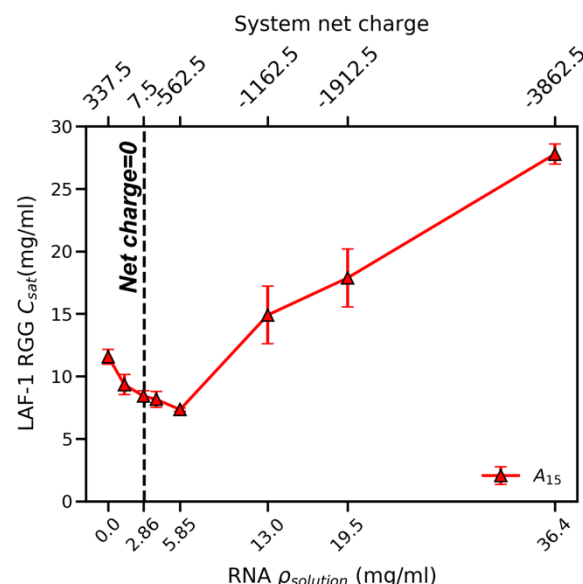
We conducted coexistence simulations of the disordered LAF-1 RGG protein sequence with poly-Adenosine (poly-A)RNA of length 15 nucleotides ( $A_{15}$ ) at a range of temperatures and concentrations of RNA, keeping the total number of LAF-1 RGG molecules fixed. In our coexistence simulations, LAF-1 RGG forms a protein-rich condensed phase in equilibrium with the dilute vapor-like phase (54), whereas RNA alone is not able to phase separate, even at very high concentrations ( $\sim 36$  mg/ml) due to electrostatic repulsion. However, in the case of a binary mixture of LAF-1 RGG and  $A_{15}$ , RNA



Next, we observe the degree to which RNA is incorporated into the condensates by using the A<sub>15</sub> RNA and protein concentrations from both the high- and low-density coexisting phases to construct the binary thermodynamic phase diagram (Fig. 3B). The region enclosed by these two lines is the two-phase coexistence region. If the state point as a function of LAF-1 RGG and A<sub>15</sub> concentrations falls within the coexistence region, the system will phase separate into two phases with relative concentrations of protein and A<sub>15</sub> RNA dictated by the tie lines joining this point to both dense and

dilute phase arms of the multicomponent phase diagram (55). Based on the computed phase diagram in Fig. 3B, we observe a co-phase separation behavior which is referred to as scaffold-client type in the recent LLPS literature (10, 31), in which the scaffold can undergo phase separation on its own and recruit client molecules (which cannot phase separate alone) into the dense phase. At the given conditions, the “scaffold”, LAF-1 RGG can phase separate on its own into a protein-rich phase due to sufficiently strong homotypic attraction. This then recruits the “client” A<sub>15</sub> RNA molecules for which homotypic interactions are either too weak or repulsive, but the heterotypic interactions between the scaffold and client molecules are sufficiently attractive (10, 56). The qualitative shape of the binary phase diagram remains unchanged with increasing temperature and reflects a shrinking coexisting region when approaching the critical temperature for LAF-1 RGG with an upper critical solution temperature (UCST) (34).

So far, we have mostly focused on the condensed phase arm of the phase diagram. From a biological standpoint, it is quite important to quantify the changes in the scaffold (LAF-1 RGG) concentration of the low-density phase, also referred to as the saturation concentration ( $C_{\text{sat}}$ ), with the changes in client (RNA) solution concentration. The significance of  $C_{\text{sat}}$  is that this is the minimum protein concentration required to undergo LLPS, so it is important to understand how the presence of RNA can be used to modulate this. In Fig. 4, we show how  $C_{\text{sat}}$  changes with increasing RNA concentration. At low RNA concentrations, there is an initial decrease in the  $C_{\text{sat}}$  values (by as much as a factor of two) highlighting the LAF-1 RGG’s enhanced propensity to phase separate in the presence of RNA. Similar LLPS behavior has also been observed previously in many other cases (19, 57, 58) which in this case is due to favorable attraction between cationic Arginine residues and anionic nucleotides.



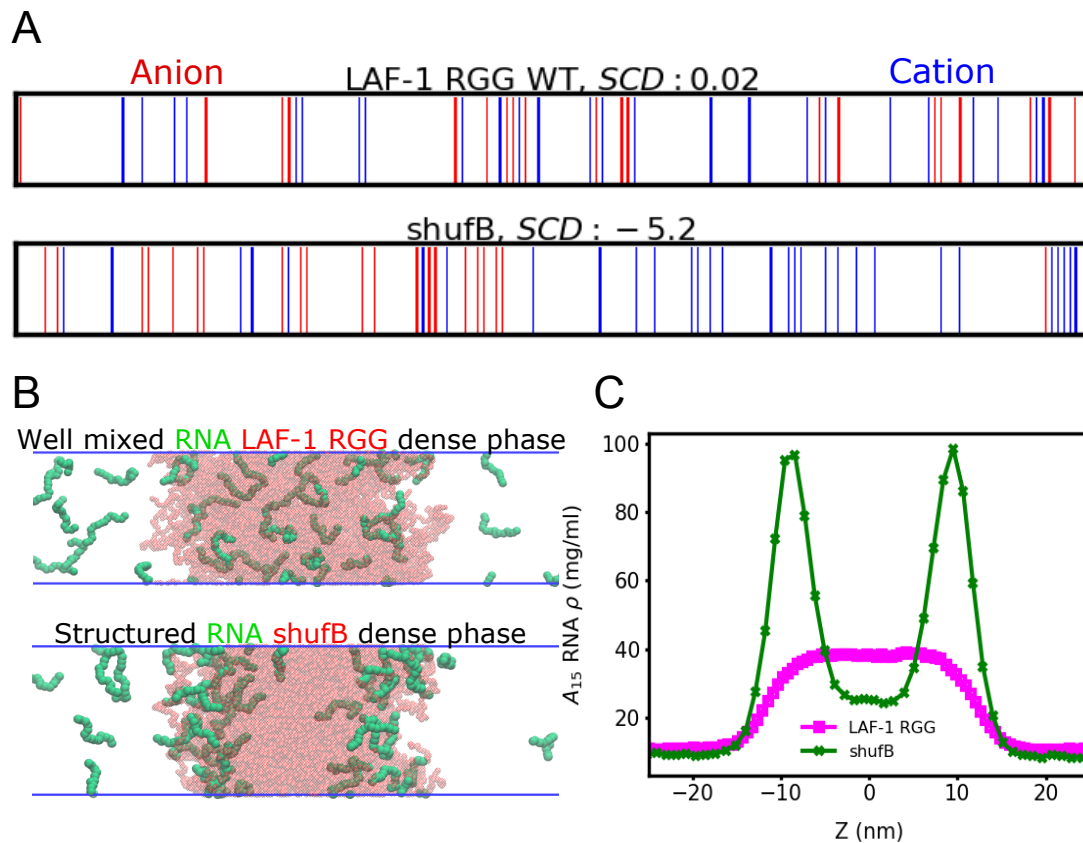
**Figure 4: RNA shows concentration dependent protein LLPS promoter/disruptor roles for different chain lengths:** Effect of increasing RNA solution concentration on LAF-1 RGG saturation concentration shows switching between RNA LLPS promoter/disruptor roles (non-monotonic curve). Vertical line represents a net charge = 0 system.



Upon a further increase in the RNA concentration beyond a certain threshold value, we observe an increase in  $C_{\text{sat}}$ , highlighting LAF-1 RGG's reduced propensity to phase separate. In fact, at much higher RNA concentrations, the LAF-1 RGG  $C_{\text{sat}}$  can be higher than in the absence of RNA. The concentration at which RNA switches its role as a promoter to a disruptor of protein LLPS may be expected to be near the net neutral state of the solution. However, we see a small shift in this, possibly due to the non-electrostatic, multivalent hydrophobic interactions between nucleotides and amino acids. Previous experimental studies by the Banerjee group (59) have provided remarkable insights into this type of non-monotonic change in  $C_{\text{sat}}$  with increasing RNA concentration. They showed that in the cases of the FUS/RNA, protamine/RNA mixtures, the system displays a reentrant phase separation behavior, similar to what is observed in the case of LAF-1 RGG. This reentrant behavior emerges from a competition between the attractive and repulsive electrostatic interactions between unlike and like charges, respectively, and depends on the net system charge. Banerjee and co-workers also developed a theoretical approach to explain the observed behavior in terms of a charge inversion mechanism which suggests increasing instability of the condensate as more RNA is added to a mixture beyond a certain point due to charge over screening (19). Their theory predicts a switch in the surface charge of the condensates after the critical point which was confirmed with electrophoretic mobility measurements in experiments.

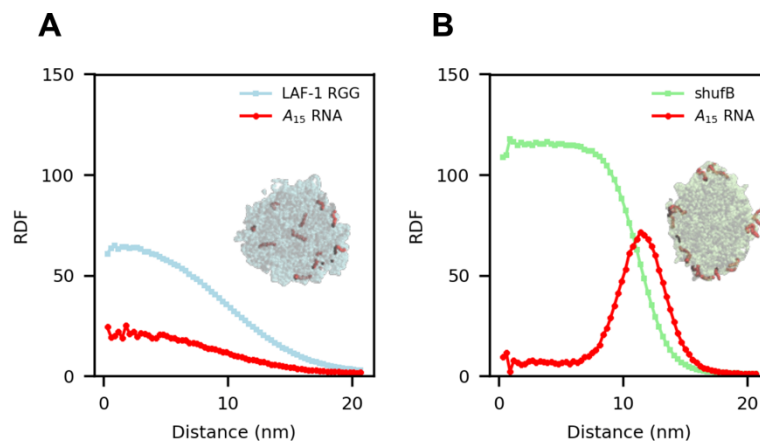
### ***Effect of protein sequence charge patterning on RNA partitioning inside the condensed phase***

In the last section, we showed how presence of RNA could modulate the LLPS of the LAF-1 RGG sequence. As our CG modelling strategy can be used to study sequence-dependent effects, our next goal was to investigate the role of the protein sequence in modulating RNA-protein LLPS. In our recent work on determining the sequence determinants of LAF-1 RGG phase separation, we showed that the patterning of the charged amino acids can play a critical role in controlling the LLPS properties. Based on earlier work highlighting the role of charged amino acid patterning on the conformational properties of IDPs (60), Sawle and Ghosh proposed a sequence charge decoration (SCD) parameter (61) to quantify the distribution of charged amino acids in a protein sequence. Highly negative SCD values for a protein sequence indicate segregation of positive and negative charges within the sequence and can be observed in sequences having patches of similarly charged amino acids. This generally enhances electrostatic intra and inter-molecular interactions because of the cooperativity of interactions between oppositely charged patches (31). This will result in an enhanced phase separation propensity for sequences with lower SCD values, which was tested successfully by a combination of *in silico*, *in vitro*, and *in vivo* techniques for the LAF-1 RGG sequence in our recent work (54). Smaller negative, or positive values of SCD indicate a sequence with evenly distributed positive and negative charges, or a large net charge, which would be the case for an unstructured RNA chain.



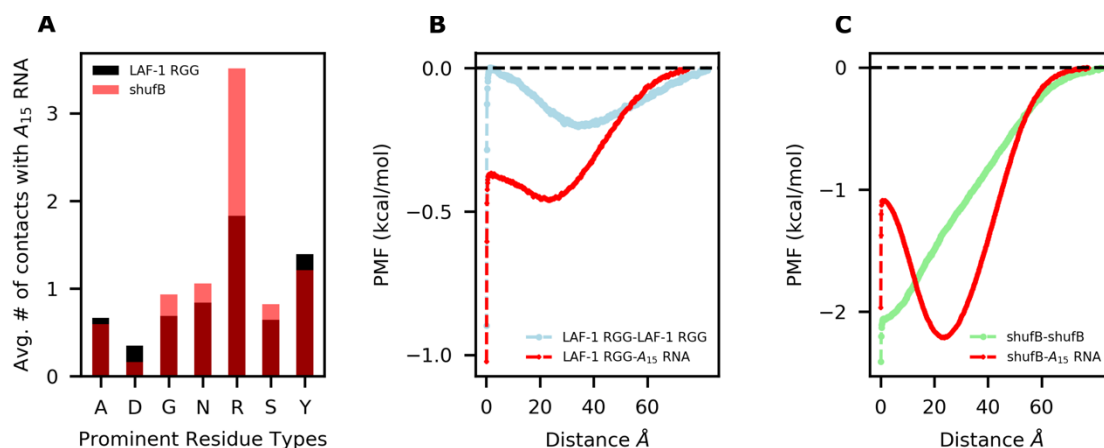
**Figure 5: Shuffling protein sequence shows sequence dependence of RNA-protein co-phase separation:** Increased charge patterning (lowering SCD) on LAF-1 RGG WT sequence to create shufB (B) Slab snapshots show RNA with preferential localization to the interface for shufB in contrast to the well mixed RNA/LAF-1 RGG case (C) Density profiles from MD simulations at the same simulation temperature and bulk concentrations show sharp interfacial peaks for RNA in shufB whereas RNA in LAF-1 RGG has a central flat peak.

To study the effects of SCD on a binary mixture of protein and RNA molecules, we use a variant of LAF-1 RGG from our previous work (37) where we had shuffled the sequence and selected an instance where positive and negative charges were highly segregated, having a large negative SCD (Fig. 5A). In the case of shufB, we observe significant accumulation of RNA molecules at the interface between the dense and dilute phases as opposed to a relatively well-mixed WT/RNA condensate (Fig. 5B, C). We then conducted additional simulations in a cubic simulation box of a spherical condensed phase in order to ascertain whether the observed microstructural features may be an artefact of the simulation geometry (Fig. 6, insets). We calculate the protein and RNA radial distribution functions (RDFs) with respect to the center of mass of the condensed phase as a spherical density profile (Fig. 6). As we found previously, the WT LAF-1 RGG condensate formed a well-mixed protein-RNA droplet (Fig. 6A), and the shufB condensate formed a core-shell structure, where RNA mostly forms a "shell" around the "core" consisting of shufB protein (Fig. 6B). This is reminiscent of other examples where sub compartmentalization has been observed (62–64), including ones that show similarly structured organization of different components in a mixture to create sub compartments within a condensate (57, 58).



**Figure 6: Protein sequence dependent co-phase separation persists in a droplet simulation geometry with intermolecular interactions causing sequence dependent behavior:** (A) Radial distribution functions (RDF) from droplet simulations also show a well-mixed condensate for the (A) RNA/LAF-1 RGG mixture and interfacial peaks for the (B) RNA/shufB mixture. Inset images show cross sections of these droplets where we can see RNA molecules (red) located throughout and at the interface for LAF-1 RGG and shufB respectively.

We then ask the question about the driving forces of these two distinct condensate architectures. Our hypothesis is that the observed core-shell architecture emerges from a complex interplay of homotypic protein-protein and heterotypic A<sub>15</sub> RNA - protein interactions. Looking into the RNA-protein interactions based on the proximity of beads during the simulation we see that the positively charged Arginine residues seem to be making the most contacts (Fig. 7 A) with the negatively charged nucleotides and this is even more pronounced with the shufB sequence as was expected due to the presence of higher charge patterning. Since RNA-RNA interactions will always be repulsive within our model, they would not be expected to contribute to stabilizing of any condensed phase.



**Figure 7: Intermolecular homotypic and heterotypic interactions play a role in modulating RNA-protein co-phase separation:** (A) Intermolecular contacts calculated on the basis of proximity of beads in the MD simulation show electrostatic interactions play a major role (B) Potentials of Mean Force for homotypic LAF-1 RGG and heterotypic LAF-1 RGG - RNA interactions and (C) for homotypic shufB and heterotypic shufB - RNA interactions show the formation of a second minima away from zero center of mass distance for shufB - A<sub>15</sub> RNA interactions linking intermolecular interactions to the observed structure in the shufB-RNA condensate.

We also quantify interactions between a pair of protein-protein or RNA-protein molecules using umbrella sampling Monte Carlo simulations similar to our previous work (30). We thus calculated the

potential of mean force (PMF) between the two molecules (protein-protein and A<sub>15</sub> RNA-protein) as a function of the distance between their centers of mass (Fig. 7 B, C). In all four cases, we see net attractive interactions between the two components, with a strong minimum at zero distance, which is accessible due to the disordered nature of the components. The free energy minimum is deeper in the case of shufB-shufB compared to the LAF-1 RGG – LAF-1 RGG, as would be expected based on their single-component phase behavior (30). We note that most of the PMFs have a secondary minimum at ~20-30Å, with the exception of shufB-shufB. Interestingly, this second minimum for A<sub>15</sub> RNA-shufB (~25 Å) is deeper than the first minimum at zero distance (Fig. 7 C), which is not the case for the other pairs A<sub>15</sub> RNA - LAF-1 RGG. The homotypic interactions between LAF-1 RGG is similar to the heterotypic interactions between LAF-1 RGG and A<sub>15</sub> RNA, so A<sub>15</sub> RNA is well mixed with LAF-1 RGG. However, the homotypic interactions between shufB is quite different from heterotypic interactions between shufB and A<sub>15</sub> RNA. This is mainly driven by shuffling the LAF-1 RGG sequence towards higher charge patterning, introducing prominent attractive shufB-shufB interactions. This explains the observation that shufB does not form a well-mixed condensate with A<sub>15</sub> RNA, and that in this case, RNA is majorly situated on the periphery of the shufB condensate (Fig. 5B, C). Comparing how similar the interactions between the homotypic scaffold interactions and the heterotypic scaffold-client interactions might be a useful strategy for checking the microscopic organization in MLOs (65).

## CONCLUSION

In this work we have developed a novel coarse-grained (CG) simulation model for capturing RNA-protein interactions and co-phase separation. We show how this simple RNA model is sufficient to reveal insights into the role RNA plays in modulating protein phase separation. We demonstrate this using the N-terminal disordered RGG domain of the LAF-1 protein, which has been shown to phase separate itself, and in the presence of disordered RNA molecules (23). Our simulations allow us to quantify the degree to which RNA is incorporated into the condensed protein phase and also show effects of RNA on the phase separation propensity of the protein in concordance with experiments (19).

In addition, as our protein model captures the sequence level detail of the disordered protein, it allows us to identify the role that the protein sequence plays in modulating RNA-protein co-phase separation. While keeping the overall amino acid composition constant we shuffled the LAF-1 sequence to identify variants with enhanced electrostatic interactions by creating a more charge-segregated protein sequence. We find that the enhanced electrostatic interactions resulting from this shuffling causes the protein to phase separate more strongly. Interestingly, even though protein-RNA interactions are also enhanced by this charge segregation, we do not observe an increase in RNA incorporation into the condensed phase. Instead, we see interfacial peaks indicating RNA accumulating around the shuffled protein condensate forming a layer between the molecule rich and molecule depleted vapor region in our simulations, and a lower overall concentration of RNA in the condensed phase.

We then looked at 2-molecule interactions by calculating the PMFs between the different molecular pairs. We propose that a complex interplay of homotypic and heterotypic interactions controls co-phase separation leading to the formation of structured droplets. The different structuring of the condensates can be explained by the difference in heterotypic shufB-RNA interactions, and homotypic shufB-shufB interactions, which increase to a greater degree upon shuffling, while heterotypic and homotypic interaction strengths are more balanced in the case of unshuffled LAF-1. Hence, our work here suggests the possibility of using charge patterning of protein sequence to control not only the phase diagram of co-phase separation of proteins and RNA molecules, but also introduce nonhomogeneous microstructures within the liquid droplet.

Even though our simplistic unstructured CG representation of RNA provides us with results in significant agreement with experiments, there are still some effects which one might not be able to capture due to the complex nature of interaction modes which can allow RNA to interact with proteins and itself. Hence, further work on this RNA CG representation is warranted as we would like to capture the more nuanced RNA interactions arising from the  $\pi$  character and secondary structure of RNA molecules which would also play important roles in modulating RNA-protein co-phase separation. Other aspects like improving the charge representation in the CG model can also be important as we show above that electrostatic interactions are a crucial part of how RNA interacts with proteins.

## FUNDING

Research was supported in part by NINDS and NIA R01NS116176, NSF 2004796, and NSF MRSEC Grant DMR 1720530. W.Z acknowledges the support from NSF 2015030. Y.C.K. is supported by the Office of Naval Research via the U.S. Naval Research Laboratory base program. Use of the high-performance computing capabilities of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the NSF grant TG-MCB-120014, is gratefully acknowledged.

## CONFLICT OF INTEREST

None declared.

## REFERENCES

1. Robertson,M.J., Tirado-Rives,J. and Jorgensen,W.L. (2017) Improved treatment of nucleosides and nucleotides in the OPLS-AA force field. *Chem. Phys. Lett.*, 10.1016/j.cplett.2017.02.049.
2. Brangwynne,C.P. (2013) Phase transitions and size scaling of membrane-less organelles. *J Cell Biol*, **203**, 875–881.
3. Gomes,E. and Shorter,J. (2019) The molecular language of membraneless organelles. *J. Biol. Chem.*, **294**, 7115–7127.
4. Anderson,P. and Kedersha,N. (2006) RNA granules. *J. Cell Biol.*, **172**, 803–808.
5. Eulalio,A., Behm-Ansmant,I. and Izaurralde,E. (2007) P bodies: At the crossroads of post-transcriptional pathways. *Nat. Rev. Mol. Cell Biol.*, **8**, 9–22.
6. Anderson,P. and Kedersha,N. (2002) Stressful initiations. *J. Cell Sci.*, **115**, 3227–3234.
7. Navarro,R.E. and Blackwell,T.K. (2005) Requirement for P granules and meiosis for accumulation of the germline RNA helicase CGH-1. *Genesis*, **42**, 172–180.
8. Krichevsky,A.M. and Kosik,K.S. (2001) Neuronal RNA granules: A link between RNA localization and stimulation-dependent translation. *Neuron*, **32**, 683–696.
9. Brangwynne,C.P., Eckmann,C.R., Courson,D.S., Rybarska,A., Hoege,C., Gharakhani,J., Jülicher,F. and Hyman,A.A. (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (80-. )*, **324**, 1729–1732.
10. Dignon,G.L., Best,R.B. and Mittal,J. (2020) Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annu. Rev. Phys. Chem.*, **71**, 53–75.
11. Elbaum-Garfinkle,S., Kim,Y., Szczepaniak,K., Chen,C.C.-H., Eckmann,C.R., Myong,S. and Brangwynne,C.P. (2015) The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 7189–94.
12. Burke,K.A., Janke,A.M., Rhine,C.L. and Fawzi,N.L. (2015) Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell*, **60**, 231–241.
13. Uversky,V.N. (2017) Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.*, 10.1016/j.sbi.2016.10.015.
14. Lin,Y., Protter,D.S.W., Rosen,M.K. and Parker,R. (2015) Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell*, **60**, 208–219.
15. Smith,J., Calidas,D., Schmidt,H., Lu,T., Rasoloson,D. and Seydoux,G. (2016) Spatial patterning of P granules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3. *Elife*, **5**, e21337.
16. Maharana,S., Wang,J., Papadopoulos,D.K., Richter,D., Pozniakovsky,A., Poser,I., Bickle,M., Rizk,S., Guillén-Boixet,J., Franzmann,T.M., *et al.* (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science (80-. )*, **360**, 918–921.
17. Aumiller,W.M., Pir Cakmak,F., Davis,B.W. and Keating,C.D. (2016) RNA-Based Coacervates as a Model for Membraneless Organelles: Formation, Properties, and Interfacial Liposome Assembly. *Langmuir*, **32**, 10042–10053.
18. Poudyal,R.R., Pir Cakmak,F., Keating,C.D. and Bevilacqua,P.C. (2018) Physical Principles and



Extant Biology Reveal Roles for RNA-Containing Membraneless Compartments in Origins of Life Chemistry. *Biochemistry*, **57**, 2509–2519.

19. Banerjee, P.R., Milin, A.N., Moosa, M.M., Onuchic, P.L. and Deniz, A.A. Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets *Angewandte*. 10.1002/anie.201703191.
20. Murthy, A.C., Dignon, G.L., Kan, Y., Zerbe, G.H., Parekh, S.H., Mittal, J. and Fawzi, N.L. (2019) Molecular interactions underlying liquid–liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.*, **26**, 637–648.
21. Martin, E.W., Holehouse, A.S., Peran, I., Farag, M., Incicco, J.J., Bremer, A., Grace, C.R., Soranno, A., Pappu, R. V. and Mittag, T. (2020) Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* (80-. ), **367**, 694–699.
22. Wang, J., Choi, J.M., Holehouse, A.S., Lee, H.O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., *et al.* (2018) A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*, **174**, 688–699.e16.
23. Wei, M.T., Elbaum-Garfinkle, S., Holehouse, A.S., Chen, C.C.H., Feric, M., Arnold, C.B., Priestley, R.D., Pappu, R. V. and Brangwynne, C.P. (2017) Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat. Chem.*, **9**.
24. Lin, Y., McCarty, J., Rauch, J.N., Delaney, K.T., Kosik, K.S., Fredrickson, G.H., Shea, J.E. and Han, S. (2019) Narrow equilibrium window for complex coacervation of tau and RNA under cellular conditions. *Elife*, **8**, 1–31.
25. Mittag, T. and Forman-Kay, J.D. (2007) Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.*, **17**, 3–14.
26. Mitrea, D.M., Chandra, B., Ferrolino, M.C., Gibbs, E.B., Tolbert, M., White, M.R. and Kriwacki, R.W. (2018) Methods for Physical Characterization of Phase-Separated Bodies and Membrane-less Organelles. *J. Mol. Biol.*, **430**, 4773–4805.
27. Rauscher, S. (2017) The liquid structure of elastin aggregates. *Biophys. Struct. Biol.*
28. Conicella, A.E., Dignon, G.L., Zerbe, G.H., Schmidt, H.B., D'Ordine, A.M., Kim, Y.C., Rohatgi, R., Ayala, Y.M., Mittal, J. and Fawzi, N.L. (2019) TDP-43  $\alpha$ -helical structure tunes liquid-liquid phase separation and function. *bioRxiv*, 10.1101/640615.
29. Dignon, G.L., Zheng, W. and Mittal, J. (2019) Simulation methods for liquid–liquid phase separation of disordered proteins. *Curr. Opin. Chem. Eng.*, **23**, 92–98.
30. Dignon, G.L., Zheng, W., Best, R.B., Kim, Y.C. and Mittal, J. (2018) Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci.*, **115**, 9929–9934.
31. Lin, Y.H., Brady, J.P., Forman-Kay, J.D. and Chan, H.S. (2017) Charge pattern matching as a ‘fuzzy’ mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J. Phys.*, **19**.
32. Espinosa, J.R., Joseph, J.A., Sanchez-Burgos, I., Garaizar, A., Frenkel, D. and Collepardo-Guevara, R. (2020) Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. *Proc. Natl. Acad. Sci.*, **117**, 201917569.



33. Dignon, G.L., Zheng, W., Kim, Y.C., Best, R.B. and Mittal, J. (2018) Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.*, **14**, e1005941.
34. Dignon, G.L., Zheng, W., Kim, Y.C. and Mittal, J. (2019) Temperature-Controlled Liquid-Liquid Phase Separation of Disordered Proteins. *ACS Cent. Sci.*, **5**, 821–830.
35. Kapcha, L.H. and Rossky, P.J. (2014) A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.*, **426**, 484–498.
36. Perdikari, T.M., Jovic, N., Dignon, G.L., Kim, Y.C., Fawzi, N.L. and Mittal, J. (2020) A predictive coarse-grained model for position-specific effects of post- translational modifications on disordered protein phase separation.
37. Schuster, B.S., Dignon, G.L., Tang, W.S., Kelley, F.M., Ranganath, A.K., Jahnke, C.N., Simpkins, A.G., Regy, R.M., Hammer, D.A., Good, M.C., *et al.* (2020) Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proc. Natl. Acad. Sci.*, **117**, 202000223.
38. Creighton, T.E. (1993) Proteins: structures and molecular properties Macmillan.
39. Debye, P. and Hückel, E. (1923) De la theorie des electrolytes. I. abaissement du point de congelation et phenomenes associes. *Phys. Zeitschrift*, **24**, 185–206.
40. Ashbaugh, H.S. and Hatch, H.W. (2008) Natively unfolded protein stability as a coil-to-globule transition in charge/hydrophathy space. *J. Am. Chem. Soc.*, **130**, 9536–9542.
41. Šulc, P., Romano, F., Ouldrige, T.E., Doye, J.P.K. and Louis, A.A. (2014) A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.*, **140**, 235102.
42. Denesyuk, N.A. and Thirumalai, D. (2013) Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B*, **117**, 4901–4911.
43. Pasquali, S. and Derreumaux, P. (2010) HiRE-RNA: A high resolution coarse-grained energy model for RNA. *J. Phys. Chem. B*, **114**, 11957–11966.
44. Ghobadi, A.F. and Jayaraman, A. (2016) Effect of backbone chemistry on hybridization thermodynamics of oligonucleic acids: A coarse-grained molecular dynamics simulation study. *Soft Matter*, **12**, 2276–2287.
45. Robertson, M.J., Qian, Y., Robinson, M.C., Tirado-Rives, J. and Jorgensen, W.L. (2019) Development and Testing of the OPLS-AA/M Force Field for RNA. *J. Chem. Theory Comput.*, **15**, 2734–2742.
46. De Ruyter, A. and Zagrovic, B. (2015) Absolute binding-free energies between standard RNA/DNA nucleobases and amino-acid sidechain analogs in different environments. *Nucleic Acids Res.*, **43**, 708–718.
47. Potoff, J.J. and Panagiotopoulos, A.Z. (1998) Critical point and phase behavior of the pure fluid and a Lennard-Jones mixture. *J. Chem. Phys.*, **109**, 10914–10920.
48. Panagiotopoulos, A.Z. (2000) Monte Carlo methods for phase equilibria of fluids. *J. Phys. Condens. Matter*, **12**, R25.
49. Kim, J., Keyes, T. and Straub, J.E. (2010) Generalized replica exchange method. *J. Chem. Phys.*, **132**, 224107.
50. Blas, F.J., MacDowell, L.G., de Miguel, E. and Jackson, G. (2008) Vapor-liquid interfacial properties

- of fully flexible Lennard-Jones chains. *J. Chem. Phys.*, **129**, 144703.
51. Silmore, K.S., Howard, M.P. and Panagiotopoulos, A.Z. (2017) Vapour–liquid phase equilibrium and surface tension of fully flexible Lennard–Jones chains. *Mol. Phys.*, **115**, 320–327.
52. Anderson, J.A., Glaser, J. and Glotzer, S.C. (2020) HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput. Mater. Sci.*, **173**, 109363.
53. Mittal, J. (2020) Theory of biological phase separation, Liquid-Liquid Phase Coexistence and Membraneless Organelles. In Keating, C.D. (ed), *Liquid-Liquid Phase Coexistence and Membraneless Organelles*.
54. Schuster, B.S., Dignon, G.L., Tang, W.S., Kelley, F.M., Ranganath, A.K., Jahnke, C.N., Simpkins, A.G., Regy, R.M., Hammer, D.A., Good, M.C., *et al.* (2020) Identifying Sequence Perturbations to an Intrinsically Disordered Protein that Determine Its Phase Separation Behavior. *bioRxiv*, 10.1101/2020.01.06.894576.
55. Dill, K.A. and Bromberg, S. (2010) Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience Garland Science.
56. Choi, J.-M., Holehouse, A.S. and Pappu, R. V. (2020) Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.*, **49**, 107–133.
57. Ghosh, A., Zhang, X. and Zhou, H.X. (2020) Tug of war between condensate phases in a minimal macromolecular system. *J. Am. Chem. Soc.*, **142**, 8848–8861.
58. Alshareedah, I., Moosa, M.M., Raju, M., Potoyan, D. and Banerjee, P.R. (2020) Phase Transition of RNA-protein Complexes into Ordered Hollow Condensates. *bioRxiv*, 10.1101/2020.01.10.902353.
59. Alshareedah, I., Kaur, T., Ngo, J., Seppala, H., Kounatse, L.A.D., Wang, W., Moosa, M.M. and Banerjee, P.R. (2019) Interplay between Short-Range Attraction and Long-Range Repulsion Controls Reentrant Liquid Condensation of Ribonucleoprotein-RNA Complexes. *J. Am. Chem. Soc.*, **141**, 14593–14602.
60. Das, R.K. and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 13392–13397.
61. Sawle, L. and Ghosh, K. (2015) A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.*, **143**.
62. Feric, M., Vaidya, N., Harmon, T.S., Mitrea, D.M., Zhu, L., Richardson, T.M., Kriwacki, R.W., Pappu, R. V. and Brangwynne, C.P. (2016) Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, **165**, 1686–1697.
63. Putnam, A., Cassani, M., Smith, J. and Seydoux, G. (2019) granules in *Caenorhabditis elegans* embryos. *Nat. Struct. Mol. Biol.*, **26**.
64. Fei, J., Jadaliha, M., Harmon, T.S., Li, I.T.S., Hua, B., Hao, Q., Holehouse, A.S., Reyher, M., Sun, Q., Freier, S.M., *et al.* (2017) Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.*, **130**, 4180–4192.
65. Gasior, K., Zhao, J., McLaughlin, G., Forest, M.G., Gladfelter, A.S. and Newby, J. (2019) Partial

demixing of RNA-protein complexes leads to intradroplet patterning in phase-separated biological condensates. *Phys. Rev. E*, **99**, 1–11.