Published online 2 December 2020 *Nucleic Acids Research, 2020, Vol. 48, No. 22* **12593–12603**
doi: 10.1093/nar/gkaa1099

Sequence dependent phase separation of protein-polynucleotide mixtures elucidated using molecular simulations

Roshan Mammen Regy[1], Gregory L. Dignon[1], Wenwei Zheng[2], Young C. Kim[3] and Jeetain Mittal[1,*]

[1]Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, PA 18015, USA, [2]College of Integrative Sciences and Arts, Arizona State University, Mesa, AZ 85212, USA and [3]Center for Materials Physics and Technology, Naval Research Laboratory, Washington, DC 20375, USA

Received July 07, 2020; Revised October 21, 2020; Editorial Decision October 26, 2020; Accepted October 27, 2020

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 and other polynucleotides 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 polynucleotides play the role of a modulator/promoter of LLPS in cells using computational methods. Here, we present a coarse-grained polynucleotide model developed to fill this gap, which together with our recently developed HPS model for protein LLPS, allows us to capture the factors driving protein-polynucleotide phase separation. We explore the capabilities of the modelling framework with the LAF-1 RGG 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 polynucleotide incorporation into condensates. With increased charge patterning we observe formation of structured or patterned condensates which suggests the possible roles of polynucleotides 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 and 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

*To whom correspondence should be addressed. Tel: +1 610 758 4791; Fax: +1 610 758 5057; Email: jeetain@lehigh.edu
Present address: Gregory L. Dignon, Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA.

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 multicomponent 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 behind RNA–protein 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 allow investigations into the formation of biomolecular condensates and 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 amino acid hydropathy (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 hydropathy values of the naturally occurring twenty amino acids (33,35). This model was recently extended to post-translationally modified amino acids as well (36).

Here, we present a simplistic polynucleotide model which can capture electrostatic interactions between RNA and protein that has been shown to be important for phase separating RNA-protein mixtures in recent experimental studies (1–3). The model can also account for non-electrostatic interactions with amino acids based on their calculated hydropathy values. The effects of RNA secondary structure such as the formation of G quadruplexes on protein phase separation, which is non-trivial to capture in an all-atom or coarse-grained models (40), can also be important in many cases (41–44). To capture such effects, more rigorous development of detailed RNA models that are compatible with our coarse-grained protein model will be necessary. Here, we instead focus on issues that can be resolved within the limitations of our current CG model.

We first define parameters for RNA nucleotides within the HPS modeling framework using a one bead per nucleotide CG representation. We then test the new model to study the 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 phase separation of polynucleotides and LAF-1 RGG protein variants with the same sequence composition but different arrangement of charged amino acids used in our previous work (45). Consistent with the expectations from previous work (31,45), we find that the phase behavior can be significantly perturbed by changes in the protein sequence, and this change in behavior also alters phase separation of the protein-polynucleotide mixture. Most importantly, we observe a significant change in terms of the co-localization of protein and polynucleotide molecules within the condensate. For the charge segregated variants of the LAF-1 RGG protein, polynucleotides adsorb on the protein condensate surface rather than mixing evenly throughout as in the condensate of the wild type (WT) LAF-1 RGG and polynucleotide mixture. We bring forth an interpretation for the observed behavior based on the potential of mean force between pairs of protein and polynucleotide molecules which provides us with detailed information about heterotypic and homotypic interactions in the mixture. Our work here provides evidence linking this microscopic information about the intermolecular heterotypic and homotypic interactions to the formation of multicomponent condensates, allowing us to better understand what determines selectivity in the incorporation of biomolecules and the morphology of MLOs.

MATERIALS AND METHODS

HPS model for proteins

Our CG modeling approach for proteins uses a one bead per amino acid level of resolution, and the 20 naturally occurring amino acids are identified in the model by the following characteristics: the mass, charge, diameter (σ) (46) and hydropathy (λ) of the amino acid. The λ 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 Rossky (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^2 and a bond length of 3.8 Å. The electrostatic interactions are modeled using the Debye–Hückel (DH) electrostatic screening term (47),

$$E_{i,j} = \frac{q_i q_j}{4\pi\, Dr}\ \exp\left(-\frac{r}{\kappa}\right) \quad (1)$$

in which κ 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 (48) as

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

where Φ_{LJ} is the standard 12–6 Lennard–Jones potential,

$$\Phi_{LJ} = 4\epsilon\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right]. \quad (3)$$

Both the hydropathy (λ_{ij}) and the diameter (σ_{ij}) are set to be the arithmetic average of the values for the interacting pair of amino acids (i,j). A value of λ_{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 (ε) in (Equation 3) is a free parameter and was adjusted to reproduce the experimentally measured radii of gyration of IDPs as 0.2 kcal/mol (33).

HPS model for polynucleotides

Our purpose for building a new polynucleotide model, as opposed to using an existing model (49–51), is to capture the qualitative features of the protein-polynucleotide phase separation behavior with the simplest possible CG representation that is also consistent with our protein model. Hence, we decided to model polynucleotides with each nucleotide represented as a single particle (Figure 1A). The potential energy 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^2 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 (52).

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 (Equation 2). The hydropathy 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 (53). The RNA nucleotide hydropathy 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. 20 amino acids or four RNA nucleotides) is used for characterizing the short-range pairwise interactions. Most of the protein–polynucleotide short range interactions are weakly attractive with the most significant attraction between positively charged amino acids and nucleotides (Supplementary Tables S1 and S2).

Table 1. Polynucleotide model parameters for CG beads, each representing a single nucleotide

Nucleotide	Mass (amu)	Charge (1)	Diameter, σ (Å)	Hydropathy, λ (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

Simulation strategy for sampling protein–polynucleotide 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 (58) to sample the phase behavior of proteins undergoing LLPS successfully and efficiently (20,32). Here, we use the same strategy, as shown in Figure 1A, 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 (Figure 1A). 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 (30,57,58). By conducting such simulations for several temperatures, we can use the co-existing densities (Figure 1B) to map the binodal (limits of thermodynamic stability) and the critical temperature of the phase separating component (Figure 1D). As mentioned in our 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 (33). We then extend the box z-dimension to 280 nm, and conduct coexistence simulations in the NVT ensemble for 5 μs. We have shown previously that starting the system from a configuration where all chains are dispersed throughout the box converges to the same results as when starting with a condensed phase initially (33). 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) (59). For more details about the methodology and the analysis, we refer the readers to our previous work (33,60). We provide pertinent details of the simulated systems in Supplementary Table S3. We also provide a python code (https://bitbucket.org/rom218/polynucleotide_protein_sim), which can be used to setup a protein-polynucleotide simulation

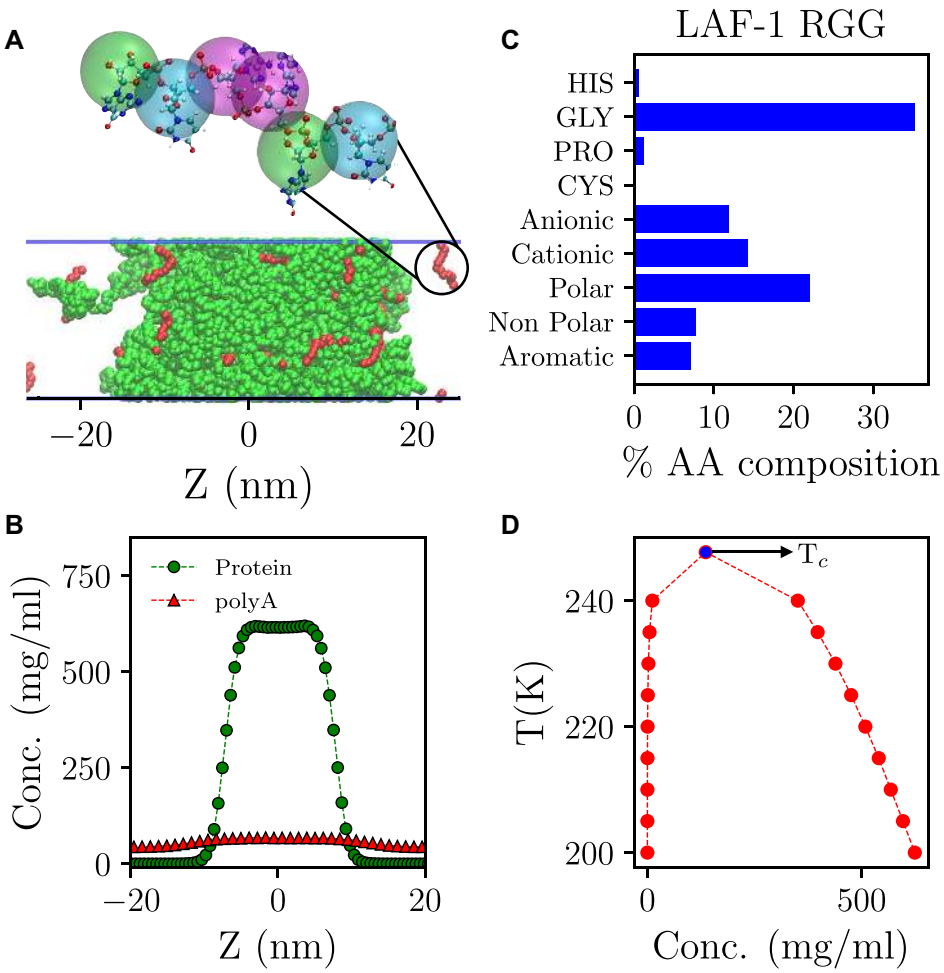

Figure 1. Slab simulation methodology for protein–polynucleotide mixtures. (**A**) Slab configuration for CG simulations of protein (green)-polynucleotide (red, zoomed inset). (**B**) Density profiles of LAF-1 RGG and polyA sampled over the course of a slab simulation centered on the largest cluster formed. (**C**) Amino acid composition of the LAF-1 RGG WT sequence. (**D**) Binodal phase diagram showing temperature regime under which LAF-1 RGG phase separation is possible.

with a user provided protein sequence, length of polyA and with the interactions parametrized according to the HPS model described in this work.

RESULTS AND DISCUSSION

Polynucleotides perturb LAF-1 RGG phase separation in a concentration-dependent manner

We conducted coexistence simulations of the disordered LAF-1 RGG protein sequence with 15 nucleotide long polyadenosine (A_{15}) at a range of temperatures and concentrations of A_{15} while 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 (aqueous) phase (45), whereas the polynucleotide alone is not able to phase separate, even at relatively high concentrations (∼36mg/ml) due to electrostatic repulsion. However, upon phase separation of a binary mixture of LAF-1 RGG and A_{15}, A_{15} molecules are preferentially found inside the LAF-1 RGG-rich phase rather than the dilute region of the simulation box (Figure 2B). We therefore quantify the preference of molecules by calculating the partition coefficient of the A_{15} molecules - the ratio of the concentration of a component in the dense phase to that in the dilute phase. A high partition coefficient suggests the A_{15} molecule prefers to partition into the protein condensate whereas a small value suggests the A_{15} molecule tends to stay dispersed in the protein-deficient phase, and a value of 1 would indicate that A_{15} molecules disperse evenly inside and outside the condensate. The partitioning of the A_{15} molecule into the LAF-1 RGG condensed phase can lead to condensed phase A_{15} concentrations up to two orders of magnitude higher than its vapor phase concentration (Figure 2C). This is likely due to the presence of a relatively large number of cationic Arginine residues in the LAF-1 RGG sequence (Figure 1C) which can form favorable interactions with the anionic adenosine nucleotides. We also confirmed that the simulation run lengths are long enough to obtain converged data with respect to the densities of both components in the dense and dilute phases (see Movies M1, M2).

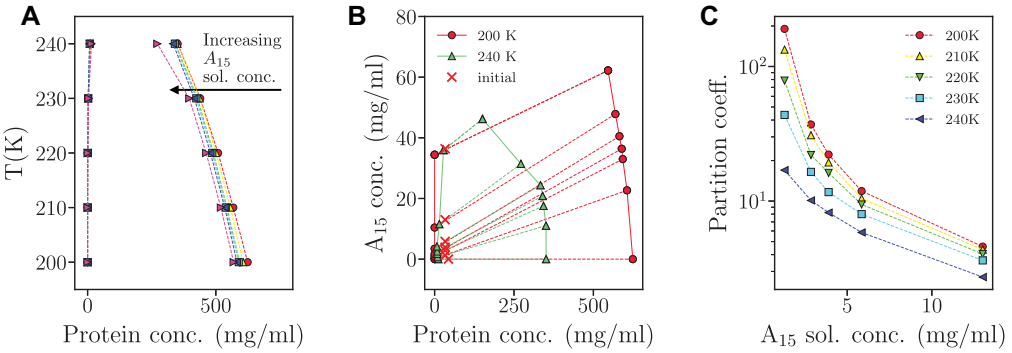

Figure 2. A_{15} and LAF-1 RGG phase separate together to form A_{15} modulated protein condensates: (**A**) LAF-1 RGG binodal phase diagram shifts with increasing A_{15} solution concentration. (**B**) Binary phase diagram from simulations with varying solution A_{15} concentrations (colored (dashed) tie lines) joining the coexistence concentrations of A_{15} and protein at different temperatures (colored (solid lines) phase envelopes). Red X's indicate total concentrations of simulation systems. Tie lines will always pass through these as long as the system separates into two phases. (**C**) Partition coefficients quantifying incorporation of A_{15} into the LAF-1 RGG condensed phase from the dilute phase at different temperatures and increasing solution concentrations of A_{15}.

To observe the effect of A_{15} on the thermodynamics of LAF-1 RGG phase separation, we calculated the coexistence densities of the protein at different temperatures and A_{15} solution concentrations (Figure 2A). With increasing A_{15} solution concentration, the right arm of the binodal that represents protein density in the condensed phase shifts to the left. This suggests that the protein concentration inside this dense phase decreases which is presumably due to the increasing A_{15} concentration inside. This reduction in protein concentration with increasing A_{15} in the case of LAF-1 RGG is also consistent with previous experimental work on the same system by Elbaum-Garfinkle *et al.* (23).

It is interesting to note that the total biomolecular concentration (protein and polynucleotide) in the condensed phase is mostly unchanged in our simulations at a set temperature (Supplementary Figure S2). There is significant interest in understanding the physicochemical characteristics of biomolecular condensates as their function is expected to be perturbed by changes in the underlying microstructural features and dynamics. At this point, it is unclear if the lack of significant changes in the condensed phase concentration due to polynucleotide incorporation is due to the simplicity of our CG model in which excluded volume interactions are relatively similar between different components or a more general consequence of the liquid-like nature of biomolecular condensates.

Next, we determined the degree to which A_{15} is incorporated into the condensates by using the A_{15} and protein concentrations from both the high- and low-density coexisting phases to construct the binary thermodynamic phase diagram (Figure 2B). 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_{15} concentrations falls within the coexistence region, the system will phase separate into two phases with relative concentrations of protein and A_{15} dictated by the tie lines joining this point to both dense and dilute phase arms of the multicomponent phase diagram (61). Based on the computed phase diagram in Figure 2B, we observe a 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 attractions. This condensed protein phase then recruits the 'client' A_{15} molecules for which homotypic interactions are either too weak or repulsive, but the heterotypic interactions between the scaffold and client molecules are sufficiently attractive (10,62). 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 also 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_{sat}), with the changes in client (A_{15}) solution concentration. The significance of C_{sat} is that this is the minimum protein concentration required to undergo LLPS, so it is important to understand how the presence of polynucleotides can be used to modulate this at a fixed temperature. Multivalency of the two components in our system is also an aspect which requires investigation as multivalent interactions have been shown to be responsible (20,45,62,63) for imparting increased phase separation propensity to biomolecules and one would like to see if this also applies to phase separation of protein-polynucleotide mixtures. To observe the effects of perturbing the multivalency of protein-polynucleotide interactions, we also simulated various lengths of polyA, i.e. A_5, A_{10} and A_{30} apart from the A_{15} case (which we discuss throughout this work) at the same nucleotide 'bead' solution concentration by changing the number of chains in the mixture to account for the change in length and at a fixed protein solution concentration and temperature (Supplementary Table S4). In Figure 3, we show how the LAF-1 RGG C_{sat} changes with increasing polynucleotide solution concentration for different lengths of the polynucleotide. At low polynucleotide solution concentrations, there is an initial decrease in the C_{sat} values (by as much as a factor of two for the longer A_{15}, A_{30} polynucleotides) highlighting LAF-1 RGG's enhanced propensity to phase separate

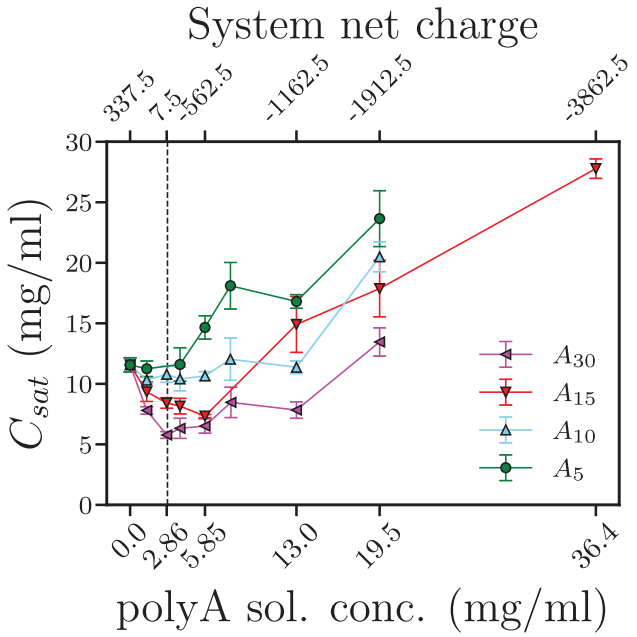

Figure 3. Polynucleotides show concentration dependent protein LLPS promoter/disruptor roles for various chain lengths. Effect of increasing polyA solution concentration on LAF-1 RGG saturation concentration shows switching between polyA LLPS promoter/disruptor roles (non-monotonic curve) for different chain lengths at a temperature of 240 K. Vertical dashed line represents a net charge = 0 system.

in the presence of polynucleotides. This effect is less pronounced for shorter polynucleotides (A_5, A_{10}), hinting towards the important role multivalency plays in modulating protein–polynucleotide interactions which can affect LAF-1 RGG's phase separation propensity. Similar LLPS behavior has also been observed previously in many other cases (19,37,64).

Upon a further increase in the polynucleotide concentration beyond a certain threshold value, we observe an increase in C_{sat} for all lengths of the polynucleotide, highlighting LAF-1 RGG's reduced propensity to phase separate beyond an optimum polynucleotide solution concentration. In fact, at much greater polynucleotide solution concentrations, the LAF-1 RGG C_{sat} can be higher than in the absence of polynucleotides. The concentration at which the polynucleotide 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 find that it is usually shifted away from this point, possibly due to added contributions from the non-electrostatic, multivalent hydrophobic interactions between nucleotides and amino acids (Supplementary Tables S1 and S2). Previous experimental studies by the Banerjee group have provided remarkable insights into this type of non-monotonic change in C_{sat} with increasing RNA concentration (38). 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 *et al.* 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 (38).

Effect of protein sequence charge patterning on polynucleotide partitioning inside the condensed phase

In the previous section, we showed how presence of polynucleotides 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 protein–polynucleotide 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 (45). Based on earlier work highlighting the role of charged amino acid patterning on the conformational properties of IDPs (65), Sawle and Ghosh proposed a sequence charge decoration (SCD) parameter (66) to quantify the distribution of charged amino acids in a protein sequence. SCD is calculated based solely on the amino acid sequence as $SCD = \frac{1}{N}\sum_{i=1}^{N-1}\sum_{j=2}^{N} q_i q_j \sqrt{j-i}$ where q_i and q_j are the charges of residues i and j. 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 (45). Smaller negative values of SCD indicate a sequence with evenly distributed positive and negative charges, and positive values indicate a significant net charge, which would be the case for an unstructured RNA chain.

To study the effects of SCD on a binary mixture of protein and polynucleotide molecules, we use a variant of LAF-1 RGG from our previous work (45) called LAF-1 RGG_{Cshuf} or RGG_{Cshuf} where we had shuffled the sequence and selected an instance where positive and negative charges were highly segregated, having a large negative SCD (Figure 4A). We conducted coexistence simulations at the same temperature and solution concentrations of RGG/A_{15} and RGG_{Cshuf}/A_{15} to isolate the effect of changing the protein sequence and charge distribution. We observe that in the case of LAF-1 RGG_{Cshuf}, there is significant accumulation of A_{15} molecules at the interface between the dense and dilute phases as opposed to a relatively well-mixed LAF-1 RGG WT/A_{15} condensate (Figure 4B, C). This phenomenon was consistently observed for different polynucleotide solution concentrations (Sup-

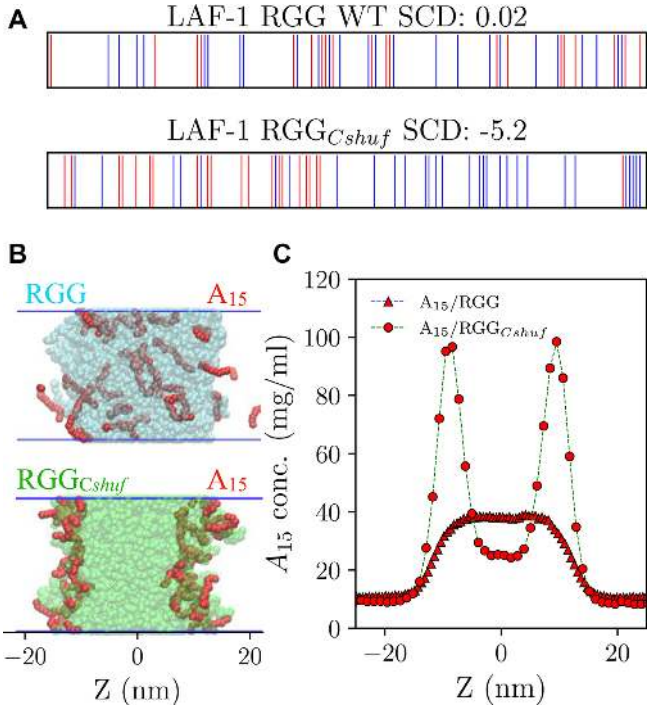

Figure 4. Shuffling protein sequence shows sequence dependence of protein-polynucleotide phase separation: (**A**) Increased charge patterning (lowering SCD) on LAF-1 RGG WT sequence to create LAF-1 RGG$_{Cshuf}$ where vertical lines represent anions (red) and cations (blue) (**B**) Slab snapshots show A_{15} with preferential localization to the interface for LAF-1 RGG$_{Cshuf}$ in contrast to the well mixed A_{15}/LAF-1 RGG case (**C**) Density profiles from MD simulations at the same simulation temperature (230 K) and bulk concentrations (protein: 33.33 mg/ml and A_{15}: 13.1 mg/ml) show sharp interfacial peaks for A_{15} in LAF-1 RGG$_{Cshuf}$ whereas A_{15} in LAF-1 RGG has a central flat peak.

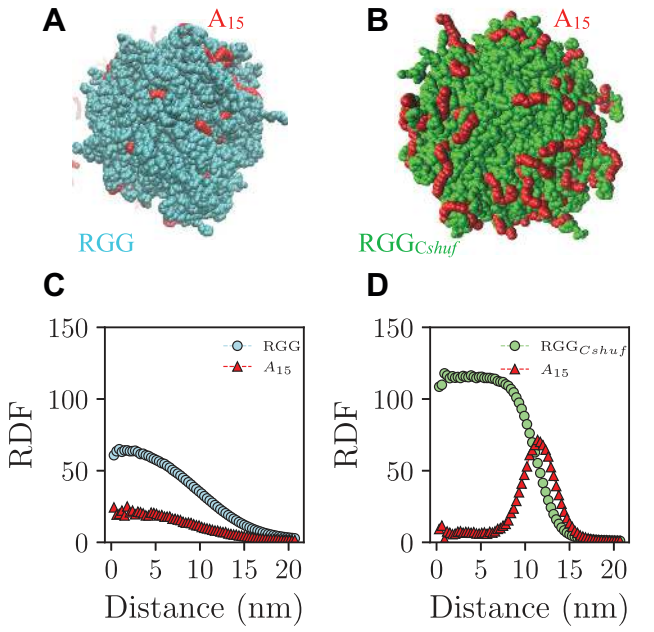

Figure 5. Protein sequence dependent protein–polynucleotide phase separation persists in a droplet simulation: (**A**) Snapshot of the A_{15}/LAF-1 RGG condensate and (**B**) snapshot of the A_{15}/LAF-1 RGG$_{Cshuf}$ condensate. Radial distribution functions (RDF) from droplet simulations also show a well-mixed condensate for the (**C**) A_{15}/LAF-1 RGG mixture and interfacial peaks for the (**D**) A_{15}/LAF-1 RGG$_{Cshuf}$ mixture at 220K.

plementary Figure S6) at different temperatures and different chain lengths of the nucleotide (Supplementary Figures S4 and S5). We then conducted additional simulations in a cubic simulation box of a spherical condensed phase at a fixed temperature for both sequences in order to ascertain whether the observed microstructural features were may be an artefact of the simulation geometry (Figure 5 A, B). We calculated the protein and polynucleotide radial distribution functions (RDFs) with respect to the center of mass of the condensed phase as a spherical density profile (Figure 5C, D). As we found previously, the WT LAF-1 RGG condensate formed a well-mixed protein-polynucleotide droplet (Figure 5A,C), and the LAF-1 RGG$_{Cshuf}$ condensate formed a core-shell structure, where A_{15} mostly forms a 'shell' around the 'core' consisting of LAF-1 RGG$_{Cshuf}$ protein (Figure 5B,D). This is reminiscent of other examples where sub compartmentalization has been observed (67–69), including ones that show similarly structured organization of different components in a mixture to create sub compartments within a condensate (37,64). This prompted us to investigate further into the driving forces which make these distinct condensate architectures possible.

Our initial hypothesis was that the observed core-shell architecture emerges from a complex interplay of homotypic protein-protein and heterotypic protein-A_{15} interactions. Looking into the protein–A_{15} 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 (Figure 6A) with the negatively charged nucleotides and this is even more pronounced with the LAF-1 RGG$_{Cshuf}$ sequence as was expected due to the presence of higher charge patterning. Since A_{15}–A_{15} interactions are always repulsive within our model, they would not be expected to contribute to the stabilization of the condensed phase.

To delve further into investigating the role of homotypic and heterotypic interactions in our protein–polynucleotide mixtures we also quantified interactions between a pair of protein-protein or protein-polynucleotide 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 protein–A_{15}) as a function of the distance between their centers of mass (Figure 6 B, C) at 230 K.

In all four cases, we see net attractive interactions between the two components. There are significant differences in the PMF profiles though, both in terms of the minimum well-depth and the location of this minimum. Homotypic protein-protein interactions are considerably stronger in the case of RGG$_{Cshuf}$ compared to RGG (Figure 6B, C; black). The stronger homotypic interactions are expected for LAF-1 RGG$_{Cshuf}$ due to the greater segregation of similarly charged residues in this sequence that leads to stronger intra- and intermolecular attractions in polyampholytic proteins composed of cationic and anionic residues (66,70). This behavior is also consistent with the changes

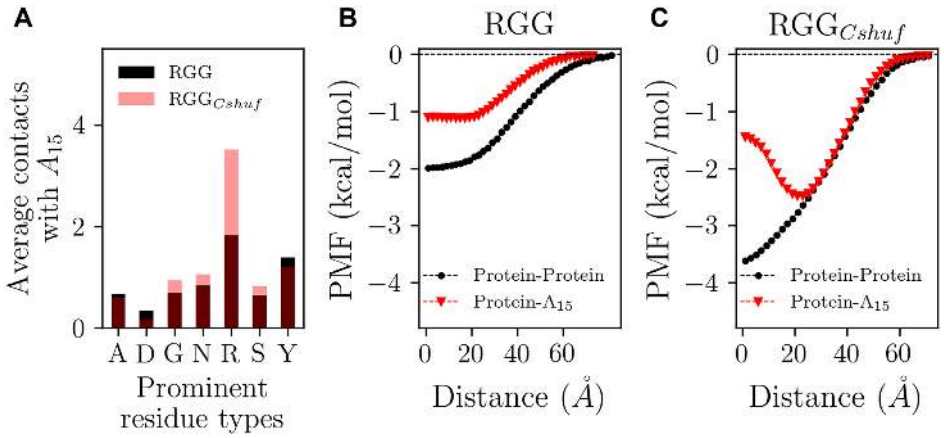

Figure 6. Intermolecular homotypic and heterotypic interactions play a role in modulating protein-polynucleotide phase separation: (**A**) Intermolecular contacts calculated on the basis of proximity of beads in the coexistence simulation at 230 K show electrostatic interactions play a major role (**B**) Potentials of Mean Force for homotypic LAF-1 RGG and heterotypic LAF-1 RGG–polynucleotide interactions at 230 K and (**C**) for homotypic LAF-1 RGG$_{Cshuf}$ and heterotypic LAF-1 RGG$_{Cshuf}$–polynucleotide interactions at 230 K show the formation of a second minima away from zero center of mass distance for LAF-1 RGG$_{Cshuf}$– polynucleotide interactions linking intermolecular interactions to the observed structure in the LAF-1 RGG$_{Cshuf}$–polynucleotide condensate.

in the phase diagram between the LAF-1 RGG and LAF-1 RGG$_{Cshuf}$ sequences (Figure 2 and Supplementary Figure S1), as would also be expected based on correlation between two-body interactions and LLPS (30). Notably, the free energy minima for homotypic interactions are at zero distance, which is possible due their disordered polymeric nature (71,72).

For the heterotypic interactions of these proteins with A_{15} we see a marked difference in the free energy profiles (Figure 6B,C; red) which may provide clues toward the driving forces behind the observed polynucleotide patterning behavior in Figures 4 and 5. In the case of LAF-1 RGG WT(Figure 6B; red), the free energy minimum is a uniform region from zero to about 20Å distance between the protein and polynucleotide chain which agrees with the uniform incorporation of the polynucleotide into the LAF-1 RGG WT condensate seen in Figure 4B,C. In the case of LAF-1 RGG$_{Cshuf}$, the heterotypic interactions have a more peculiar profile (Figure 6C; red) where we see a prominent free energy minimum at ∼20 Å distance between the centers of mass of the two chains. This indicates the existence of a more favorable configuration for these two components where they are a certain distance away from each other. The presence of this more favorable configuration away from zero distance could be linked to why we observe a prominent presence of polynucleotides on the outer surface or shell, which has a lower concentration of protein, compared to the protein-enriched core of the condensate. We find this free energy minimum forming away from zero distance for heterotypic interactions when we shuffled the residues on the LAF-1 RGG WT to create patches of similarly charged residues (Figure 4A) in LAF-1 RGG$_{Cshuf}$ which increased segregation of charged residues and cooperative electrostatic interactions (Figure 6A). For heterotypic interactions with negatively charged polynucleotides, this charge patterning would imply both stronger attractive and repulsive interactions with positive and negatively charged patches respectively on the LAF-1 RGG$_{Cshuf}$ sequence. Hence a more favorable configuration for the LAF-1 RGG$_{Cshuf}$–A_{15} chains would be at a certain distance where the attractive interaction contributions to the free energy are more prominent (The net charge of LAF-1 RGG and hence LAF-1 RGG$_{Cshuf}$ being positive might also be playing a role here) than repulsive interaction contributions which could be the cause of the presence of the free energy minima forming at a certain distance away from the zero distance point between their centers of mass.

As multivalency of the polynucleotide is another factor which could potentially change these intermolecular heterotypic interaction strengths and free energy profiles we also performed two body interaction studies with different polynucleotide lengths we previously studied and presented in Figure 3. Increasing the chain length and hence multivalency of the polynucleotide while keeping nucleotide solution concentration constant for different chain lengths showed their increased incorporation into the LAF-1 RGG condensate (Supplementary Figure S4A) with a corresponding decrease in the heterotypic free energy minima (Supplementary Figure S3A). On performing the same study with the LAF-1 RGG$_{Cshuf}$ condensate we observed an enhanced presence of polynucleotides with increasing length on the periphery or 'shell' of the LAF-1 RGG$_{Cshuf}$ rich core (Supplementary Figure S5A) while the displaced free energy minimum decreases further (Supplementary Figure S3B) without a significant change in its location.

This study into the microscopic peculiarities of intermolecular interactions between the entities present in a multicomponent mixture provided us with significant insight towards explaining the phase separation behavior of the mixture highlighting the importance of studying interactions at the molecular level to understand how these molecules phase separate en masse. Inferences gained from estimating two body interaction strengths and interaction free energy profiles can be used to compare and rank LLPS-promoting biomolecules in the mixture and also possibly allow one to predict or explain the formation of microscopic

organization or sub-compartmentalization previously observed in condensates or MLOs (73). Another aspect which one must consider is that although one can use two body interaction PMFs to gain a better understanding of how two molecules would prefer to interact with each other, one should be careful with the application of this information as these two body interaction PMFs which are calculated at infinite dilution might not reflect the thermodynamics of molecules in the crowded condensed phase.

CONCLUSION

In this work, we have developed a novel coarse-grained (CG) simulation model for capturing protein–polynucleotide interactions and their phase separation. We show how this simple polynucleotide model is sufficient to reveal insights into the role polynucleotides like RNA play 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 by itself, and in the presence of disordered RNA molecules (23). Our coexistence simulations allow us to quantify the degree to which polynucleotides are incorporated into the condensed protein phase and also show effects of polynucleotides 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 protein-polynucleotide phase separation. While keeping the overall amino acid composition constant we shuffled the LAF-1 RGG 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-polynucleotide interactions are also enhanced by this charge segregation, we do not observe a marked increase in polynucleotide incorporation into the condensed phase. Instead, we see interfacial peaks indicating polynucleotides accumulating around the charge patterned protein condensate forming a layer between the molecule rich and molecule depleted vapor region in our simulations.

To explain the occurrence of this phenomenon, we then looked at two-body interactions by calculating the PMFs between the different molecular pairs using which we propose that the interplay and peculiarity of homotypic and heterotypic interactions controls phase separation of the protein–polynucleotide mixture leading to the formation of microstructure in condensates. The different architectures of the condensates are explained by the difference in heterotypic LAF-1 RGG_{Cshuf}–polynucleotide and LAF-1 RGG–polynucleotide interaction energy profiles which suggests that the microstructure in the condensate stems from the presence of a spaced configuration of LAF-1 RGG_{Cshuf}–polynucleotide chains which is more stable than the configuration where the chains' centers of mass are zero distance apart which is the favorable configuration for the LAF-1 RGG–polynucleotide pair. Hence, our work here suggests the possibility of using charge patterning of the protein sequence to not only modulate the phase separation of the protein-polynucleotide mixture but also introduce nonhomogeneous microstructures within the multicomponent condensate and also suggest how one can obtain a better grasp of the driving forces for this behavior by closely studying interactions present in the mixture at the molecular level.

Even though our simplistic unstructured CG representation of polynucleotides provides us with results in significant agreement with experimental data on RNA-protein complexes, 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 CG representation is warranted as we would like to capture the more nuanced RNA interactions arising from the π character and secondary structure of RNA molecules which would also play important roles in modulating RNA-protein 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 polynucleotides like RNA interact with proteins.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Institute of General Medical Sciences of the National Institutes of Health [R01GM136917]; NINDS; NIA [R01NS116176], NSF [2004796]; NSF MRSEC Grant [DMR 1720530] (in part); 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 [TG-MCB-120014]. Funding for open access charge: NIH [R01GM136917].

Conflict of interest statement. None declared.

REFERENCES

1. Banani,S.F., Rice,A.M., Peeples,W.B., Lin,Y., Jain,S., Parker,R. and Rosen,M.K. (2016) Compositional Control of Phase-Separated Cellular Bodies. *Cell*, **166**,3651–663.
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. Lee,CY. and Seydoux,G. (2019) Dynamics of mRNA entry into stress granules. *Nat. Cell. Biol.*, **21**, 116–117.
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**, 7189–7194.
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.*, **44**, 18–30.
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. (2017) Reentrant phase transition drives dynamic substructure formation in ribonucleoprotein droplets. *Angew. Chem. Int. Ed. Engl.*, **56**, 11354–11359.
20. Murthy,A.C., Dignon,G.L., Kan,Y., Zerze,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*, **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.
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**, 1118–1125.
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**, e42571.
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. and Pomès,R.2017) The liquid structure of elastin aggregates. *Elife*.**6**, e26526.
28. Conicella,A.E., Dignon,G.L., Zerze,G.H., Schmidt,H.B., D'Ordine,A.M., Kim,Y.C., Rohatgi,R., Ayala,Y.M., Mittal,J. and Fawzi,N.L. (2020) TDP-43 α-helical structure tunes liquid–liquid phase separation and function. *Proc. Natl. Acad. Sci. USA*, **117**, 5883–5894.
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. U.S.A.*, **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**, doi:10.1088/1367-2630/aa9369.
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. U.S.A.*, **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. bioRxiv doi: https://doi.org/10.1101/2020.06.12.148650, 12 June 2020, preprint: not peer reviewed.
37. Alshareedah,I., Moosa,M.M., Raju,M., Potoyan,D.A. and Banerjee,P.R. (2020) Phase transition of RNA–protein complexes into ordered hollow condensates. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 15650–15658.
38. 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.
39. Kaur,T., Raju,M., Alshareedah,I., Davis,R.B., Potoyan,D.A. and Banerjee,P.R. (2020) Sequence-encoded and composition-dependent protein-RNA interactions control multiphasic condensate topologies. bioRxiv doi: https://doi.org/10.1101/2020.08.30.273748, 31 August 2020, preprint: not peer reviewed.
40. Smith,L.G., Zhao,J., Mathews,D.H. and Turner,D.H. (2017) Physics-based all-atom modeling of RNA energetics and structure. *WIREs RNA*, **8: e1422**.
41. Zhang,Y., Yang,M., Duncan,S., Yang,X., Abdelhamid,M.A.S., Huang,L., Zhang,H., Benfey,P.N., Waller,Z.A.E. and Ding,Y. (2019) G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Res.*, **47**, 11746–11754.
42. Fay,M.M., Anderson,P.J. and Ivanov,P. (2017) ALS/FTD-Associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell Rep.*, **21**, 3573–3584.
43. Boeynaems,S., Holehouse,A.S., Weinhardt,V., Kovacs,D., Van Lindt,J., Larabell,C., Van Den Bosch,L., Das,R., Tompa,P.S., Pappu,R. V. *et al.* (2019) Spontaneous driving forces give rise to protein–RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 7889–7898.
44. Langdon,E.M., Qiu,Y., Niaki,A.G., Mclaughlin,G.A., Weidmann,C.A., Gerbich,T.M., Smith,J.A., Crutchley,J.M., Termini,C.M., Weeks,K.M. *et al.* (2018) mRNA structure determines specificity of a polyQ-driven phase separation. *Science*, **1**, 922–927.
45. 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. U.S.A.*, **117**, 202000223.
46. Creighton,T.E. (1993) Proteins: structures and molecular properties Macmillan.

47. 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.
48. Ashbaugh,H.S. and Hatch,H.W. (2008) Natively unfolded protein stability as a coil-to-globule transition in charge/hydropathy space. *J. Am. Chem. Soc.*, **130**, 9536–9542.
49. Šulc,P., Romano,F., Ouldridge,T.E., Doye,J.P.K. and Louis,A.A. (2014) A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.*, **140**, 235102.
50. Denesyuk,N.A. and Thirumalai,D. (2013) Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B*, **117**, 4901–4911.
51. Pasquali,S. and Derreumaux,P. (2010) HiRE-RNA: a high resolution coarse-grained energy model for RNA. *J. Phys. Chem. B*, **114**, 11957–11966.
52. 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.
53. 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.
54. 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.
55. Panagiotopoulos,A.Z. (2000) Monte Carlo methods for phase equilibria of fluids. *J. Phys. Condens. Matter*, **12**, R25.
56. Kim,J., Keyes,T. and Straub,J.E. (2010) Generalized replica exchange method. *J. Chem. Phys.*, **132**, 224107.
57. 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.
58. 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.
59. 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.
60. Regy,R.M., Zheng,W. and Mittal,J. (2020) Using a sequence-specific coarse-grained model for studying protein liquid–liquid phase separation. In: Keating,C.D. (ed). *Liquid-Liquid Phase Coexistence and Membraneless Organelles, Methods in Enzymology*. Academic Press, Vol. **646**.
61. Dill,K.A. and Bromberg,S. (2010) Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience.Garland science.
62. 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.
63. Vernon,R.M., Chong,P.A., Tsang,B., Kim,T.H., Bah,A., Farber,P., Lin,H. and Forman-Kay,J.D. (2018) Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife*, **7**, e31486.
64. 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.
65. 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.
66. Sawle,L. and Ghosh,K. (2015) A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.*, **143**, 085101.
67. 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.
68. Putnam,A., Cassani,M., Smith,J. and Seydoux,G. (2019) A gel phase promotes condensation of liquid P granules in Caenorhabditis elegans embryos. *Nat. Struct. Mol. Biol.*, **26**, 220–226.
69. Fei,J., Jadaliha,M., Harmon,T.S., Li,I.T.S., Hua,B., Hao,Q., Holehouse,A.S., Reyer,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.
70. Madinya,J.J., Chang,L., Perry,S.L. and Sing,C.E. (2020) Sequence-dependent self-coacervation in high charge-density polyampholytes. *Mol. Syst. Des. Eng.*, **5**, 632–644.
71. Borgia,A., Zheng,W., Buholzer,K., Borgia,M.B., Schu,A., Hofmann,H., Soranno,A., Nettels,D., Gast,K., Grishaev,A. *et al.* (2016) Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J. Am. Chem. Soc.*, **138**, 11714–11726.
72. Zheng,W., Borgia,A., Buholzer,K., Grishaev,A., Schuler,B. and Best,R.B. (2016) Probing the action of chemical denaturant on an intrinsically disordered protein by simulation and experiment. *J. Am. Chem. Soc.*, **138**, 11702–11713.
73. 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**, 012411.