Charge-Driven Phase Separation of RNA and Proteins without Disorder

Bercem Dutagaci[1], Grzegorz Nawrocki[1], Joyce Goodluck[2], Lisa J. Lapidus[2*], and Michael Feig[1*]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, USA

[2]Department of Physics, Michigan State University, East Lansing, MI, USA

* Co-corresponding authors:
Lisa Lapidus
567 Wilson Rd., Room 4227 BPS
East Lansing, MI 48824, USA
lapidus@msu.edu
517-884-5656

Michael Feig
603 Wilson Rd., Room 218 BCH
East Lansing, MI 48824, USA
mfeiglab@gmail.com
517-432-7439

RUNNING TITLE

Phase Separation in RNA-Protein Mixtures

ABSTRACT

Phase separation of globular RNA and positively charged proteins is reported from a combination of coarse-grained simulations parametrized based on atomistic simulations, theory informed by the coarse-grained simulations, and experimental validation via confocal microscopy and FRET spectroscopy. Phase separation is found to depend on concentration, size, and charge of the proteins, requiring a minimum protein size, minimum protein charge, and minimum protein concentration before condensates can form. The general principle for phase separation is based on electrostatic complementarity rather than invoking polymer character as in most previous studies. Simulation results furthermore suggest that such phase separation may occur in heterogenous cellular environment, not just between tRNA and cellular proteins but also between ribosomes and proteins where there may be competition for positively charged proteins.

STATEMENT OF SIGNIFICANCE

Liquid-liquid phase separation has been recognized as a key mechanism for forming membrane-less organelles in cells. Commonly discussed mechanisms invoke a role of disordered peptides and specific multi-valent interactions. We report here phase separation of RNA and proteins based on a more universal principle of charge complementarity that does not require disorder or specific interactions. The findings are supported by coarse-grained simulations, theory, and experimental validation via microscopy and FRET spectroscopy. The implications of this work are that condensate formation may be an even more universal phenomenon in biological systems than thought to date.

KEYWORDS

Liquid-liquid phase separation; coarse-grained simulation; confocal microscopy; electrostatics; FRET spectroscopy; biomolecular condensates; RNA; proteins

INTRODUCTION

Biological cells compartmentalize to support specific functions such as stress response (1, 2), regulation of gene expression (3, 4) and signal transduction (5). Compartmentalization by organelles that are surrounded by lipid membranes is well known. In addition, membrane-less organelles that result from biomolecular condensation have been described (6-9). In the nucleus they include the nucleolus (10, 11), nuclear speckles (12, 13), and cajal bodies (14-16); stress granules (2, 17, 18), germ granules (19, 20), and processing bodies (21, 22) have been found in the cytoplasm. The formation of such condensates depends on the composition and concentration of biomolecules (9) as well as environmental conditions such as pH, temperature, and the concentration of ions (6, 23, 24). Multivalent interactions, the presence of intrinsically disordered peptides (IDPs) (25), and electrostatic interactions between highly charged molecules such as RNA (25, 26) are generally considered the key factors that promote phase separation (PS) (8, 10, 18, 27-33). Experiments such as fluorescence recovery after photobleaching (FRAP) suggest that these condensates retain liquid-like behavior (34, 35) and therefore can fully support biological function. However, diffusion inside condensates (10, 18, 36) may be retarded and transitions from the liquid to gel or solid phases including aggregation to fibrils may occur (37-39).

Biomolecular condensates have been studied extensively (40). Microscopy (23, 41, 42), nuclear magnetic resonance (NMR) spectroscopy (17, 43), fluorescence spectroscopy (10, 44, 45), X-ray diffraction (46, 47) , and scattering methods (44, 48, 49) have characterized *in vitro* (10, 17, 23, 42, 43, 45-47, 49) and *in vivo* systems (19, 50, 51). Theoretical studies have complemented experiments (40, 52), including particle-based simulations (53) and analytical approaches based on polymer (54, 55) and colloid theories (56-58). Additional insights into specific interactions have come from molecular dynamics (MD) simulation studies (45, 59, 60). Polymer aspects of IDPs and unstructured RNA were emphasized in applications of Flory-Huggins theory in combination with simulations (10, 53, 61, 62). Related studies in the colloid field have described the phase behavior of macromolecules and nanoparticles as single spherical particles (56-58). However, most of the latter studies so far have focused on liquid-solid transitions and the formation of finite size clusters in monodisperse systems. Despite progress, it has remained unclear what components can lead to condensation, especially in highly heterogeneous cellular environments.

In this study, we focus specifically on the question of whether disorder, in the form of IDPs or unstructured RNA is essential for PS and what factors may determine the propensity for PS in a heterogeneous system. The starting point is a molecular model of a bacterial cytoplasm that was established by us previously (63, 64) and that was simulated here using colloid-like spherical particles with a potential parameterized against atomistic MD simulations of concentrated protein solutions. Distinct phases enriched with highly negatively charged RNA and positively charged proteins were formed in the simulations, suggesting a generic electrostatic mechanism that does not require polymer-like components. The phase behavior seen in the cytoplasmic

system was reproduced in reduced five- and two-component models and described by an analytical model where we could systematically vary molecular charge, size, and concentrations. The main prediction from the simulations and theory is that mixtures of globular RNA and globular proteins of opposite charge can form phase separated states dependent on concentration, charge, and size of the proteins. The predictions were confirmed experimentally by imaging condensates between globular RNA and a series of globular proteins via confocal microscopy.

RESULTS

Liquid condensates enriched in tRNA and ribosomes form in a model bacterial cytoplasm

A model of the cytoplasm of *Mycoplasma genitalium* established previously (63, 64) was simulated at a coarse-grained (CG) level with one sphere per macromolecule or complex (SI Methods and SI Sheet 1). CG particle interactions were calibrated against results from atomistic MD simulations of concentrated protein solutions. The parameters involve only two particle-dependent properties, namely size and charge (SI Methods). Droplet-like condensates formed spontaneously within 20 μs (Fig. 1A/B) and remained present during 1 ms simulation time. Two types of condensates were observed: one type contained predominantly tRNA and positively charged proteins; the other type contained ribosomes (RP) and positively charged proteins. The RP condensates also attracted the weakly negatively charged GroEL particles at the surface (Fig. 1A). The condensates increased in size as the system size was increased from 100 to 300 nm (Fig. 1A). This is consistent with PS vs. finite-size cluster formation. The presence of multiple droplets in the 300 nm system suggests incomplete convergence, but as the droplets grow in size, further merging becomes kinetically limited due to slowing diffusion. Further analysis focused on the condensates observed in the 100-nm system.

Cluster analysis (SI Methods) considered interactions between the nucleic acids and positively charged proteins to obtain trajectory-averaged cluster size distributions (Fig. 1C). Most tRNA (87%) was part of a condensate. The remaining fraction of tRNA existed as monomers or small clusters, suggesting coexistence of dilute and condensed phases. RP particles were only found in the RP condensates and therefore underwent a complete phase transition. Total macromolecular volume fractions inside tRNA and RP condensates were 0.42 and 0.58, respectively, whereas volume fractions for just tRNA and RP inside their respective condensates were 0.07 and 0.26. The condensates were significantly denser than the rest of the simulated system (Fig. S1). The moderately high volume fractions for tRNA condensates are still within the range of concentrated liquid phases (65), but the higher volume fractions in the RP condensate tend towards solid- or gel-like phases (65). Radial distribution functions of tRNA and RP from the center of the corresponding condensates show a relatively smooth decay with a soft boundary for tRNA condensates (Fig. S2), also consistent with a liquid-like phase, whereas distinct peaks and a sharper boundary for RP indicate a more ordered phase in the RP condensates.

Both, tRNA and RP condensates contained (positively charged) proteins at high concentrations. tRNA and RP interactions with those proteins were favorable as evidenced by a strong peak in the pairwise radial distribution function $g(r)$ at contact distance (Fig. S3). The charge and size of the proteins attracted to the condensates differed between tRNA and RP condensates (Fig. S4). In the tRNA condensates, large proteins with radii of 3 nm and above and with charges of 10 and above were preferred. In contrast, the proteins in the RP condensates were smaller, with radii of 3 nm or less, and many proteins had charges below 10.

The dynamics inside and outside the condensates was analyzed further in terms of translational diffusion coefficients (D_{tr}) calculated based on mean-squared displacements (Fig. S5). Diffusion during the last 1 μs of the simulation was compared with diffusion during the first 1 μs when condensates were not yet formed. Molecule-specific values of D_{tr} are given in SI Sheet 1. As a function of the radius of the macromolecules (Fig. S6), D_{tr} values follow a similar trend as observed before in atomistic simulations of the same system. Diffusion outside the condensates is like diffusion in the dispersed phase. In tRNA condensates, the diffusion of macromolecules is similar to the disperse phase or moderately retarded, depending on the molecule, and consistent with reduced diffusion in increased protein concentrations seen in experiment (66, 67). In RP condensates, diffusion is reduced more, but liquid-like dynamics is still maintained for most types of macromolecules as they diffuse around a relatively static RP cluster (SI Movie 1).

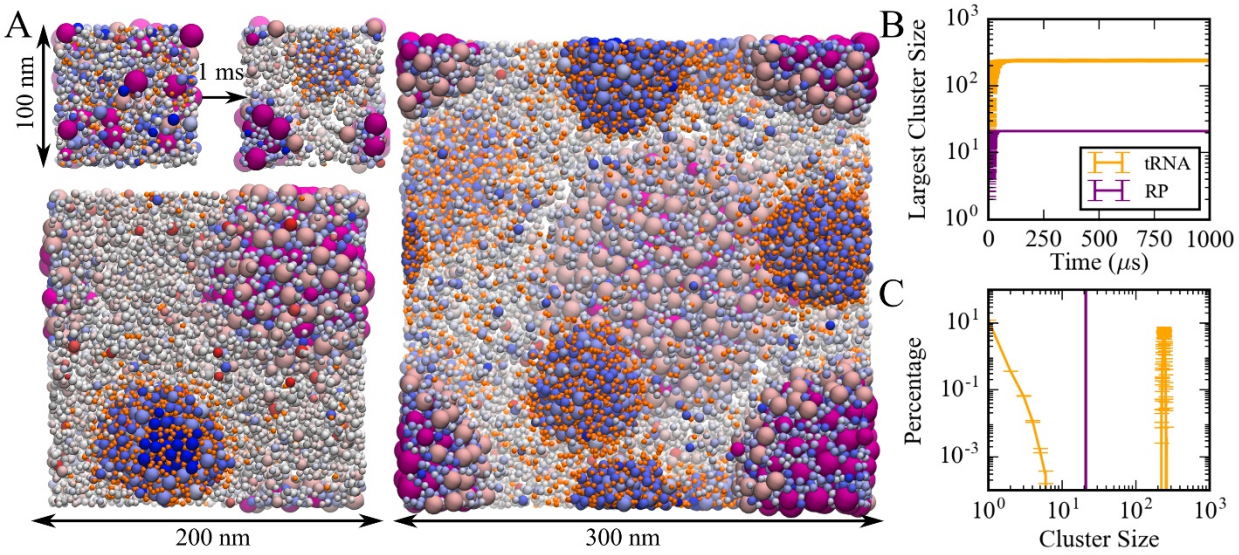

Figure 1. (A) Coarse-grained simulations of a model bacterial cytoplasm. Initial and final frames for 100 nm box and final frames for 200 and 300 nm boxes are shown with tRNAs in orange, ribosomes in magenta, and other molecules colored according to their charges (blue towards positive charges; red towards negative charges). Sphere sizes are shown proportional to molecular sizes. Large pink spheres correspond to GroEL particles. (B) Size of the largest cluster vs. simulation time in 100 nm system. (C) Cluster size distributions for tRNA and RP during the last 500 μs in 100 nm system.

Factors promoting RNA condensation in a reduced five-component model system

A simplified five-component system was constructed to reproduce the RNA condensation observed in the cytoplasmic model. The simplified model consisted of tRNA, ribosomes (RP), large (POS_L, q = 20, r = 3.5 nm) and small (POS_S, q = 1, r = 2.52 nm) positively charged proteins as well as neutral crowders (CRW, q = 0, r = 2.52 nm). tRNA and RP concentrations were initially set as in the cytoplasmic model while concentrations and parameters of the other three particle types were adjusted to match the total number of particles, total molecular volume, and total charge of the cytoplasmic system. Subsequently, series of simulations were run at different concentrations and with different parameters (SI Sheet 2).

In simulations of the five-component model, tRNA and RP condensed separately as in the cytoplasmic model (Fig. S7). Again, the condensates formed quickly, within 50 μs (Fig. S7), and cluster size distributions of tRNA and RP resembled the results from the cytoplasmic system (*cf.* Figs. 1 and S7). However, different from the cytoplasmic system, we found a small fraction (2% on average) of RP particles in the dilute phase. As in the cytoplasmic model, tRNA strongly preferred interactions with the larger POS_L particles, whereas RP interacted favorably with both POS_S and POS_L (Fig. S8). tRNA condensates remained fully liquid as in the cytoplasmic system. From the last 100 μs of the simulation, we obtained diffusion coefficients D_{tr} for tRNA of 28.3 ± 0.7 and 59.0 ± 0.5 $nm^2/\mu s$ inside and outside of the condensates, respectively, similar to values of 16.3 ± 0.1 and 55.5 ± 0.8 $nm^2/\mu s$ in the cytoplasmic system. Diffusion coefficients for RP inside and outside of the RP condensates were 0.49 ± 0.01 and 0.80 ± 0.4 $nm^2/\mu s$, respectively, compared to D_{tr} = 0.34 ± 0.01 $nm^2/\mu s$ for RP in the cytoplasmic condensates.

RP and POS_L concentrations were varied systematically, while the concentration of POS_S was kept constant and the number of CRW particles was adjusted to maintain a constant total molecular volume (SI Sheet 2). Cluster size distributions were extracted (Fig. S9) and the fraction of tRNA and RP in the large clusters was determined (Fig. 2). Some degree of clustering occurs at all concentrations, but condensation requires that a significant fraction of particles is found in the largest clusters. Based on a criterion that at least half of the particles are found in one or few large clusters, tRNA and RP condensation occurs for $[POS_L] > 100$ μM (Fig. 2).

Increasing [RP] reduces the amount of tRNA in the tRNA condensates and effectively raises the critical POS_L concentration above which tRNA forms condensates (Fig. 2). This can be understood from competition for POS_L. tRNA only interacts significantly with POS_L (Fig. S10) and needs POS_L to form condensates, whereas RP interacts with both POS_S and POS_L (Fig. S11) and therefore draws POS_L from tRNA condensates (Fig. S12). For $[POS_L] > 500$ μM, the fraction of tRNA particles in the tRNA condensates is relatively constant (Fig. 2). However, the number of POS_L particles in the condensates increases as the total $[POS_L]$ increases (Fig. S12A). This results in larger clusters and lower effective [tRNA] in the condensates at the highest values of $[POS_L]$ (Fig. S13). The effect of increasing [RP] is again a depletion of POS_L in the tRNA

condensates, so that [tRNA] in the condensates increases with [RP] for a given value of [POS_L] (Fig. S13).

In the simulations described so far, the total volume fraction of the system was kept constant by reducing the crowder (CRW) concentration as [POS_L] and [RP] increased. Therefore, an alternative explanation of the decrease in [tRNA] inside the condensates with increasing [POS_L] could be reduced crowding by the condensate environment. To test this further, we reduced [CRW] without changing [POS_L]. Reduced [CRW] also led to reduced [tRNA] in the condensate, but the effect is much smaller than when [CRW] is reduced along with an increase in [POS_L] (Fig. S14).

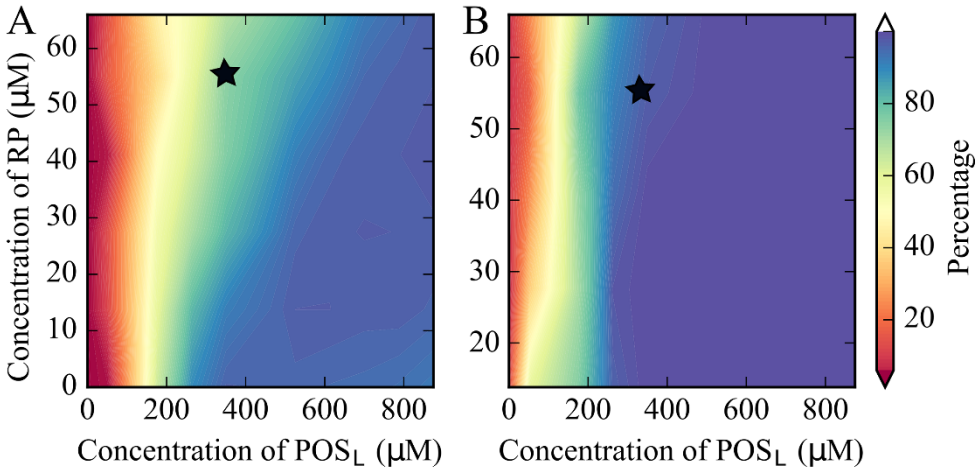

Figure 2. Percentage of tRNA (A) and RP (B) particles in largest clusters in coarse-grained simulations of the five-component model system as a function of [RP] and [POS_L]. The black star indicates the conditions that match the cytoplasmic model.

tRNA condensation is a phase separation process.

In order to construct phase diagrams, simulations of the five-component model phases were carried out at a range of temperatures for selected values of [RP] and [POS_L]. Cluster size distributions were extracted (Figs. S15-S17) and the volume fractions of tRNA in dilute and condensed phases as a function of temperature were determined based on the number of tRNA outside and inside the largest tRNA clusters. The resulting curves (Fig. 3) show the typical features of phase diagrams with phase coexistence below critical temperatures T_c of 400 to 535 K. An increase in [POS_L] raises T_c (Fig. 3C), consistent with POS_L promoting tRNA condensation. When [POS_L] = 180 μM, near the minimum needed for PS, an increase in [RP] slightly decreased T_c (Fig. 3A/D). However, at a higher concentration, i.e. [POS_L] = 880 μM, T_c increased with increasing [RP] up until 55 μM. This suggests that ribosomes also affect PS of tRNA although the effect appears to be complicated, presumably again because of the competition for POS_L.

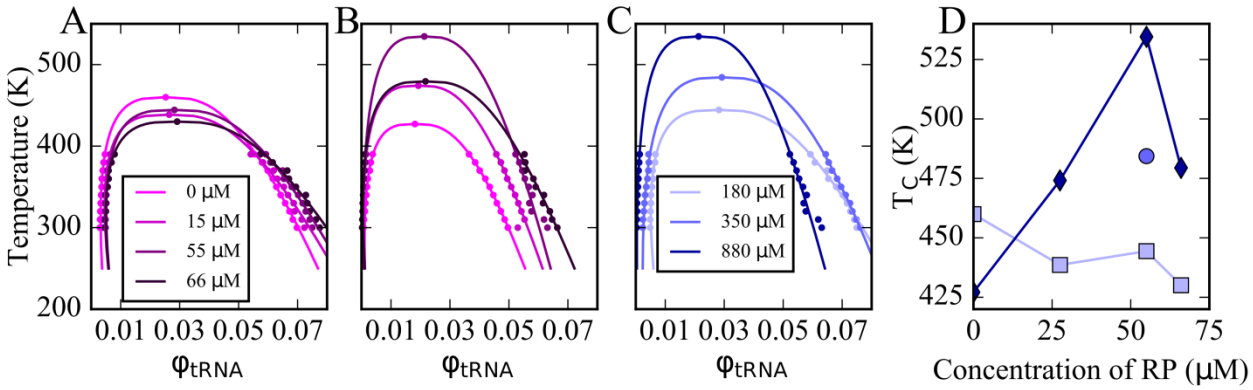

Figure 3. Phase diagrams for tRNA with [POS_L] = 180 μM and varying RP concentrations (A); with [POS_L] = 880 μM and varying RP concentrations (B); and with [RP] = 55 μM and varying POS_L concentrations (C); critical temperatures as a function of [RP] at [POS_L] = 180 μM (squares), at [POS_L] = 880 μM (diamonds), and at [POS_L] = 350 μM (sphere) (D). Lines in A-C were fitted according to Eqs. M-5 and M-6 (SI Methods).

Prediction of phase separation in binary mixtures of globular RNA and proteins

The possibility of PS in mixtures of globular RNA and positively charged proteins was studied further via CG simulations and a theoretical model that aims to capture the behavior seen in the simulations and allow further predictions. We focus here specifically on the 47-nucleotide J345 Varkud satellite ribozyme RNA, that folds into an approximately globular shape(68), in binary mixtures with common proteins with positive charges and varying sizes for which we may expect PS: myoglobin (q = +2, r = 1.64 nm), trypsin (q = +6, r = 1.81 nm), lysozyme (q = +8, r = 1.54 nm), cytochrome C (q = +11, r = 1.45 nm), lactate dehydrogenase (LDH; q = +4, r = 2.68 nm), and alcohol dehydrogenase (ADH; q = +8, r = 2.79 nm)). We also compared with bovine serum albumin (BSA; q = -17, r = 2.58 nm), for which condensate formation is not expected due to its negative charge. The RNA and proteins were selected because they allowed experimental validation as described below.

In CG simulations, we observed the formation of condensates at sufficiently high salt concentrations. With κ = 0.7 (about 20 mM salt, see SI Methods), condensates formed with lysozyme, trypsin, LDH, and ADH, but not with cytochrome C, myoglobin, or BSA (Fig. 4). This further suggests that both size and charge are determinants of RNA PS as in the cytoplasmic models. At κ = 0.5, all proteins form condensates since electrostatic interactions are screened so much that attractive close-range interactions dominate and overcome even repulsion with BSA. [RNA] and [protein] in the condensates with the larger LDH and ADH proteins are significantly lower than in the condensates with the smaller lysozyme and trypsin proteins, as expected due to packing considerations. At the same time, the concentration of RNA in the dilute phase is lower with LDH and ADH (Fig. 4), indicating a greater propensity to form condensates with the larger proteins. Interestingly, there are only very few proteins in the dilute phase suggesting phase coexistence for RNA but a complete phase transition for the proteins.

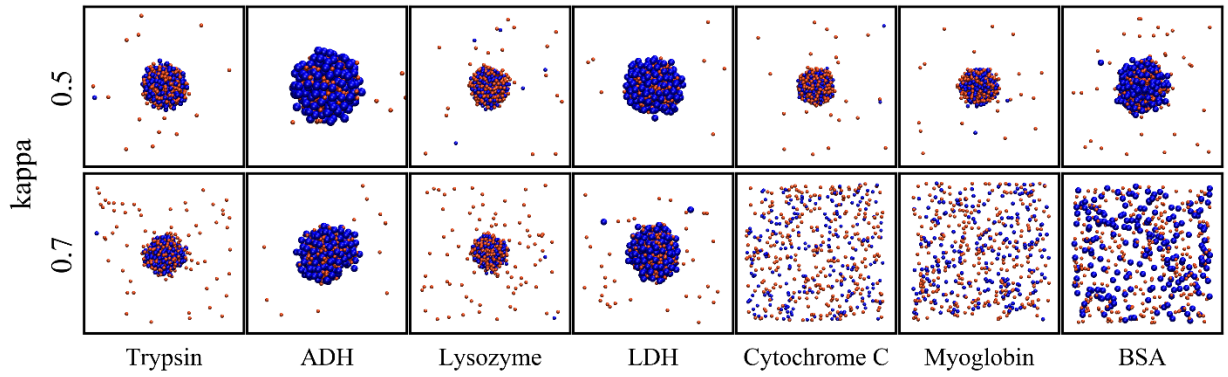

Figure 4. Snapshots after 1 ms for binary RNA-protein mixtures at T=298K, [RNA] = 0.493 mM, [protein] = 0.350 mM. Orange and blue spheres show RNA and proteins, according to size. Concentrations inside the condensates at κ = 0.7 were [RNA:lysozyme] = 20.2:20.2 mM; [RNA:trypsin] = 16.5:15.2 mM; [RNA:LDH] = 9.6:7.2 mM; [RNA:ADH] = 9.5:6.7 mM.

A theoretical model was developed based on the CG simulations to further understand the determinants of condensation. Briefly, the model approximates the chemical potential for either RNA or proteins in condensed and dilute phases based on a decomposition into enthalpy and entropy: $\mu = \Delta h - T\Delta s$. The enthalpy is determined from convoluting the coarse-grained interaction potential *U(r)* (Eq. 3) with radial distribution functions $\hat{g}(r)$ of RNA-RNA, RNA-protein, and protein-protein interactions in the condensed and dilute phases extracted from CG simulations and scaled by particle densities ρ (see SI Theory):

$$\Delta h = 2\pi\rho \int \hat{g}(r)U(r)r^2 dr \quad (1)$$

The entropy was estimated from the ratio of particle densities ρ between the entire system and either the dilute or condensed phase:

$$\Delta s = R\log\left(\frac{\rho_{total}}{\rho_{phase}}\right) \quad (2)$$

Solutions with respect to the concentrations of protein and RNA in dilute and condensed phases were determined numerically under the conditions that $\mu_{condensed} = \mu_{dilute}$ for either RNA, protein, or both, and that molecular volume packing fractions did not exceed maximum liquid packing densities (SI Theory). Total free energies were then calculated, taking also into account mixing entropy contributions between RNA and protein particles. PS was predicted based on the solution with the lowest free energy.

The theory predicts that PS should occur for a wide range of protein radii and charges if proteins are large enough and carry sufficiently positive charge (Fig. 5). With κ = 1.3, the model predicts PS with lysozyme, trypsin, LDH, and ADH, but not with myoglobin or BSA, as in the CG simulations. The theory also predicts PS for cytochrome C, for which PS was not seen in the simulations. Unexpectedly, we find additional solutions for very small radii and high charges. At the molecular level, such solutions could correspond to small clusters of basic amino acid residues either in the form of small peptides or perhaps as part of IDPs where clusters of charged amino acids may be connected by flexible segments. However, we emphasize that our theory does not consider any polymer characteristics and is therefore necessarily incomplete with respect to describing IDPs.

The model reflects an expected salt and temperature dependence of PS with protein-dependent critical maximal temperatures and critical minimal salt concentrations (Figs. S18 and S19). However, the critical salt concentrations are lower (larger values of κ) than in the CG simulations. This may be expected since the theory assumes essentially infinite-size macroscopic systems. Surface-tension effects between phases are not considered but they are expected to play a significant destabilizing role during finite-size droplet condensate formation. This may also explain why cytochrome C did not lead to condensation in the simulations at κ = 0.7, since such condensates would be smallest with the highest surface tension compared to the other proteins.

According to the theory, PS also depends on the total protein and RNA concentrations (Figs. S20-S25). A minimal protein concentration around 50-100 μM is predicted for lysozyme (Figs. S22), trypsin (Figs. S23), and cytochrome C (Figs. S24), largely independent of the RNA concentration. For LDH and ADH, no significant minimum protein concentration is predicted. On the other hand, for myoglobin, the model predicts that PS may occur for high protein concentrations (above 0.5 mM; Fig. S25).

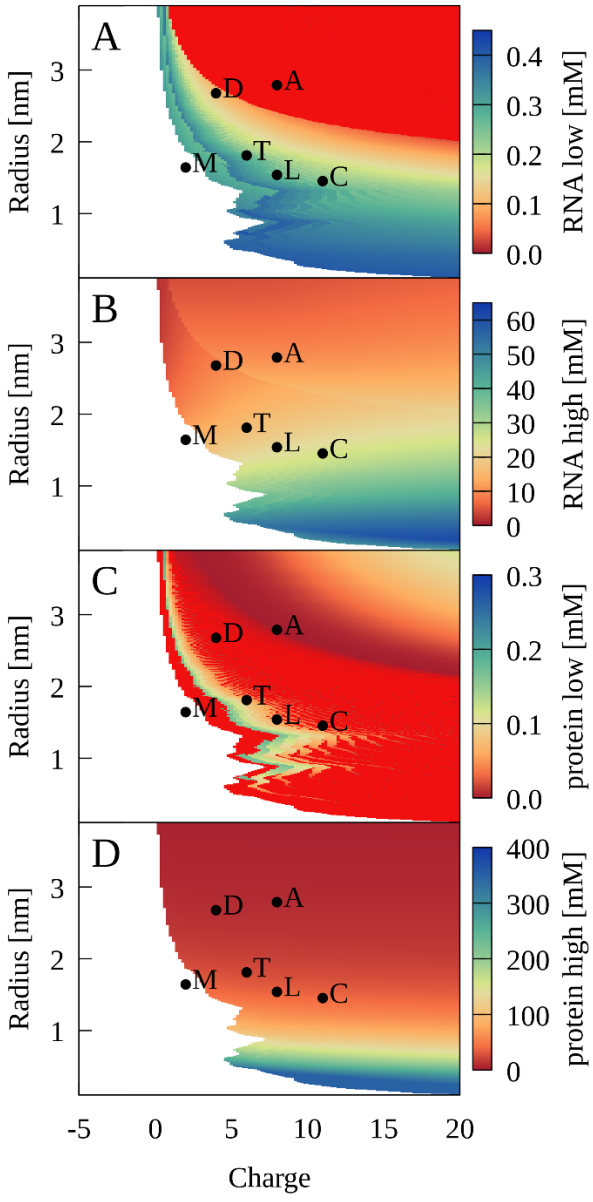

Figure 5. Phase separation for binary RNA-protein mixtures as a function of protein charge and radius from theory. Colors show [RNA] (A, B) and [protein] (C, D) in dilute (A, C) and condensed (B, D) phases. Red indicates zero concentration. [RNA] = 0.45 mM, [protein] = 0.35 mM, κ = 1.3, and T = 298 K. Proteins are projected as follows: myoglobin (M); trypsin (T); lysozyme (L); cytochrome C (C); LDH (D); ADH (A).

Phase separation of RNA and proteins in experiment

The central prediction from the simulations and theory presented so far is that mixtures of globular RNA and globular proteins can phase separate into high-density liquid condensates that coexist with dilute phases. To test this prediction, experiments were carried out for mixtures between dye-labeled J345 RNA, known to form compact, globular-like structures, and globular proteins for which PS was predicted. In addition, myoglobin and BSA were also included as controls where the theory did not predict PS. Imaging via confocal microscopy (Fig. 6 and Figs. S26 to S31) shows well defined fluorescent clusters for mixtures of RNA and trypsin, ADH, lysozyme, LDH, but not for RNA and myoglobin and BSA. Generally, the condensates are small and near or below the diffraction limit of the microscope. Trypsin/RNA mixtures exhibit some larger clusters and the sizes follow roughly an exponential distribution (Fig. S32). The background fluorescence varies significantly with protein, suggesting, for example, that only a fraction of RNA is participating in LDH condensates.

One key prediction from the theory is that condensates form only above a certain minimum protein concentration. This was confirmed for trypsin where condensates were not observed at 50 μM but appeared at 150 μM protein concentration (Fig. S33). Förster resonance energy transfer (FRET) experiments also showed a significant increase in FRET efficiencies from 50 μM to 150 μM consistent with condensation (Fig. 7A). The gradual increase in FRET efficiencies upon increase of trypsin concentrations is interpreted to result initially from increasing non-condensate cluster formation (see cluster size distributions in Fig. S9 at [RP] = 0 with increasing protein concentration), whereas a continued gradual increase after condensates are formed is expected from theory based on an increasing fraction of RNA molecules in the condensates (Fig. 7B, Fig. S34, and SI Theory). The comparison between the microscopy and FRET results furthermore establishes that RNA condensates at this RNA concentration can be recognized by FRET efficiencies above 0.26, whereas lower values may indicate a disperse phase.

Finally, we confirmed via circular dichroism spectroscopy that trypsin and RNA remain in their folded states upon condensate formation. The spectrum for a mixture of RNA and trypsin that forms condensates agrees qualitatively with individual spectra for RNA and trypsin (Fig. S35).

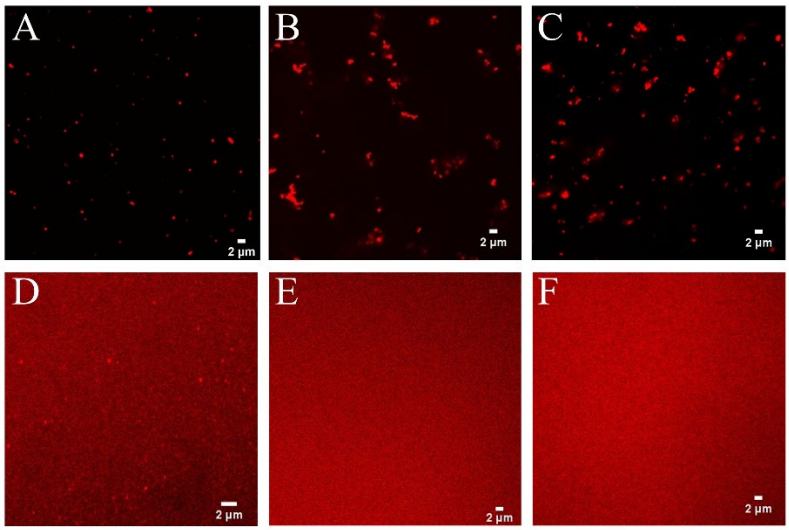

Figure 6. Phase separation in mixtures of J345 RNA at 0.45 mM and various globular proteins at 0.35 mM from confocal microscopy of labeled RNA: trypsin (A), ADH (B), lysozyme (C), LDH (D), myoglobin (E), BSA (F).

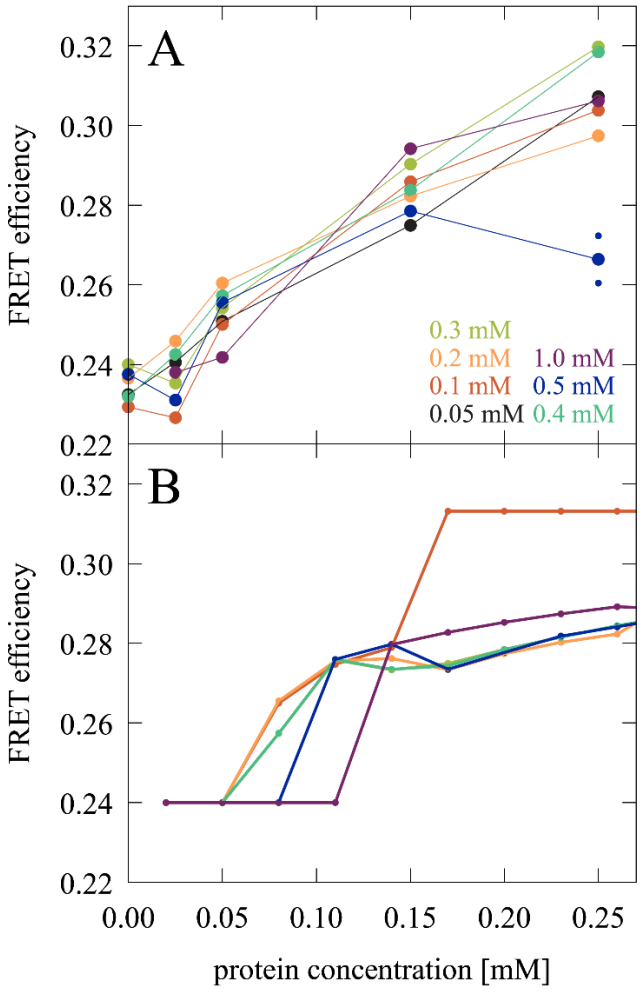

Figure 7. FRET efficiency in mixtures of J345 RNA with trypsin as a function of protein concentration at different RNA concentrations (as indicated by color). The average of two measurements is shown for 0.25 mM protein and 0.5 mM RNA concentrations with smaller points indicating individual measurements. (A). FRET efficiency estimated from the fraction of RNA in the condensed phase from theory (SI Theory) (B).

DISCUSSION

This study presents a new view on charge-driven biomolecular PS supported by simulation, theory, and experiments. More specifically, we report a potential for PS between negatively charged RNA and positively charged proteins without requiring disorder or polymer-character of either component. Our simulations and the theoretical model are based on isotropic spheres, whereas experimental validation is based on a compact, approximately globular RNA and a variety of globular proteins not known to specifically interact with RNA. This suggests that such PS could be a very general phenomenon in biological cells depending on the concentrations, charge, and size distribution of available nucleic acid and protein components. In fact, our simulations of a bacterial cytoplasm provide examples of separately forming tRNA-protein and ribosome-protein condensates in a cytoplasmic environment.

The idea of strong complementary interactions playing a major role in PS is well-established (69). While almost all of the LLPS studies to-date involve polymers and in particular IDPs (70), there are a few other examples in the literature that discuss PS involving folded proteins (41, 48, 71, 72). In those cases, the ability to form condensates was generally ascribed to specific multi-valent interactions. While those examples could be viewed as more akin to specific complex or amorphous aggregate formation, we are suggesting a more general principle here that does not require polymers, multiple specific interaction sites, or specific secondary structures as in very recent work (72). One factor is simply electrostatic complementarity, but a more generalized concept of multi-valency is implicitly assumed. Isotropic spheres without any directional preference for interactions are in fact infinitely multi-valent. On the other hand, globular proteins with basic amino acids distributed widely across their surface are effectively poly-valent particles with respect to interactions with nucleic acids. The key insight from this study is that proteins not known to interact specifically with nucleic acids under dilute conditions may form condensates with nucleic acids, if the proteins are present at sufficient amounts, simply based on such a generic poly-valency and an overall charge attraction.

Our study suggests that size and charge are essential determinants of PS between RNA and proteins. Favorable condensates require optimal packing and a balance of attractive and repulsive interactions between the oppositely charged RNA and protein particles. Fig. S36 shows a snapshot from the cytoplasmic system illustrating how such packing may be achieved. The optimal balance depends on the size of the RNA particles: Larger proteins are required for the smaller RNA molecules to phase separate whereas smaller proteins allow the larger ribosomal particles to phase-separate (Fig. S4). This can be seen more clearly in the five-component model system, where a relatively modest reduction in the radius of the larger positively charged particle leads to a loss of close tRNA contacts (Fig. S37), therefore preventing condensate formation. The theoretical model for binary RNA-protein mixtures also predicts a minimum protein radius for PS, at least at lower charges (Fig. 5). Myoglobin is outside the predicted range and although

it has a net-positive charge, PS was not observed in the experiment at protein concentrations below the RNA concentrations (Fig. 6) as predicted by the theory.

The total concentration of the protein is in fact another determinant for PS. Simulations and theory predict minimum protein concentrations depending on the protein charge and size around 0.05 mM or more (Figs. 20-25). For trypsin, this was validated experimentally via microscopy and FRET spectroscopy (Figs. 6 and 7). While many cellular proteins may not be present individually at such high concentrations, our cytoplasmic model shows that a heterogeneous mixture of similar-sized and similar-charged proteins may promote PS equally well. The RNA concentration appears to be a less critical factor for observing PS, although a larger amount of RNA allows more numerous and larger condensates to form assuming that there is enough protein available. In binary mixtures, this is simply a question of the total protein concentration. In the heterogeneous cytoplasmic model, we found competition for the larger positively charged proteins by the ribosomes forming their condensates to be another factor affecting tRNA condensate formation that would need to be considered in cellular environments (Fig. S13).

Since electrostatics is a major driving force of the PS described here, changes in salt concentration are expected to alter the tendency for PS. The theory predicts that there is a critical salt concentration depending on the protein below which PS should not be observed anymore (Fig. S18). We were not successful in testing this prediction for the RNA-protein systems studied here. However, there is only a limited range of lower salt conditions that can be applied before either the RNA or the protein structures become destabilized. On the other hand, the spherical models using Debye-Hückel theory only provide a very approximate qualitative idea of how changes in salt concentration may affect intermolecular interactions in the condensates at the atomistic level. Therefore, an accurate quantitative agreement between simulations, theory, and experiment with respect to salt effects should not be expected.

A significant interest in PS in biology is related to liquid-state condensates as they would maintain dynamics that is necessary for many biological processes as opposed to dynamically retarded gels or amorphous clusters. The simulations suggest that the condensates retain a dynamic character based on calculated self-diffusion rates. Because diffusion estimates from CG simulations, especially in the absence of hydrodynamic interactions, are not expected to be quantitatively accurate (73), but we believe that the relative change in diffusion found between dilute and condensed phases is a meaningful prediction. D_{tr} values inside the condensates are around half of the values outside the clusters. Given the higher molecular volume fractions in the denser condensates, such a reduction is in line with reduced diffusion in crowded, liquid solutions (66, 67) whereas the formation of gels or clusters would have resulted in a much more dramatic decrease in diffusion. This prediction remains to be tested experimentally. Other studies have employed fluorescence recovery after photobleaching (FRAP) to establish a liquid character of condensates (35) but there has been significant criticism of this approach, primarily because the observed range of recovery times is enormous and cannot be solely attributed to diffusion. (74-76). In addition, these measurements are problematic for the condensates observed

here because the clusters are generally smaller than the diffraction-limit of the microscope. We believe that nuclear magnetic resonance (NMR) spectroscopy (77) would be better suited to address this question in future work.

There are many limitations in this study that will require to be revisited in future studies to gain a more detailed understanding of the more universal PS between RNA and proteins described here. The advantage of the CG models and theory is that its simplicity allowed us to explore large enough spatial scales and long enough time scales to predict phase behavior and propose experimentally testable hypotheses. Although the CG models were parameterized based on high-resolution atomistic simulations of concentrated simulations, these models lack all but the most basic features of biological macromolecules. Increased levels of realism could be achieved without too much additional computational cost via patchy particles (78), whereas higher-resolution, residue-based coarse-graining (70) is possible for exploring the effects of shape anisotropy and inhomogeneous charge distributions across RNA and protein surfaces. Although such models already place significant limitations on system size and length scales that could be covered. Finally, we expect that further insights could be gained from atomistic simulations of RNA-protein clusters initiated from configurations in the CG simulations to better understand the detailed molecular interactions stabilizing the condensates. On the experimental side, we only focused on RNA without visualizing protein condensation. Moreover, there is a need to follow up on this work with *in vivo* studies to establish how ubiquitous the condensates described here are under cellular conditions.

CONCLUSIONS

We report phase separation of RNA and proteins based on a universal principle of charge complementarity that does not require polymers or multi-valency via specific interactions. The results are supported by coarse-grained simulations, theory, and experimental validation via microscopy and FRET spectroscopy. Condensate formation depends on concentration, size, and charge of the proteins but appears to be possible for typical RNA and common proteins. Simulation results furthermore suggest that such phase separation may occur in heterogenous cellular environment, not just between tRNA and cellular proteins but also, in separate condensates, between ribosomes and proteins. Further computational and experimental studies are needed to gain more detailed insights into the exact molecular nature of the condensates described here.

METHODS

Coarse-grained simulations

CG simulations were run using a modified version of a previously introduced colloid-type spherical model (79). In this model, pair interactions consist of a short-range 10-5 Lennard-Jones potential and a long-range Debye Hückel potential according to:

$$U(r_{ij}) = 4\varepsilon\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{5}\right) + \frac{(A_{ij}+A_0)\kappa\sigma_{ij}}{r_{ij}} e^{-\frac{r_{ij}}{\kappa\sigma_{ij}}} \quad (3)$$

where r_{ij} is the inter-particle distance, σ_{ij} is the distance between particles at which the potential is zero, ε is the strength of short-range attraction, $A_{ij}+A_0$ describes attractive or repulsive long-range interactions, and $\kappa\sigma_{ij}$ is the Debye-Hückel screening length. Only A_{ij} and σ_{ij} vary between different particles according to charge and size (SI Methods).

Most systems involved box sizes of 100 nm and simulations were extended to 1 ms via Langevin dynamics using OpenMM (80) on graphics processing units (GPUs). The CG simulations were analyzed via in-house scripts and the MMTSB Tool Set (81) (SI Methods).

Experimental materials and methods

The J345 RNA sequence was synthesized by Dharmacon, both with and without Cy3 or Cy5 on the 3' end. The 47-base sequence is
GCAGCAGGGAACUCACGCUUGCGUAGAGGCUAAGUGCUUCGGCACAGCACAAGCC
CGCUGCG

All measurements were made using the buffer used by Bonneau and Legault for structure determination of this sequence, 10 mM sodium cacodylate (pH 6.5), 50 mM NaCl, .05% sodium azide, 5 mM $MgCl_2$. Equine liver trypsin, equine alcohol dehydrogenase, bovine lactic dehydrogenase, equine myoglobin, hen egg lysozyme and bovine serum albumin were obtained from Sigma-Aldrich and used without further modification.

Confocal microscopy images were obtained on a Nikon A1 scanning confocal microscope with 100x magnification. The excitation wavelength was 561 nm and detection was set for Cy3 fluorescence using a GaAsP detector. The diffraction-limited spatial resolution is 260 nm. Images were processed with ImageJ and modified only for contrast and brightness. Images were cropped and enlarged to aid observation of the smallest features.

Fluorescence spectra were obtained with PTI Q4 fluorimeter, excited at 475 nm and emission observed between 525 and 700 nm. The concentration of Cy3-labeled RNA and Cy5-labeled RNA were kept constant at 8 uM and 42 uM, respectively, with the unlabeled concentration varied from 0 to 0.5 mM. The normalized FRET ratio was calculated from the total intensity between 525 and 650 nm for the donor and 650 and 700 nm for the acceptor,

$$FRET = \frac{I_A}{I_D + I_A}.$$

In the absence of protein, the RNA exhibits some baseline transfer, likely due to transient interactions between the dyes, leading to a background FRET level of ~0.24. Upon the addition of protein above the threshold concentration, the mixture is visibly turbid.

Circular dichroism measurements were made using an Applied Photophysics Chirascan spectrometer. All measurements were made using a 0.1 mm pathlength cuvette at room temperature.

ACKNOWLEDGMENTS

This study was funded by the National Science Foundation (MCB 1817307) and the National Institutes of Health (R35 GM126948). Computational resources at the Institute for Cyber-Enabled Research/High Performance Computing Cluster (ICER/HPCC) at Michigan State University were used. We thank Prof. Charles G. Hoogstraten for insightful discussions.

REFERENCES

1. S. Boulon, B. J. Westman, S. Hutten, F. M. Boisvert, A. I. Lamond, The Nucleolus under Stress. *Mol Cell* **40**, 216-227 (2010).
2. D. S. W. Protter, R. Parker, Principles and Properties of Stress Granules. *Trends Cell Biol* **26**, 668-679 (2016).
3. F. M. Boisvert, S. van Koningsbruggen, J. Navascues, A. I. Lamond, The multifunctional nucleolus. *Nat Rev Mol Cell Bio* **8**, 574-585 (2007).
4. M. Morimoto, C. F. Boerkoel, The Role of Nuclear Bodies in Gene Expression and Disease. *Biology* **2**, 976-1033 (2013).
5. X. L. Su *et al.*, Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595-599 (2016).
6. S. Alberti, A. Gladfelter, T. Mittag, Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419-434 (2019).
7. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Bio* **18**, 285-298 (2017).
8. S. Boeynaems *et al.*, Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**, 420-435 (2018).
9. J. A. Ditlev, L. B. Case, M. K. Rosen, Who's In and Who's Out-Compositional Control of Biomolecular Condensates. *J Mol Biol* **430**, 4666-4684 (2018).
10. M. Feric *et al.*, Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686-1697 (2016).
11. O. V. Iarovaia *et al.*, Nucleolus: A Central Hub for Nuclear Functions. *Trends Cell Biol* **29**, 647-659 (2019).
12. L. Galganski, M. O. Urbanek, W. J. Krzyzosiak, Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res* **45**, 10350-10368 (2017).
13. A. I. Lamond, D. L. Spector, Nuclear speckles: A model for nuclear organelles. *Nat Rev Mol Cell Bio* **4**, 605-612 (2003).
14. M. Cioce, A. I. Lamond, Cajal bodies: A long history of discovery. *Annu Rev Cell Dev Biol* **21**, 105-131 (2005).
15. J. G. Gall, Cajal bodies: The first 100 years. *Annu Rev Cell Dev Biol* **16**, 273-+ (2000).
16. M. Machyna, P. Heyn, K. M. Neugebauer, Cajal bodies: where form meets function. *Wires Rna* **4**, 17-34 (2013).
17. K. A. Burke, A. M. Janke, C. L. Rhine, N. L. Fawzi, Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol Cell* **60**, 231-241 (2015).
18. A. Molliex *et al.*, Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163**, 123-133 (2015).
19. C. P. Brangwynne *et al.*, Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **324**, 1729-1732 (2009).
20. E. Voronina, G. Seydoux, P. Sassone-Corsi, I. Nagamori, RNA Granules in Germ Cells. *Csh Perspect Biol* **3** (2011).
21. S. A. Fromm *et al.*, In Vitro Reconstitution of a Cellular Phase-Transition Process that Involves the mRNA Decapping Machinery. *Angew Chem Int Edit* **53**, 7354-7359 (2014).

22. Y. Luo, Z. K. Na, S. A. Slavoff, P-Bodies: Composition, Properties, and Functions. *Biochemistry* **57**, 2424-2431 (2018).
23. S. Elbaum-Garfinkle *et al.*, 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 (2015).
24. K. M. Ruff, S. Roberts, A. Chilkoti, R. V. Pappu, Advances in Understanding Stimulus Responsive Phase Behavior of Intrinsically Disordered Protein Polymers. *J Mol Biol* **430**, 4619-4635 (2018).
25. I. A. Sawyer, D. Sturgill, M. Dundr, Membraneless nuclear organelles and the search for phases within phases. *Wires Rna* **10** (2019).
26. M. M. Fay, P. J. Anderson, The Role of RNA in Biological Phase Separations. *J Mol Biol* **430**, 4685-4701 (2018).
27. J. S. Andersen *et al.*, Nucleolar proteome dynamics. *Nature* **433**, 77-83 (2005).
28. K. W. Fong *et al.*, Whole-genome screening identifies proteins localized to distinct nuclear bodies. *J Cell Biol* **203**, 149-164 (2013).
29. T. Chujo, T. Yamazaki, T. Hirose, Architectural RNAs (arcRNAs): A class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Bba-Gene Regul Mech* **1859**, 139-146 (2016).
30. C. M. Clemson *et al.*, An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol Cell* **33**, 717-726 (2009).
31. H. Falahati, B. Pelham-Webb, S. Blythe, E. Wieschaus, Nucleation by rRNA Dictates the Precision of Nucleolus Assembly. *Curr Biol* **26**, 277-285 (2016).
32. D. M. Mitrea *et al.*, Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *Elife* **5** (2016).
33. J. Smith *et al.*, Spatial patterning of Pgranules by RNA-induced phase separation of the intrinsically-disordered protein MEG-3. *Elife* **5** (2016).
34. Y. Shin *et al.*, Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell* **168**, 159-+ (2017).
35. N. O. Taylor, M. T. Wei, H. A. Stone, C. P. Brangwynne, Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching. *Biophys J* **117**, 1285-1300 (2019).
36. Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell* **60**, 208-219 (2015).
37. S. Alberti, A. A. Hyman, Are aberrant phase transitions a driver of cellular aging? *Bioessays* **38**, 959-968 (2016).
38. S. C. Weber, Sequence-encoded material properties dictate the structure and function of nuclear bodies. *Curr Opin Cell Biol* **46**, 62-71 (2017).
39. S. C. Weber, C. P. Brangwynne, Getting RNA and Protein in Phase. *Cell* **149**, 1188-1191 (2012).
40. D. M. Mitrea *et al.*, Methods for Physical Characterization of Phase-Separated Bodies and Membrane-less Organelles. *J Mol Biol* **430**, 4773-4805 (2018).
41. S. F. Banani *et al.*, Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663 (2016).
42. T. J. Nott *et al.*, Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol Cell* **57**, 936-947 (2015).

43. J. P. Brady *et al.*, Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc Natl Acad Sci U S A* **114**, E8194-E8203 (2017).
44. D. M. Mitrea *et al.*, Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation. *Nat Commun* **9** (2018).
45. M. T. Wei *et al.*, Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nat Chem* **9**, 1118-1125 (2017).
46. M. Kato *et al.*, Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* **149**, 753-767 (2012).
47. Y. Lin *et al.*, Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **167**, 789-+ (2016).
48. P. L. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336-340 (2012).
49. J. A. Riback *et al.*, Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **168**, 1028-1040 (2017).
50. C. P. Brangwynne, T. J. Mitchison, A. A. Hyman, Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc Natl Acad Sci U S A* **108**, 4334-4339 (2011).
51. S. Maharana *et al.*, RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918-921 (2018).
52. G. L. Dignon, W. W. Zheng, J. Mittal, Simulation methods for liquid-liquid phase separation of disordered proteins. *Curr Opin Chem Eng* **23**, 92-98 (2019).
53. G. L. Dignon, W. W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput Biol* **14** (2018).
54. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nat Phys* **11**, 899-904 (2015).
55. A. E. Posey, A. S. Holehouse, R. V. Pappu, Phase Separation of Intrinsically Disordered Proteins. *Methods Enzymol* **611**, 1-30 (2018).
56. V. Nguemaha, H. X. Zhou, Liquid-Liquid Phase Separation of Patchy Particles Illuminates Diverse Effects of Regulatory Components on Protein Droplet Formation. *Sci Rep-Uk* **8** (2018).
57. S. B. Qin, H. X. Zhou, Protein folding, binding, and droplet formation in cell-like conditions. *Curr Opin Struct Biol* **43**, 28-37 (2017).
58. M. A. Woldeyes, C. Calero-Rubio, E. M. Furst, C. J. Roberts, Predicting Protein Interactions of Concentrated Globular Protein Solutions Using Colloidal Models. *J Phys Chem B* **121**, 4756-4767 (2017).
59. S. Rauscher, R. Pomes, The liquid structure of elastin. *Elife* **6** (2017).
60. C. W. Pak *et al.*, Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell* **63**, 72-85 (2016).
61. J. Y. Fei *et al.*, Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J Cell Sci* **130**, 4180-4192 (2017).
62. Y. H. Lin, J. D. Forman-Kay, H. S. Chan, Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles. *Phys Rev Lett* **117** (2016).
63. M. Feig *et al.*, Complete atomistic model of a bacterial cytoplasm for integrating physics, biochemistry, and systems biology. *J Mol Graph Model* **58**, 1-9 (2015).

64. I. Yu *et al.*, Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* **5** (2016).
65. A. C. Dumetz, A. M. Chockla, E. W. Kaler, A. M. Lenhoff, Protein phase behavior in aqueous solutions: Crystallization, liquid-liquid phase separation, gels, and aggregates. *Biophys J* **94**, 570-583 (2008).
66. N. Muramatsu, A. P. Minton, Tracer Diffusion of Globular-Proteins in Concentrated Protein Solutions. *Proc Natl Acad Sci U S A* **85**, 2984-2988 (1988).
67. S. B. Zimmerman, A. P. Minton, Macromolecular Crowding - Biochemical, Biophysical, and Physiological Consequences. *Annu Rev Biophys Biomol Struct* **22**, 27-65 (1993).
68. E. Bonneau, P. Legault, Nuclear Magnetic Resonance Structure of the III–IV–V Three-Way Junction from the Varkud Satellite Ribozyme and Identification of Magnesium-Binding Sites Using Paramagnetic Relaxation Enhancement. *Biochemistry* **53**, 6264-6275 (2014).
69. A. Ghosh, K. Mazarakos, H.-X. Zhou, Three archetypical classes of macromolecular regulators of protein liquid–liquid phase separation. *Proc Natl Acad Sci USA* **116**, 19474 (2019).
70. G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, J. Mittal, Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc Natl Acad Sci USA* **115**, 9929 (2018).
71. S. Banjade, M. K. Rosen, Phase Transitions of Multivalent Proteins Can Promote Clustering of Membrane Receptors. *Elife* **3** (2014).
72. A. E. Conicella *et al.*, TDP-43 α-helical structure tunes liquid–liquid phase separation and function. *Proc Natl Acad Sci USA* **117**, 5883-5894 (2020).
73. T. Ando, J. Skolnick, Crowding and Hydrodynamic Interactions Likely Dominate *in vivo* Macromolecular Motion. *Proc Natl Acad Sci USA* **107**, 18457-18462 (2010).
74. M. Kang, C. A. Day, E. DiBenedetto, A. K. Kenworthy, A quantitative approach to analyze binding diffusion kinetics by confocal FRAP. *Biophys J* **99**, 2737-2747 (2010).
75. F. Mueller, D. Mazza, T. J. Stasevich, J. G. McNally, FRAP and kinetic modeling in the analysis of nuclear protein dynamics: what do we really know? *Curr Opin Cell Biol* **22**, 403-411 (2010).
76. D. T. McSwiggen, M. Mir, X. Darzacq, R. Tjian, Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes & Development* **33**, 1619-1634 (2019).
77. C. Li, Y. Wang, G. J. Pielak, Translational and Rotational Diffusion of a Small Globular Protein under Crowded Conditions. *J Phys Chem B* **113**, 13390-13392 (2009).
78. V. Nguemaha, H.-X. Zhou, Liquid-Liquid Phase Separation of Patchy Particles Illuminates Diverse Effects of Regulatory Components on Protein Droplet Formation. *Sci Rep-Uk* **8**, 6728 (2018).
79. E. Mani, W. Lechner, W. K. Kegel, P. G. Bolhuis, Equilibrium and Non-Equilibrium Cluster Phases in Colloids with Competing Interactions. *Soft Matter* **10**, 4479-4486 (2014).
80. P. Eastman *et al.*, OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *Plos Comp Biol* **13**, e1005659 (2017).
81. M. Feig, J. Karanicolas, C. L. Brooks III, MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology. *J Mol Graph Modell* **22**, 377-395 (2004).