PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions

Mrityunjoy Kar[a], Furqan Dar[b], Timothy J. Welsh[c], Laura T. Vogel[d], Ralf Kühnemuth[d], Anupa Majumdar[a], Georg Krainer[c], Titus M. Franzmann[e], Simon Alberti[e], Claus A. M. Seidel[d], Tuomas P. J. Knowles[c,f], Anthony A. Hyman[a,1], and Rohit V. Pappu[g,1]

Contributed by Anthony A. Hyman; received February 13, 2022; accepted May 26, 2022; reviewed by Ashok Deniz, Eugene Shakhnovich, and Peter Vekilov

Macromolecular phase separation is thought to be one of the processes that drives the formation of membraneless biomolecular condensates in cells. The dynamics of phase separation are thought to follow the tenets of classical nucleation theory, and, therefore, subsaturated solutions should be devoid of clusters with more than a few molecules. We tested this prediction using in vitro biophysical studies to characterize subsaturated solutions of phase-separating RNA-binding proteins with intrinsically disordered prion-like domains and RNA-binding domains. Surprisingly, and in direct contradiction to expectations from classical nucleation theory, we find that subsaturated solutions are characterized by the presence of heterogeneous distributions of clusters. The distributions of cluster sizes, which are dominated by small species, shift continuously toward larger sizes as protein concentrations increase and approach the saturation concentration. As a result, many of the clusters encompass tens to hundreds of molecules, while less than 1% of the solutions are mesoscale species that are several hundred nanometers in diameter. We find that cluster formation in subsaturated solutions and phase separation in supersaturated solutions are strongly coupled via sequence-encoded interactions. We also find that cluster formation and phase separation can be decoupled using solutes as well as specific sets of mutations. Our findings, which are concordant with predictions for associative polymers, implicate an interplay between networks of sequence-specific and solubility-determining interactions that, respectively, govern cluster formation in subsaturated solutions and the saturation concentrations above which phase separation occurs.

mesoscale clusters | stickers and spacers | phase separation | associative polymers | sol–gel transitions

Phase separation of RNA-binding proteins with disordered prion-like domains (PLDs) and RNA-binding domains (RBDs) is implicated in the formation and dissolution of membraneless biomolecular condensates such as RNA–protein (RNP) granules (1–9). Macroscopic phase separation is a process whereby a macromolecule in a solvent separates into a dilute, macromolecule-deficient phase that coexists with a dense, macromolecule-rich phase (10, 11). In a binary mixture, the soluble phase, comprising dispersed macromolecules that are well mixed with the solvent, becomes saturated at a concentration designated as c_{sat}. Above c_{sat}, for total macromolecular concentrations c_{tot} that are between the binodal and spinodal, phase separation of full-length RNA-binding proteins and PLDs is thought to follow classical nucleation theory (12–15).

In classical nucleation theories, clusters representing incipient forms of the new dense phase form within dispersed phases of supersaturated solutions defined by $c_{tot} > c_{sat}$ (16, 17). In the simplest formulation of classical nucleation theory (16–18), the free energy of forming a cluster of radius a is $\Delta F = -\frac{4\pi}{3}a^3\Delta\mu\rho_n + 4\pi a^2\gamma$. Here, $\Delta\mu$ is the difference in the chemical potential between the one-phase and two-phase regimes (see discussion in *SI Appendix*), which is negative in supersaturated solutions and positive in subsaturated solutions; ρ_n is the number of molecules per unit volume, and γ is the interfacial tension between dense and dilute phases. At temperature T, in a seed-free solution, the degree of supersaturation s is defined as $s \equiv \frac{\Delta\mu}{RT} = \ln\left(\frac{c_{tot}}{c_{sat}}\right)$, where R is the ideal gas constant. Here, s is positive for $c_{tot} > c_{sat}$, and, as s increases, cluster formation becomes more favorable. Above a critical radius a^*, the free energy of cluster formation can overcome the interfacial penalty, and the new dense phase grows in a thermodynamically downhill fashion. Ideas from classical nucleation theory have been applied to analyze and interpret the dynamics of phase separation in supersaturated solutions (12, 13, 15). Classical nucleation theories stand in contrast to two-step nucleation theories that predict the existence of prenucleation clusters in supersaturated solutions (19–22). These newer theories hint at the prospect of there being interesting features in subsaturated solutions, where $c_{tot} < c_{sat}$ and $s < 0$.

Significance

Membraneless biomolecular condensates are molecular communities with distinct compositional preferences and functions. Considerable attention has focused on macroscopic phase separation as the process that gives rise to condensates. This focus implicitly presumes that coexisting dilute phases and subsaturated solutions are featureless. Here, we show that this is not the case. Surprisingly, the FUS-EWSR1-TAF15 family of RNA-binding proteins form heavy-tailed distributions of clusters in subsaturated solutions. Given the low endogenous concentrations of phase-separating proteins, it stands to reason that the clusters we have discovered in subsaturated conditions will be present in cellular milieus. Knowing that such clusters exist opens the door to assessing the functional relevance of clusters of different sizes and shapes that form in subsaturated solutions or coexist with condensates.

Reviewers: A.D., The Scripps Research Institute; E.S., Harvard University; and P.V., University of Houston.

Competing interest statement: A.A.H. is a founder of and consultant for Dewpoint Therapeutics Inc. and a founder of Caraway Therapeutics. S.A. and R.V.P. are members of the scientific advisory board of Dewpoint Therapeutics Inc. These companies are not currently involved in the work in this paper. C.A.M.S. was involved in a collaborative review article published alongside A.D.

See online for related content such as Commentaries.

[1]To whom correspondence may be addressed. Email: hyman@mpi-cbg.de or pappu@wustl.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2202222119/-/DCSupplemental.

Published July 5, 2022.

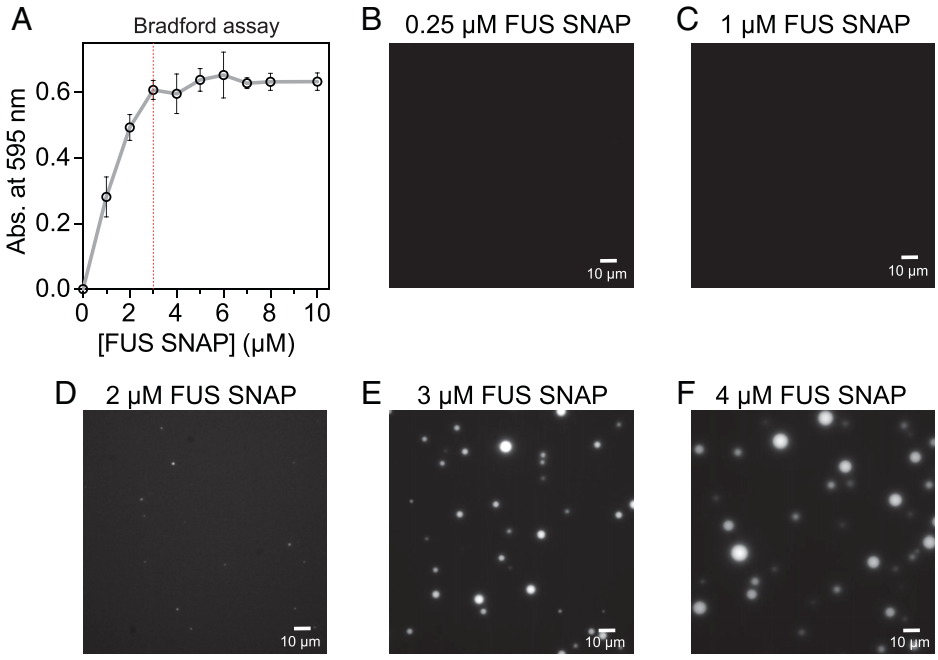

Fig. 1. FUS-SNAP has a quantifiable c_{sat} whereby condensates form only in supersaturated solutions. (*A*) Sample data for absorbance-based spin-down assays. Data shown here are for FUS-SNAP in 20 mM Tris, pH 7.4, and 100 mM KCl at ~25 °C. The red dashed line intersects the abscissa at 3 μM, which is the inferred c_{sat} for this construct. (*B–F*) Microscopy images collected at the 18-h time point for solutions containing different concentrations of FUS-SNAP in 20 mM Tris, pH 7.4, and 100 mM KCl at ~25 °C. For imaging purposes, 5% of the total mixture in each sample is made up of FUS-eGFP. (Scale bar, 10 μm.) The data show the presence of micron-scale condensates at or above c_{sat} of ~3 μM. The FUS constructs used were expressed and purified using method A (*SI Appendix*).

The subsaturated regime, where s is negative, corresponds to the one-phase regime. Ignoring the interfacial tension, the free energy of realizing clusters with n molecules in subsaturated solutions is: $\Delta F = -n\Delta\mu$. Therefore, the probability $P(n)$ of forming a cluster of n molecules in a subsaturated solution is proportional to $\exp(sn)$. Accordingly, the relative probability $P(n)/P(1)$ of forming clusters with n molecules will be $\exp(s(n-1))$. This quantity, which may be thought of as the concentration of clusters with n molecules, is negligibly small for clusters with more than a few molecules. This is true irrespective of the degree of subsaturation, s. Is this expectation from classical nucleation theories valid? We show here that subsaturated solutions feature a rich distribution of species not anticipated by classical nucleation theories. We report results from measurements of cluster size distributions in subsaturated solutions of phase-separating RNA-binding proteins from the FUS-EWSR1-TAF15 (FET) family. We find that these systems form clusters in subsaturated solutions, and that the cluster sizes follow heavy-tailed distributions. The abundant species are always small clusters. However, as total macromolecular concentration (c_{tot}) increases, the distributions of cluster sizes shift continuously toward larger values. We discuss these findings in the context of theories for associative polymers (9, 23–30).

Results

Macroscopic Phase Separation Is Not Observed in Subsaturated Solutions. We quantified c_{sat} using a spin-down-based absorbance assay. At low concentrations, any prior aggregates that are present in solution are removed upon centrifugation at 20,000 relative centrifugal force (RCF) units. Results from this assay are summarized in Fig. 1*A* for FUS-SNAP. At low c_{tot}, the concentration in the supernatant increases monotonically with the bulk concentration of SNAP-tagged FUS. However, above a threshold concentration, which we designate as c_{sat}, the concentration in the supernatant remains fixed at a plateau value. This is consistent with the establishment of phase equilibrium between dilute and dense phases (Fig. 1*A*).

To test whether c_{sat} is a true saturation concentration, we performed microscopy-based measurements for solutions containing different amounts of FUS-SNAP and untagged FUS (*SI Appendix*, Fig. S1). The concentrations investigated range from 0.25 μM to 4 μM. Data were collected ~30 min after sample preparation, and, for each sample, a series of images were collected at different time points over an 18-h period. Results at the 18-h time point are shown in Fig. 1 *B–F* for different concentrations of FUS-SNAP. Irrespective of the time point of interrogation, condensate formation of FUS-SNAP is only detectable at or above $c_{sat} \approx 3$ μM. Similar data were obtained for untagged FUS (*SI Appendix*, Fig. S1). Based on these results, we conclude that macroscopic phase separation is not realized in subsaturated solutions and that the FUS molecules are defined by construct- and solution condition–specific saturation concentrations.

Clusters Spanning a Range of Sizes Form in Subsaturated Solutions of FUS. First, we used dynamic light scattering (DLS) to characterize subsaturated solutions of untagged FUS (Fig. 2*A* and *SI Appendix*, Fig. S2) and FUS-SNAP (Fig. 2 *B* and *C* and *SI Appendix*, Fig. S3). Unless otherwise specified, all measurements were performed in 20 mM Tris, pH 7.4, and 100 mM KCl at ~25 °C. The DLS results we obtain are robust to the protocols used to express and purify FUS and FUS-SNAP molecules (*SI Appendix*, Fig. S4).

The DLS data are shown as scattering intensities plotted against the apparent hydrodynamic diameter (d_h). For calibration, the d_h of untagged FUS monomers is ~7 nm. The intensity profiles for untagged FUS collected at 0.25 μM show two peaks (Fig. 2*A*). The peak at ~8 nm corresponds to monomers and oligomers. The second peak at ~200 nm corresponds to mesoscale clusters. The observed bimodality could imply clusters of fixed size forming via either microphase separation (31) or micellization (32). To test for this possibility, we investigated

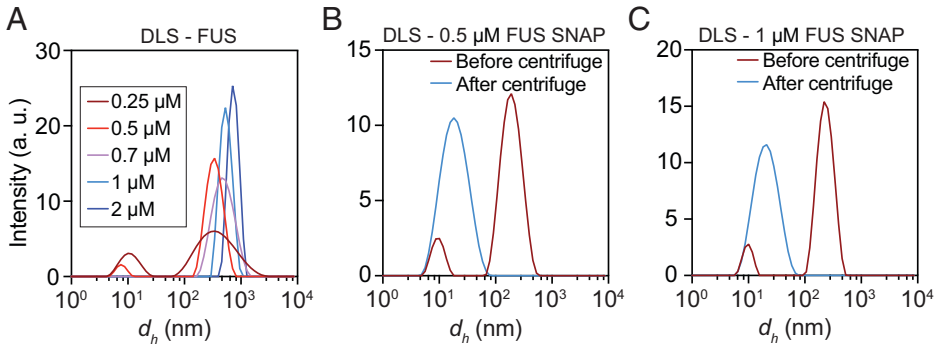

Fig. 2. DLS data show that clusters spanning a range of sizes form in subsaturated solutions. (*A*) Measurements were performed at different bulk concentrations of untagged FUS (see legend) that represent different degrees of subsaturation $-2 \leq s \leq 0$. Scattering intensities, measured 8 min after sample preparation, shift toward larger values as the total macromolecular concentration c_{tot} approaches the c_{sat} of ~2 μM. (*B* and *C*) Scattering intensity was measured at 0.5 μM (*B*) and 1 μM (*C*) of FUS-SNAP at 8 min after sample preparation and before centrifugation (brown curves). The solutions were centrifuged at 20,000 RCF for 5 min at room temperature. Then, scattering intensities of the supernatant were measured at the 8-min time point. These data are shown in blue curves in *B* and *C*. They show the presence of species of sizes ranging between 7 and 100 nm. The FUS constructs were expressed and purified using method A (*SI Appendix*).

how the DLS signals change with increasing protein concentrations (Fig. 2*A*). For microphase separation or micellization, the locations of the peaks should stay in place while the intensities of the peaks change with respect to one another. We do not observe this behavior. Instead, the location of the peak corresponding to larger d_h values shifts to the right. At higher concentrations, DLS becomes blind to the presence of smaller species. As a result, the peak at lower values of d_h vanishes above a concentration of 0.7 μM.

The apparent bimodality seen in intensity profiles can also arise from a heavy-tailed distribution of cluster sizes, whereby the most abundant species are monomers and oligomers. In this scenario, bimodality at low concentrations below 0.7 μM can result from smaller species being the most abundant, and the largest, mesoscale species having the highest scattering cross-sections (33). This would mask the presence of species of intermediate sizes. To test for the presence of species of intermediate sizes, we subjected subsaturated solutions of FUS-SNAP molecules to centrifugation at 20,000 RCF. This removes species larger than 100 nm in diameter. Although the bulk concentrations prior to centrifugation were 0.5 and 1 μM , the centrifugation step lowers the total macromolecular concentration. The top of the centrifuged solution was then collected for DLS measurements. The DLS data, collected after centrifugation, are shown in Fig. 2 *B* and *C* for FUS-SNAP. The scattering intensity profiles, measured after centrifugation, show the presence of species in the size range of 7 nm to 100 nm for the apparent d_h. These species include monomers and an assortment of higher-order species. The upper limit on the number of FUS molecules per cluster (n_{mol}) is estimated as the ratio of the hydrodynamic volume of a cluster of size d_h to that of the monomer. This yields upper bounds on estimates of n_{mol} of 10, 10^2, and 10^3 for d_h values of ~15, ~30, and ~70 nm, respectively.

Taken together with data obtained prior to centrifugation, the parsimonious interpretation of the data is that the distributions of cluster sizes are likely to be exponentials with heavy tails, such as Weibull distributions (34). This would be concordant with reversible processes such as isodesmic associations that lack a threshold concentration, with average sizes evolving continuously as concentrations increase (35, 36).

Morphologies of the Largest Clusters That Form in Subsaturated Solutions. We used negative stain transmission electron microscopy (TEM) to characterize morphologies of the largest clusters that form in subsaturated solutions of FUS-SNAP. Given their sizes, we refer to these species as mesoscale clusters. These clusters have roughly spherical morphologies (Fig. 3*A* and *SI Appendix*, Fig. S5). This was further confirmed using DLS data collected for untagged FUS at two different angles (Fig. 3*B*). Based on their quasi-spherical morphologies and plausible densities within clusters, we estimate that mesoscale clusters with diameters of ~150 nm encompass roughly 10^2 to 10^3 molecules.

Mesoscale Clusters Are of Low Overall Abundance. We quantified the abundance of mesoscale clusters using nanoparticle tracking analysis (NTA) (37). In these measurements, we track the Brownian motion of scatterers using dark-field microscopy (Movies S1 and S2). We collected NTA data for untagged FUS (Fig. 4*A*) and FUS-SNAP (Fig. 4*B*). The distributions of mesoscale cluster sizes shift toward larger values with increasing protein concentrations. NTA also enables quantification of the percentage of molecules in solution that make up mesoscale clusters (*SI Appendix*, Fig. S6). In a 250-nM solution of FUS-SNAP, the total volume fraction of proteins is $\phi \approx 1.4 \times 10^{-5}$, and the fraction of the solution volume taken up by mesoscale clusters is ϕ_{meso} $\sim 4 \times 10^{-8}$. This translates to 0.15% of the protein molecules in solution being a part of mesoscale clusters (Fig. 4*C*). For a concentration just below c_{sat}, the relative abundance of mesoscale clusters

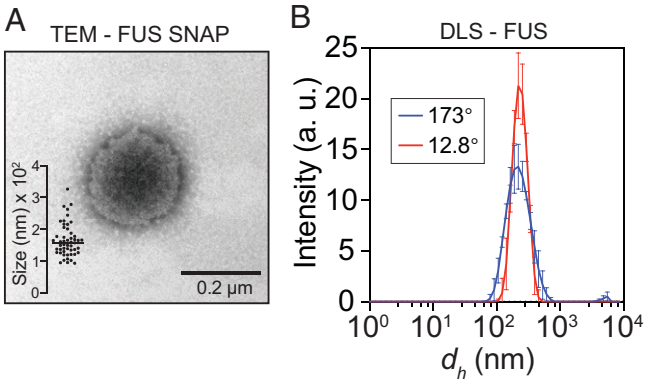

Fig. 3. Mesoscale clusters that form in subsaturated solutions have quasi-spherical morphologies. (A) Representative TEM image of 2 μM FUS-SNAP. The collections of TEM images were used to quantify the distribution of sizes (*Inset* and *SI Appendix*, Fig. S6). (*B*) DLS data collected at two different angles corroborate the spherical morphologies of clusters observed in subsaturated solutions. Samples used here were prepared using method A (*SI Appendix*).

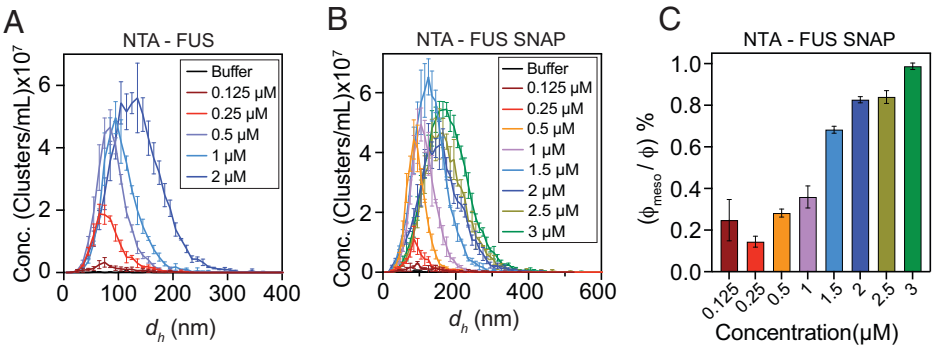

Fig. 4. Mesoscale clusters have low overall abundance. (*A*) NTA data for untagged FUS (c_{sat} ≈ 2 μM) were collected at different concentrations that represent different degrees of subsaturation. (*B*) NTA data for FUS-SNAP (c_{sat} ≈ 3 μM) at different degrees of subsaturation. (*C*) Relative abundance of mesoscale clusters, quantified at different degrees of subsaturation for FUS-SNAP. The constructs used were expressed and purified using method A (*SI Appendix*).

increases to ∼1%. Similar results, highlighting the low abundance of mesoscale clusters,were obtained for untagged FUS (*SI Appendix*, Fig. S6).

Clusters Increase in Size with Increasing Concentrations, and the Size Distributions Have Heavy Tails. Are non-mesoscale clusters monomers? To answer this question, we probed the distributions across a broader range of sizes using orthogonal, fluorescence-based approaches to characterize subsaturated solutions and access length scales that are distinct from those accessed using DLS and NTA measurements. Fluorescence anisotropy can be applied for detecting the presence of clusters with different numbers of molecules. These experiments use a small fraction of fluorescently labeled molecules (tracers) and a majority fraction of molecules that lack a fluorescent label. We used confocal multiparameter detection (MFD), which measures the fluorescence anisotropies of diffusing particles (38). This allows us to monitor the size-dependent rotational diffusion of molecules. We performed experiments with a fixed concentration (100 nM) of FUS-eGFP and added FUS-SNAP as a titrant with the concentration of FUS-SNAP ranging from 0 μM to 5 μM. The anisotropy increases monotonically at low concentrations of FUS-SNAP (Fig. 5*A*). This indicates the presence of smaller clusters such as dimers, trimers, tetramers, etc. each with increased volumes that result in longer times for molecular rotational diffusion. The anisotropy of eGFP is especially sensitive to the presence of small clusters. However, this signal saturates for larger clusters. This is because the measured apparent volume has a nonlinear scaling due to the inverse relationship between anisotropy and volume. Further, the flexible FUS backbone and long eGFP linker may diminish the actual change in anisotropy. To test for the presence of clusters larger than oligomers, we used the environmentally sensitive dye Nile Red, which offers a higher dynamic range in anisotropy. With Nile Red, we see a strong continuous increase in anisotropy as the concentration of FUS-SNAP is increased (Fig. 5*A*). The signal does not saturate as the concentrations of FUS-SNAP increase. That this lack of saturation emanates from large bright clusters is confirmed in the binned histogram of fluorescence intensity of the photon trace (Fig. 5*B*). These data show that the size distributions of clusters that form in subsaturated solutions have heavy tails, andthe distributions shift continually toward larger cluster sizes as concentrations increase.

Next, we turned to confocal detection under flow by using microfluidic confocal spectroscopy (MCS). This is a brightness-based method that combines microfluidic mixing and flow with confocal detection (39–41). MCS measurements are sensitive to the presence of species that span a broad size range, from tens to hundreds of nanometers. Due to convective flow, the measurement is not reliant on diffusion alone for sampling species of different sizes in solution (*SI Appendix*, Fig. S7). We used FUS-eGFP (c_{sat} ≈ 4 μM; *SI Appendix*, Fig. S8) for the MCS measurements. Data obtained at concentrations of 0.25, 0.5, and 1 μM were plotted as the brightness per event. These

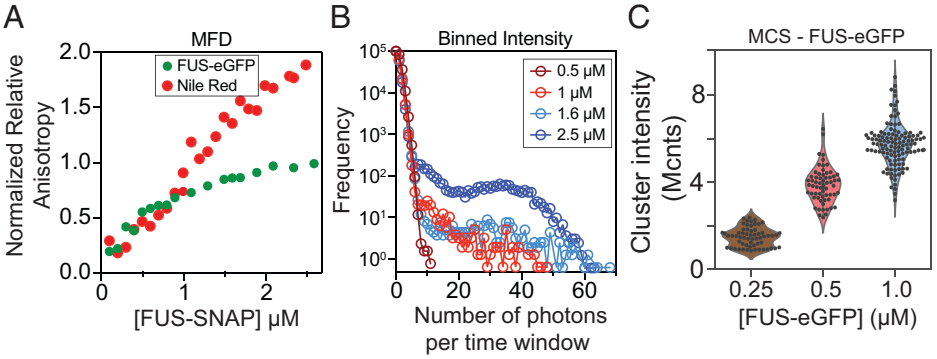

Fig. 5. Fluorescence anisotropy and brightness-based measurements show the presence of clusters in subsaturated solutions. (*A*) Normalized fluorescence anisotropy from confocal MFD plotted against the concentration FUS-SNAP, which is the titrant. Data were collected at 50 mM KCl. The sample was prepared using method B. We also titrated FUS-SNAP from 0 μM to 3 μM in the presence of 20 nM Nile Red. Unlike the anisotropy measured using eGFP fluorescence (green), the measurements based on Nile Red (red) do not show saturation behavior, implying a continuous growth of cluster sizes with the concentration of FUS-SNAP. (*B*) With increasing FUS-SNAP concentrations in the presence of 20 nM Nile Red, we observe bright particles, with apparent heavy-tailed distributions. (*C*) MCS data showing the formation of clusters in subsaturated solutions comprising FUS-GFP, prepared using method A. The apparent sizes of clusters, quantified in terms of photon intensity per cluster where 1 Mcnt = 1 million photons per millisecond, increase with increasing protein concentration. The FUS-GFP constructs used in the MCS experiments were expressed and purified using method A (*SI Appendix*).

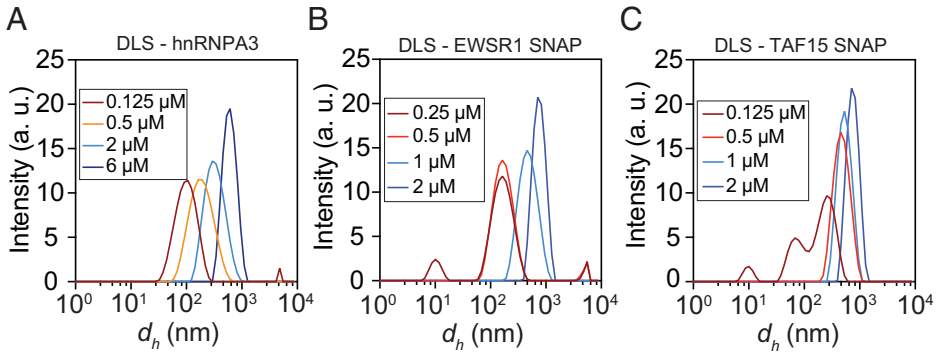

Fig. 6. Clusters also form in subsaturated solutions of FET family proteins. DLS data show evidence for mesoscale clusters in subsaturated solutions for (*A*) hnRNP-A3 ($c_{sat} \approx 6$ μM), (*B*) SNAP-tagged TAF15 ($c_{sat} \approx 2$ μM), and (*C*) SNAP-tagged EWSR1 ($c_{sat} \approx 2$ μM).

data clearly show the formation of clusters in subsaturated solutions whose average sizes evolve continuously with concentration (Fig. 5*C*).

Other FET Family Proteins Also Form Clusters in Subsaturated Solutions. Next, we asked whether the formation of clusters in subsaturated solutions is unique to FUS or whether this is a feature that is shared by other members of the FET family of RNA-binding proteins. We collected DLS data in subsaturated solutions for untagged hnRNP-A3 ($c_{sat} \approx 6$ μM; *SI Appendix*, Fig. S9*A* and Fig. 6*A*), EWSR1-SNAP ($c_{sat} \approx 2$ μM; *SI Appendix*, Fig. S9*B* and Fig. 6*B*), and TAF15-SNAP ($c_{sat} \approx 2$ μM; *SI Appendix*, Fig. S9*C* and Fig. 6*C*). These data show that other FET family proteins also form clusters in subsaturated solutions. The average sizes of these clusters increase with protein concentration. Raw DLS data in the form of autocorrelation functions are shown in *SI Appendix*, Figs. S10–S12.

Clusters Are Reversible, and Molecules Exchange between Clusters. Using DLS, we find that the sizes of clusters decrease upon dilution and increase with increased concentration in the subsaturated regime (Fig. 7 *A*–*C* and *SI Appendix*, Fig. S13). This observation points to the reversibility of cluster formation in subsaturated solutions whereby they shrink upon dilution and grow when concentrations increase. For cluster formation to be reversible, the molecules must exchange between clusters or between clusters and the bulk solution. To test for this, we used Förster resonance energy transfer (FRET) experiments (setup shown in Fig. 7*D*). These experiments show that FUS molecules readily exchange across clusters (Fig. 7*E* and *SI Appendix*, Fig. S14).

The data presented in Figs. 2–7 show the presence of clusters in subsaturated solutions. The average cluster sizes increase as concentrations increase and approach c_{sat}. The clusters are equilibrium species that form and dissolve via reversible

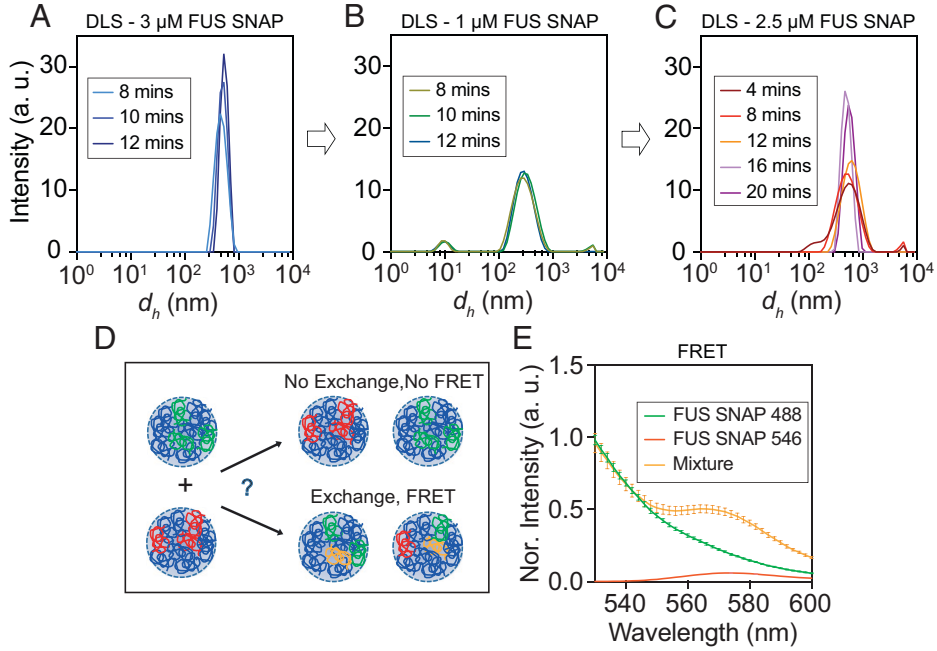

Fig. 7. Clusters in subsaturated solution form via reversible associations, and molecules readily exchange between clusters. (*A*) DLS data were collected at different time points for SNAP-tagged FUS at 3 μM. (*B*) Upon dilution to 1 μM, the DLS data show changes to the intensity profiles, with the appearance of smaller species. (*C*) Increasing the concentration from 1 μM to 2.5 μM leads to an increased preference for larger species. (*D*) Schematic showing the design of the FRET assay. Here, two sets of clusters are formed, each using a total concentration of 1 μM SNAP-tagged FUS molecules. In each set, 5% of the molecules carry a fluorescent label (Alexa Fluor 488 [green] or Alexa Fluor 46 [red]). The clusters are mixed to achieve a total concentration of 1 μM. The mixture is excited using a 488-nm laser, and the emission spectrum is measured from 520 nm to 600 nm. If molecules exchange between the clusters, then we expect to see a peak at the excitation maximum of 573 nm. (*E*) Fluorescence emission spectra show the decay of fluorescence for molecules with the Alexa 488 label and for the mixture. The latter shows a maximum at 573 nm, which is indicative of the exchange of molecules between the clusters. The FUS constructs used in this experiment were expressed and purified using method A.

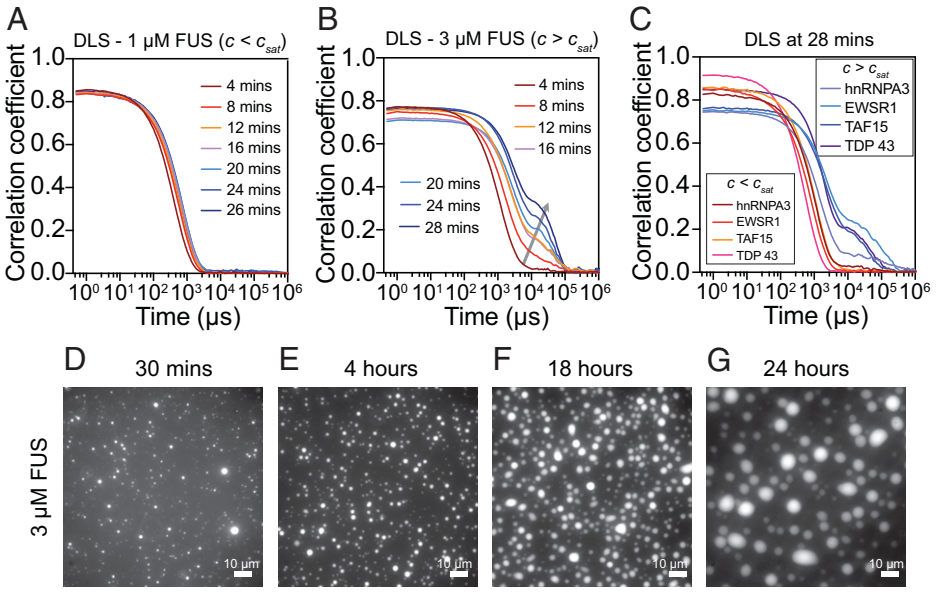

Fig. 8. Clusters grow into micron-scale bodies above c_{sat}. (*A*) Temporal evolution of autocorrelation functions from DLS measurements for untagged FUS. Below c_{sat}, the sizes of mesoscale clusters reach a steady state. (*B*) Above c_{sat}, we observe increased amplitudes of the autocorrelation function at longer times, (annotated using a gray arrow). (*C*) DLS data for hnRNPA3, EWSR1-SNAP, and TAF15-SNAP. Below c_{sat}, the cluster sizes reach a steady state. (*D*–*G*) Evidence of the growth into micrometer-scale condensates, displaying coarsening whereby fewer condensates grow by absorbing smaller species. This is made clear by the long-time evolution of micrometer-scale condensates formed by 3 μM untagged FUS containing 5% FUS-eGFP.

associations. The low abundance of mesoscale clusters quantified using NTA, the presence of smaller species, readily detected using fluorescence anisotropy, and the presence of a broad spectrum of species indicated by MFD, fluorescence intensity distribution analysis, and MCS suggest that the distributions of cluster sizes are concentration dependent and heavy tailed.

Growth of Macroscopic Phases above c_{sat}. In the physics of stochastic processes, it is known that random variables with broad distributions, specifically, heavy-tailed distributions, can be generators of condensation phenomena (42). Therefore, guided by precedents in the literature (43, 44), we used DLS to ask whether the heavy-tailed distributions of cluster sizes generate discernible signatures of transitions from finite-sized clusters to condensates just above c_{sat}. Specifically, we used DLS to probe the presence of slow modes (44) in the temporal evolution of autocorrelation functions for concentrations of untagged FUS that are below and above c_{sat}. Below c_{sat}, the autocorrelation functions reach a steady state and do not change with time after a few minutes (Fig. 8*A*). The timescales interrogated here are roughly seven orders of magnitude longer than the time it takes for individual FUS molecules to diffuse across 10 nm. Just above c_{sat} (Fig. 8*B*), the autocorrelation functions show the presence of slow modes. Such modes have been observed for polymers in dense phases, and have been attributed to reptation (44). However, we interpret our observation of slow modes to imply that clusters grow into micron-scale condensates above c_{sat}. The relevant data are shown for hnRNPA3, EWSR1-SNAP, and TAF15-SNAP (Fig. 8*C* and *SI Appendix*, Fig. S15). That the slow modes in autocorrelation functions point to the onset of condensation processes was independently verified using microscopy. These data show that condensation as a function of time leads to the formation of dense phases (Figs. 8*D*-8*G*). The observed temporal evolution is consistent with a coarsening process whereby larger condensates grow at the expense of smaller ones (45, 46). Our analysis also shows that the lowest concentration at which one observes the onset of slow modes in autocorrelation functions measured using DLS can be used as an efficient, centrifugation-free protocol for estimating c_{sat}. Further, these data suggest that c_{sat} in systems that feature heavy-tailed distributions of clusters in subsaturated solution may be viewed as the threshold concentration for the onset of coarsening through coalescence, a key signature of macroscopic phase separation (47).

Cluster Formation and Macroscopic Phase Separation Can Be Decoupled. Solutes such as 1,6-hexanediol (HD) can suppress phase separation and dissolve micron-scale condensates (30, 48). We used DLS to measure the impact of increased HD concentration on low-abundance mesoscale clusters vs. macroscopic phase separation. We observe a dose-dependent response of macroscopic phase separation on HD. Specifically, the presence of a dense phase, characterized by the appearance of slow modes in the autocorrelation functions, is weakened and abrogated at increased concentrations of HD (Fig. 9 *A*–*C*). However, profiles of autocorrelation functions, namely, the presence of fast modes that are consistent with the presence of clusters, persist even upon the addition of up to 1% wt/vol of HD. Similar results were obtained when we queried the effects of adenosine 5′-triphosphate (ATP), which is thought to be a condensate dissolving hydrotrope at the concentrations used here (49). Further, the distributions of clusters formed at c_{sat} (~2 μM) show minimal changes in the presence of HD and/or ATP (Fig. 9*D*).

We interpret the results of solute titrations to mean that there are at least two separable energy scales that contribute to condensate formation. Solutes such as HD and ATP primarily affect solvent quality, whereby the Flory χ parameter is altered to impact the overall solubility profiles of FUS molecules. This is consistent with the observations that polyols such as HD lower the macroscopic surface tension of water (50). Accordingly, we reason that cluster formation is likely to be influenced by distinct, sequence- or chemistry-specific interactions, that can be separable using solutes but are generally strongly coupled to driving forces for macroscopic phase separation.

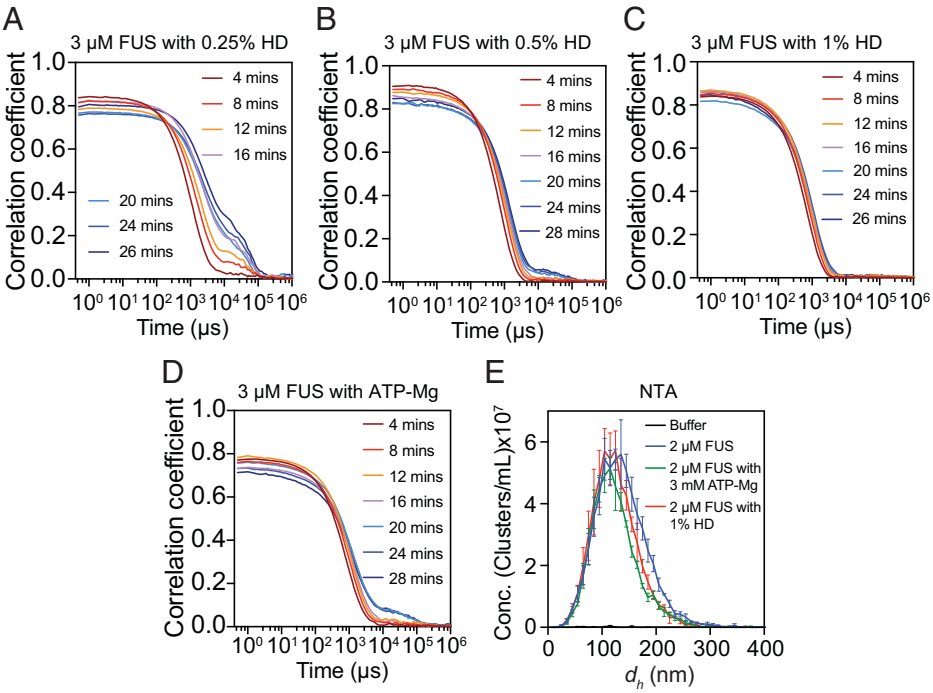

Fig. 9. Solutes dissolve condensates while having a minimal effect on clusters. For untagged FUS, above c_{sat}, there are slow modes in the autocorrelation functions; see *B*. (*A–C*) This feature is weakened in (*A*) 0.25% and (*B*) 0.5% and lost in (*C*) 1% 1,6 hexanediol (HD), which dissolves condensates for concentrations above c_{sat}. However, clusters that form via fast modes persist upon the addition of HD. (*D*) Similar observations were obtained in the presence of 3 mM ATP-Mg. (*E*) NTA data, collected at a concentration of 2 μM for untagged FUS, show that the solutes have discernible effects on the distribution of cluster sizes. Within statistical error of the measurements, the effects of solutes on clusters in subsaturated solutions are small when compared with the impact of solutes on the loss of slow modes above c_{sat}.

Impacts of Mutagenesis on Cluster Formation and Phase Separation. We used a combination of mutagenesis experiments and experiments based on changes to solution conditions to query the extent of coupling between interactions that drive cluster formation vs. macroscopic phase separation. We measured the sensitivity of cluster formation and phase separation of full-length, untagged FUS to changes in pH (*SI Appendix*, Fig. S16). These experiments show that increasing the net charge weakens cluster formation in subsaturated solutions (*SI Appendix*, Fig. S16 *A* and *B*). This leads to a clear upshift of the c_{sat} for phase separation *(SI Appendix*, Fig. S16*D*). Likewise, increasing the concentration of monovalent salts also weakens cluster formation (*SI Appendix*, Fig. S17). Therefore, increasing the net charge above a system-specific threshold or screening of electrostatic interactions weakens cluster formation and macroscopic phase separation.

To test the contributions of sequence-specific interactions to the coupling between cluster formation and macroscopic phase separation, we replaced 24 Arg residues in the RBD of full-length FUS with Gly residues. These mutations shift c_{sat} up by at least an order of magnitude (9) and concomitantly lower the abundance of mesoscale clusters by over an order of magnitude (Fig. 10*A* and *SI Appendix*, Fig. S18*A*). Likewise, substitution of aromatic Tyr residues with Ser in the PLD of full-length FUS abrogates phase separation in the low micromolar range while also lowering the abundance of mesoscale clusters by over an order of magnitude (Fig. 10*B* and *SI Appendix*, Fig. S18 *B–D*). These results highlight a strong coupling between clustering and macroscopic phase separation when Arg or aromatic residues are removed *(SI Appendix*, Figs. S19 and S20). In contrast, we find that substitution of 10 Asp and 4 Glu residues with Gly within the RBD of full-length FUS stabilizes cluster formation (Fig. 11*C* and *SI Appendix*, Fig. S18*E*). Specifically, at a concentration of 2 μM, the abundance of clusters increases by a factor of 1.8 when compared with wild-type FUS. However, the increase in abundance of clusters is not accompanied by evidence of condensate formation, even at a bulk concentration of 4 μM (*SI Appendix*, Fig. S21).

Next, we queried the impact of replacing six of the Phe residues within the RBD of full-length FUS with Gly (Fig. 10*D* and *SI Appendix*, Fig. S18*F*) vs. replacing six of the Tyr residues in the RBD with Ser (Fig. 10*E* and *SI Appendix*, Fig. S18*G*). Although both Phe and Tyr are aromatic residues, the effects of replacing these moieties are quantitatively different. At equivalent concentrations, the abundance of mesoscale clusters, detected using NTA, is at least threefold lower when Tyr residues are substituted with Ser when compared with substituting Phe residues with Gly. In both cases, macroscopic phase separation is also not observed for bulk concentrations up to 10 μM (*SI Appendix*, Fig. S22). The differences between Phe and Tyr are reminiscent of recent results for the PLD of hnRNP-A1 (51).

Impact of the Diversity of Chemistry-Specific Interactions on Cluster Formation and Phase Separation. The distribution of residue types is different across full-length FUS compared with the PLD and RBD alone (Fig. 11*A*). The measured c_{sat} values of PLDs are at least two orders of magnitude higher than those of full-length proteins (9, 52). We asked whether this change in c_{sat} was accompanied by changes in cluster formation. NTA measurements show that, at equivalent molar concentrations, the abundance of mesoscale clusters formed by the PLD of FUS is lower by a factor of 60 when compared with the abundance of such clusters formed by full-length FUS (Fig. 11*B* and *SI Appendix*, Fig. S23*A*). The DLS measurements show that macroscopic phase separation of the PLD was not observed for concentrations up to 150 μM (*SI Appendix*, Fig. S23*B*). The sequence of the RBD features a significant number of aromatic and anionic residues, and its measured c_{sat} is ~20 μM (*SI Appendix*, Fig. S23*C*). This is an order of magnitude lower than that of the PLD. NTA data show that, at a molar concentration of 1 μM, mesoscale clusters formed

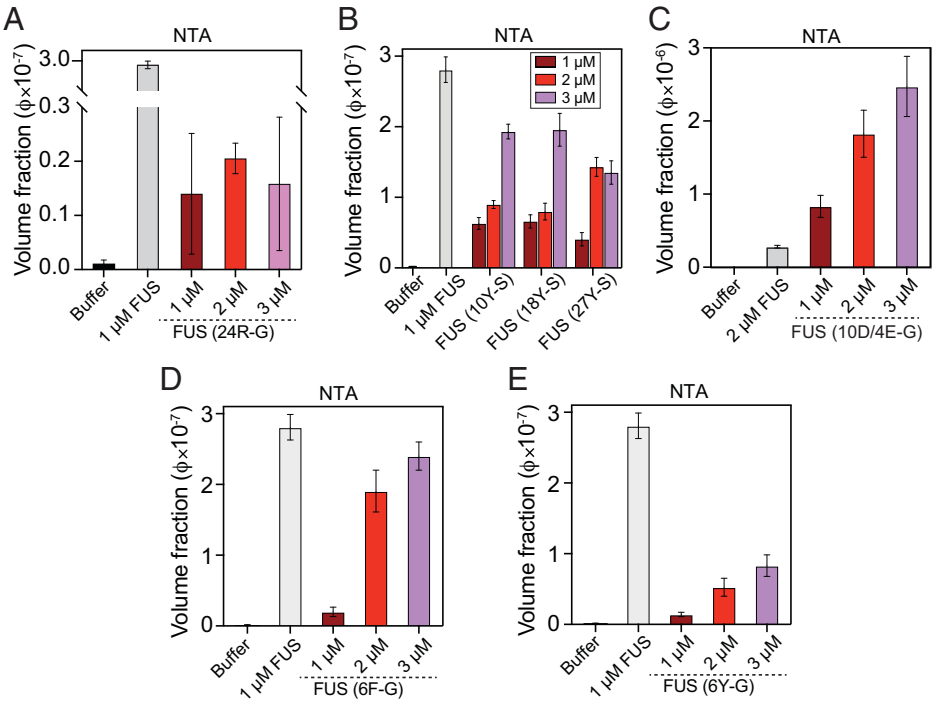

Fig. 10. Mutagenesis experiments reveal chemistry-specific effects of different residue types on the formation of mesoscale clusters in subsaturated solutions of FUS. (*A*) NTA data show that substituting 24 Arg residues with Gly within the RBD reduces the abundance of mesoscale clusters. (*B*) Replacing aromatic residues within the PLD of full-length FUS shows a valence-dependent reduction in the abundance of mesoscale clusters. (*C*) Replacing Asp/Glu with Gly in the RBD of full-length FUS enhances chemistry-specific interactions that stabilize clusters. (*D*) Substitution of six Phe residues with Gly in the RBD of full-length FUS lowers the abundance of clusters. (*E*) Likewise, substitution of six Tyr with Ser in the RBD of full-length FUS also lowers the abundance of clusters.

by the RBD (Fig. 11*C* and *SI Appendix*, Fig. S23*D*) are at least 10 times more abundant than clusters formed by the PLD. This highlights the contributions of sequence- or chemistry-specific interactions to cluster formation and phase separation.

Simulations Show That Systems Featuring At Least Two Energy Scales Behave Very Differently from Those with a Single Energy Scale. One way to introduce distinct energy scales into a polymeric system is via the framework of associative polymers (9, 23, 26–29, 53–57). These are macromolecules with attractive groups known as "stickers" that are interspersed by "spacers" (23, 25, 26, 55, 58). Stickers can form noncovalent reversible cross-links with one another (30, 51, 59, 60). The strengths of intersticker interactions are governed by the functional groups that define stickers. Therefore, in a stickers-and-spacers framework, intersticker interactions are specific interactions, whereas spacers define the intrinsic χ and contribute to the overall solubility profile of the polymer (30, 51, 59, 60). In recent computational work, FUS molecules were modeled as diblock polymers (61). Note that associative polymers represent a superset that encompasses specific instantiations such as diblock systems.

We used the LaSSI (**La**ttice **S**imulation **E**ngine for **S**ticker and **S**pacer **I**nteractions) engine (27), to model lattice polymers featuring a single energy scale vs. models for associative polymers with at least two energy scales. Our phenomenological models are intended to uncover possibilities that come from competing energy scales. First, we considered a model homopolymer where all units are identical (Fig. 12, *Left*). Here, all interactions are uniform and isotropic, and they are described by a single energy scale, ε. For the finite-sized homopolymer considered here, phase separation requires that ε be negative (χ must be positive), and the magnitude of ε must be greater than 0.1 k_BT. Results for ε values of -0.15 k_BT and -0.2 k_BT are shown in

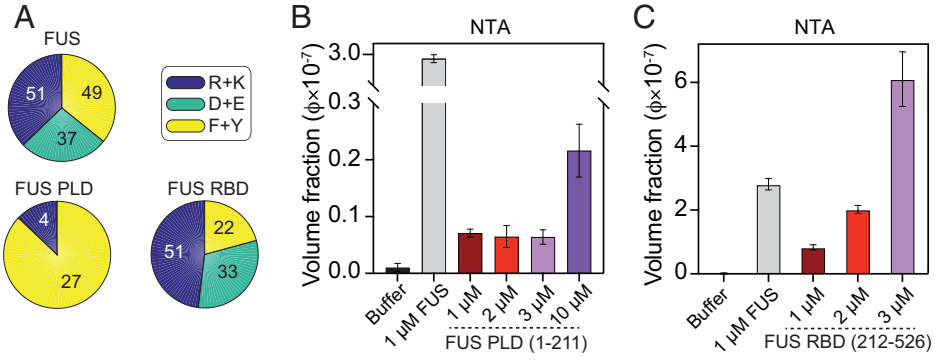

Fig. 11. Alterations to chemistry-specific interactions impact the abundance of clusters. (*A*) RBDs and PLDs of FUS feature differences in amino acid compositions. (*B*) NTA data show that the lower valence and absence of Arg-π interactions reduces the abundance of mesoscale clusters for the PLD of FUS by roughly two orders of magnitude when compared with full-length FUS. (*C*) In contrast, the RBD of FUS forms mesoscale clusters that are an order of magnitude more abundant than the PLD.

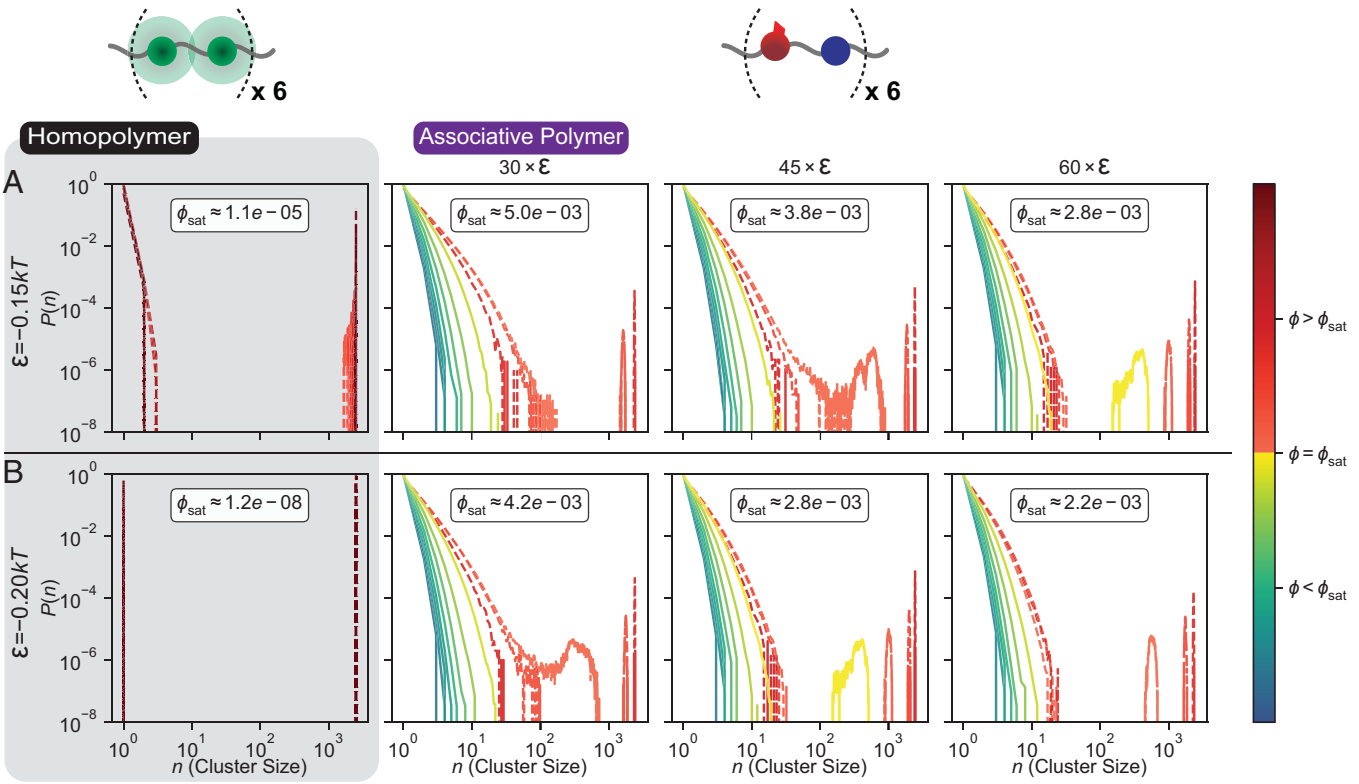

Fig. 12. Results from LaSSI simulations highlight the distinctions between homopolymers governed by a single energy scale vs. associative polymers with two energy scales. Each polymer has 12 beads. For homopolymers (results shown in *Left*), all beads are the same. For associative polymers, the red beads are stickers, and the blue beads are spacers. We quantify the probability $P(n)$ of observing clusters comprising n molecules. The data are presented in two rows (A) and (B), with four panels per row. In each panel, the dashed lines represent distributions computed from LaSSI simulations that were performed above the model-specific c_{sat}, and the solid lines represent distributions computed in subsaturated solutions. The colors of the dashed and solid lines represent the extent of supersaturation ($s > 0$ for $\phi > \phi_{sat}$) or subsaturation ($s < 0$ for $\phi < \phi_{sat}$), respectively. Results in *Left* are for the homopolymer described by a single energy scale ε, whereas results in *Left Center*, *Right Center*, and *Right* are for associative polymers featuring two energy scales, namely, the excluded volumes of spacers and the specific sticker–sticker interactions with anisotropic interactions worth $g\varepsilon$, where g = 30, 45, and 60, respectively.

Fig. 12, *Left*. Making ε stronger by a factor of 4/3, realized by lowering ε from −0.15 k_BT and −0.2 k_BT, decreases ϕ_{sat} by three orders of magnitude (Fig. 12, *Left*). This is because phase separation is a highly cooperative process when it is governed by a single energy scale (62). Above a threshold value, small changes to the energy scale ε will dramatically alter the overall solubility, and this is manifest as large changes to c_{sat}. The dense phases formed by homopolymers are essentially polymer melts with volume fractions of $\phi_{den} \approx 1$ (*SI Appendix*, Fig. S24).

The introduction of distinct energy scales alters the overall phase behavior. In our model for associative polymers, each bead is designated as being a sticker or a spacer (red vs. blue beads in Fig. 12). Stickers form specific anisotropic interactions with one another. For a given value of ε, the strengths of anisotropic inter-sticker interactions are set to be $g\varepsilon$, where g equals 30, 45, or 60 (Fig. 12, *Left Center*, *Right Center*, and *Right*). The spacers do not engage in attractive spacer–spacer or spacer–sticker interactions. Their only contribution is to the generic excluded volume. For values of ε where the homopolymers undergo phase sepa ration, we observe three distinct effects of introducing two different energy scales. First, the value of the saturation concentration ϕ_{sat} is renormalized to be larger than that of the equivalent homopolymer. Although ϕ_{sat} decreases with increasing g, the decrease is considerably smaller when compared with what is observed for the homopolymer when the magnitude of ε is increased by similar extents. Second, clusters form in subsaturated solutions. The size distributions of clusters are heavy tailed, and their quantitative features are sensitive to both g and ε. This highlights the coupling between the effects of specific sticker–sticker interactions and generic, solubility-determining interactions. Third, the interplay between stickers and spacers causes a dilution of polymer concentrations in the dense phase (*SI Appendix*, Fig. S24). Instead of a polymer melt, the dense phases are condensate-spanning networks defined by intersticker cross-links that are diluted by spacers. The overall thrust of our simulation results is concordant with results from simulations performed by Ranganathan and Shakhnovich (63) of systems with similar architectures.

Discussion

A simple way to describe the driving forces for phase separation is using the mean field Flory–Huggins formalism (10, 11). These simple theories have been the mainstay for quantitative descriptions and analysis of phase transitions driven by multivalent proteins and nucleic acids in vitro and in cells (12–15, 45, 47, 62, 64–68). In this formalism, there is only one effective energy scale as captured by the parameter χ (10). For a purely phase separation–based process (10, 11), subsaturated solutions should feature dispersed monomers and very few small clusters, if any, at any given time. In this work, we report results from our investigations of the types of species that form in subsaturated solutions of phase-separating RNA-binding proteins with disordered PLDs and RBDs. We find that subsaturated solutions do not conform to expectations based on systems that are characterized by a single energy scale such as the Flory χ-parameter. Therefore, the biophysical principles underlying biomolecular condensate formation and dissolution should not be described using simple, Flory-Huggins-stylemodels for macroscopic phase separation.

The presence of different energy scales is best illustrated by the response to solutes, in which clusters and condensates respond differently to HD and ATP. Phase separation is likely driven mainly by composition-specific interactions that determine the solubility profiles of FUS and FUS-like molecules. However, our data also show that the sequence features of FUS and FUS-like molecules can engender a strong coupling between the driving forces for cluster formation and phase separation. Mutations to stickers weaken cluster formation and increase c_{sat}. This indicates that the mechanism of phase separation is intimately tied to the structure of the underlying size distribution of clusters that form in subsaturated solutions. The picture that emerges is of networks of chemistry-specific, interresidue interactions (7, 9, 51, 52, 69) driving cluster formation and determining the extent of coupling between cluster formation and phase separation. It is important to emphasize that the clusters that we report here are not the prenucleation clusters that have been reported for various systems in supersaturated solutions (19–22). Nor are they micellar species or facsimiles of microphase separation. Instead, they are distinct species that appear to form via mechanisms that are paralogous to isodesmic reversible associations (36) that generate heavy-tailed distributions of cluster sizes.

Recently, Zhao et al. (70) showed that the N protein of the severe acute respiratory syndrome coronavirus 2 virus, which has been studied for its macroscopic phase separation (71, 72), also forms an assortment of clusters in subsaturated solutions. Likewise, Seim et al. (73) discovered that the interplay between specific homotypic and heterotypic interactions of the fungal protein Whi3 influences the distribution of protein cluster sizes that forms in dilute phases that coexist with Whi3-RNA granules. Mutations to the specific interactions that generate clusters through homotypic interactions can either enhance or diminish clustering, and this has a concomitant effect on the concentrations of Whi3 proteins in the dense phase (73). Both reports highlight the importance of dilute phase clusters that are either modulated by heterotypic interactions or influence phase behavior driven by heterotypic interactions.

Mesoscale clusters of low overall abundance have been reported for subsaturated solutions of folded proteins that form crystalline solids (74, 75). These clusters were found to have fixed size, irrespective of protein concentration (74), and are likely to be manifestations of microphase separation (31). In contrast, we observe continuous growth in the sizes and abundance of clusters with concentration. As discussed below, the framework of associative polymers provides a plausible explanation for our observations.

Associative polymers undergo two types of transitions namely, percolation without phase separation—also known as sol–gel transitions—or phase separation coupled to percolation (25–28, 56, 57). Above a system-specific threshold concentration known as the percolation threshold, or c_{perc} (27–29, 57), nonstoichiometric clusters become connected into a percolated or system-spanning network (27–29, 57). Quantitatively, c_{perc} is determined by the numbers (valence) of stickers, diversity of sticker types (29), and strengths of intersticker interactions (28, 29). If $c_{perc} < c_{sat}$, then associative polymers undergo percolation without phase separation. Alternatively, if $c_{sat} < c_{perc} < c_{den}$, where c_{den} is the concentration of the dense phase formed via phase separation (57, 76), then phase separation and percolation are coupled . In this scenario, the condensates form via PSCP and are akin to spherical microgels (77), whereby the percolated network is condensate spanning. The prepercolation clusters that form in subsaturated solutions are direct manifestations of clusters predicted to define sols formed by associative polymers undergoing either percolation without phase separation or phase separation coupled to percolation.

Precedents for Clusters in Subsaturated Solutions. The multivalent domain linker systems studied by Rosen and coworkers (57, 76, 78–81) are exemplars of associative polymers featuring an interplay between site-specific domain–motif interactions and linker-mediated solubility profiles. Li et al. (76) investigated the phase behaviors of systems where phase transitions are driven by multivalent interactions among poly-SH3 and poly-PRM (proline-rich module) molecules. Existence of a threshold concentration for phase transitions requires that there be at least three SH3 domains and three PRMs within the associating molecules. This is consistent with the presence of a valence-dependent percolation threshold (82, 83). Li et al. (76) showed that, as the valence (numbers) of SH3 domains and PRMs increases, the threshold concentration for phase transitions decreases. Importantly, Li et al. (76) also reported the presence of clusters in subsaturated solutions. These were detectable using DLS and small-angle X-ray scattering. The phase transitions were referred to by Li et al. as macroscopic liquid–liquid phase separation that is thermodynamically coupled to sol–gel transitions. This phrasing is synonymous with the process we describe here as phase separation coupled to percolation or PSCP.

Biological Relevance of Clusters That Form in Subsaturated Solutions. Our results lead us to propose that intrinsically disordered RNA-binding proteins from the FET family belong to the class of polymers known as associative polymers. Our findings show that chemistry- or site-specific sticker-mediated interactions and solubility-determining spacer-mediated interactions are separable in systems with sticker and spacer architectures (28, 30, 51, 52, 56, 57, 84–90). Prepercolation clusters, with size distributions featuring heavy tails, are likely to be present at endogenous concentrations that tend to be in the nanomolar to low micromolar range in live cells (91). The clusters we observe are likely to be the bridge that connects to recent findings regarding dynamic clusters comprising multivalent molecules (92, 93). Importantly, the recent work of Cho et al. (94) identified cytoplasmic clusters formed by several RNA-binding proteins under endogeneous, unstressed conditions. These clusters are distinct from cytoplasmic stress granules, which are bona fide condensates that form in response to stress through the process of phase separation coupled to percolation (5–7). Now that we know what to look for, we expect there to be many more accounts of clusters, precursors of condensates, forming at endogenous levels in live cells. The clusters of interest will lack a phase boundary and will form via reversible associations, and the distributions of their sizes and shapes will be determined by a complex network of homotypic and heterotypic interactions.

What are the potential consequences of clusters defined by heavy-tailed size distributions forming in subsaturated solutions in cells? Process control via biomolecular condensates requires that phase separation be robust and reproducible (3). In a classical phase separation system, the barrier to nucleation will determine the response time. Clusters in subsaturated solutions will likely speed up the response time by lowering the barrier for phase separation. Accordingly, we postulate that clusters poise RNA-binding proteins for robust condensate formation in response to stimuli. Our discovery of the presence of clusters following heavy-tailed size distributions in subsaturated solutions suggests that regulation can also occur at the level of shaping the distribution of cluster sizes. This can happen through a

variety of heterotypic interactions, specifically with RNA molecules. Additionally, chaperones are known to modulate size distributions of self-associating molecules (12), and one could imagine chaperones acting to reshape the size distributions of clusters that form in subsaturated solution and in dilute phases that coexist with condensates. Additionally, deleterious interactions with cellular components that lead to disease (95), and the dynamical arrest of condensate growth (63), could also occur at the level of clusters in dilute phases and / or subsaturated solutions.

Materials and Methods

All proteins were first expressed in 1 L SF9 (1 million/mL) insect cells with 5 mL of P2 virus and harvested 72 h postinfection. We used a centrifugation approach to determine saturation concentrations. The samples at the indicated protein concentration were incubated for 30 min, the solutions were spun down at 20,000 RCF on a benchtop centrifuge unit, and the concentration was measured in the clarified supernatant by using a Bradford assay. Details of all the constructs used in the experiments, the protein expression and purification protocols that distinguish methods A and B, the setup and analysis of DLS, NTA, MFD, MCS, and TEM experiments, raw data from DLS measurements, and the LaSSI simulations are described in *SI Appendix*.

Data Availability. All study data are provided explicitly in the article and/or Supporting Information Appendix. All raw data for the main text figures are available for free download at: https://zenodo.org/record/6652106#.YqseeuxOnAw (96).

ACKNOWLEDGMENTS. We thank Mina Farag, Tyler Harmon, Adam Klosin, Alex Holehouse, Stephen Michnick, Tanja Mittag, Doayuan Qian, Kiersten Ruff, Vijay Rangachari, Samuel Safran, Peter Schuck, Jie Wang, and members of the A.A.H. and R.V.P. laboratories for helpful discussions. This work was funded by a direct grant from the Max Planck Society (to A.A.H.), a grant from the NOMIS Foundation (to A.A.H), the Wellcome Trust (209194/Z/17/Z to A.A.H.), the European Research Council (ERC) through ERC Grant PhysProt (to T.P.J.K, Agreement 337969), the Wellcome Trust and the Frances and Augustus Newman Foundation (to T.P.J.K.), priority program SPP2191 from the Deutsche Forschungsgemeinschaf (to S.A. and C.A.M.S), the US NIH (grants 5R01NS1056114 and R01NS121114 to R.V.P), and the St. Jude Children's Research Hospital collaborative research consortium on membraneless organelles (to R.V.P). We are grateful for technical support provided by Régis Lemaitre and Barbara Borgonovo of the protein expression, purification, and characterization facility of the Max Planck Institute for Cell Biology and Genetics (MPI-CBG), as well as Michaela Wilsch-Bräuninger and Jana Mesenser of the Transmission Electron Microscopy facility at MPI-CBG. We thank Andrei Pozniakovsky for DNA constructs of all proteins.

Author affiliations: [a]Max Planck Institute of Cell Biology and Genetics, 01307 Dresden, Germany; [b]Department of Physics, Washington University in St. Louis, St. Louis, MO 63130; [c]Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, CB2 1EW Cambridge, United Kingdom; [d]Department of Molecular Physical Chemistry, Heinrich Heine University, 40225 Düsseldorf, Germany; [e]Biotechnology Center, Center for Molecular and Cellular Bioengineering, Technische Universität Dresden, 01069 Dresden, Germany; [f]Cavendish Laboratory, University of Cambridge, CB3 0HE Cambridge, United Kingdom; and [g]Department of Biomedical Engineering, Center for Science & Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO 63130

Author contributions: Conceptualization: M.K., F.D., A.A.H, and R.V.P. Fig. 1 M.K., and A.M., Figs. 2-4 M.K., Fig. 5 L.V., R.K., C.A.M.S., T.J.W., G.K., and T.P.J.K., Figs. 6-11 M.K., and Fig. 12 F.D., and R.V.P. Overall analysis: M.K., F.D., A.A.H., T.P.J.K., R.V.P. Reagents: M.K., T.M.F., and S.A. Writing: R.V.P. Editing and revising: M.K., F.D., C.A.M.S., T.P.J.K., A.A.H., and R.V.P. *SI Appendix*: M.K., F.D., A.M., T.M.F., and R.V.P. Funding acquisition: S.A., C.A.M.S., T.P.J.K., A.A.H., and R.V.P.

1. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
2. 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).
3. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
4. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382 (2017).
5. D. W. Sanders *et al.*, Competing protein-RNA interaction networks control multiphase intracellular organization. *Cell* **181**, 306–324.e28 (2020).
6. P. Yang *et al.*, G3BP1 is a tunable switch that triggers phase separation to assemble stress granules. *Cell* **181**, 325–345.e28 (2020).
7. J. Guillén-Boixet *et al.*, RNA-induced conformational switching and clustering of G3BP drive stress granule assembly by condensation. *Cell* **181**, 346–361.e17 (2020).
8. S. Alberti, A. Gladfelter, T. Mittag, Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419–434 (2019).
9. J. Wang *et al.*, A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.e616 (2018).
10. P. J. Flory, Thermodynamics of high polymer solutions. *J. Chem. Phys.* **10**, 51–61 (1942).
11. M. L. Huggins, Solutions of long chain compounds. *J. Chem. Phys.* **9**, 440 (1941).
12. A. Narayanan *et al.*, A first order phase transition mechanism underlies protein aggregation in mammalian cells. *eLife* **8**, e39695 (2019).
13. S. F. Shimobayashi, P. Ronceray, D. W. Sanders, M. P. Haataja, C. P. Brangwynne, Nucleation landscape of biomolecular condensates. *Nature* **599**, 503–506 (2021).
14. E. W. Martin *et al.*, A multi-step nucleation process determines the kinetics of prion-like domain phase separation. *Nat. Commun.* **12**, 4513 (2021).
15. A. E. Posey *et al.*, Mechanistic inferences from analysis of measurements of protein phase transitions in live cells. *J. Mol. Biol.* **433**, 166848 (2021).
16. D. W. Oxtoby, Homogeneous nucleation: Theory and experiment. *J. Phys. Condens. Matter* **4**, 7627–7650 (1992).
17. D. W. Oxtoby, D. Kashchiev, A general relation between the nucleation work and the size of the nucleus in multicomponent nucleation. *J. Chem. Phys.* **100**, 7665–7671 (1994).
18. D. Turnbull, Kinetics of heterogeneous nucleation. *J. Chem. Phys.* **18**, 198–203 (1950).
19. P. G. Vekilov, The two-step mechanism of nucleation of crystals in solution. *Nanoscale* **2**, 2346–2357 (2010).
20. A. Vitalis, R. V. Pappu, Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides. *Biophys. Chem.* **159**, 14–23 (2011).
21. O. Gliko *et al.*, A metastable prerequisite for the growth of lumazine synthase crystals. *J. Am. Chem. Soc.* **127**, 3433–3438 (2005).
22. D. S. Yang *et al.*, Mesoscopic protein-rich clusters host the nucleation of mutant p53 amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015618118 (2021).
23. F. Tanaka, "Theory of molecular association and thermoreversible gelation" in *Molecular Gels: Materials with Self-Assembled Fibrillar Networks*, R. G. Weiss, P. Terech, Eds. (Springer, Dordrecht, The Netherlands, 2006), pp. 17–78.
24. F. Tanaka, *Polymer Physics: Applications to Molecular Association and Thermoreversible Gelation* (Cambridge University Press, Cambridge, United Kingdom, 2011).
25. M. Rubinstein, A. V. Dobrynin, Solutions of associative polymers. *Trends Polym. Sci. (Regul. Ed.)* **5**, 181–186 (1997).
26. A. N. Semenov, M. Rubinstein, Thermoreversible gelation in solutions of associative polymers. 1. Statics. *Macromolecules* **31**, 1373–1385 (1998).
27. J.-M. Choi, F. Dar, R. V. Pappu, LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Comput. Biol.* **15**, e1007028 (2019).
28. J.-M. Choi, A. S. Holehouse, R. V. Pappu, Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
29. J.-M. Choi, A. A. Hyman, R. V. Pappu, Generalized models for bond percolation transitions of associative polymers. *Phys. Rev. E* **102**, 042403 (2020).
30. S. Frey, R. P. Richter, D. Görlich, FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815–817 (2006).
31. L. Leibler, Theory of microphase separation in block copolymers. *Macromolecules* **13**, 1602–1617 (1980).
32. C. Tanford, Thermodynamics of micelle formation: Prediction of micelle size and size distribution. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1811–1815 (1974).
33. S. Bhattacharjee, DLS and zeta potential - What they are and what they are not? *J. Control. Release* **235**, 337–351 (2016).
34. R. Jiang, D. N. P. Murthy, A study of Weibull shape parameter: Properties and significance. *Reliab. Eng. Syst. Saf.* **96**, 1619–1626 (2011).
35. T. F. A. De Greef *et al.*, Supramolecular polymerization. *Chem. Rev.* **109**, 5687–5754 (2009).
36. M. R. Marzahn *et al.*, Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. *EMBO J.* **35**, 1254–1275 (2016).
37. V. Filipe, A. Hawe, W. Jiskoot, Critical evaluation of nanoparticle tracking analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. *Pharm. Res.* **27**, 796–810 (2010).
38. J. Widengren *et al.*, Single-molecule detection and identification of multiple species by multiparameter fluorescence detection. *Anal. Chem.* **78**, 2039–2050 (2006).
39. P. Arosio *et al.*, Microfluidic diffusion analysis of the sizes and interactions of proteins under native solution conditions. *ACS Nano* **10**, 333–341 (2016).
40. W. E. Arter *et al.*, Rapid generation of protein condensate phase diagrams using combinatorial droplet microfluidics. bioRxiv [Preprint] (2020). https://doi.org/10.1101/2020.06.04.132308. Accessed Jun 05, 2020.
41. G. Krainer *et al.*, Direct digital sensing of proteins in solution through single-molecule optofluidics. bioRxiv [Preprint] (2020). https://doi.org/10.1101/2020.05.24.113498. Accessed May 28, 2020
42. G. Gradenigo, E. Bertin, Participation ratio for constraint-driven condensation with superextensive mass. *Entropy (Basel)* **19**, 517 (2017).
43. P. Malo de Molina, M.-S. Appavou, M. Gradzielski, Oil-in-water microemulsion droplets of TDMAO/decane interconnected by the telechelic C18-EO150-C18: Clustering and network formation. *Soft Matter* **10**, 5072–5084 (2014).
44. J. Li, T. Ngai, C. Wu, The slow relaxation mode: From solutions to gel networks. *Polym. J.* **42**, 609–625 (2010).
45. J. Berry, S. C. Weber, N. Vaidya, M. Haataja, C. P. Brangwynne, RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5237–E5245 (2015).
46. A. Baldan, Review Progress in Ostwald ripening theories and their applications to nickel-base superalloys Part I: Ostwald ripening theories. *J. Mater. Sci.* **37**, 2171–2202 (2002).

47. J. Berry, C. P. Brangwynne, M. Haataja, Physical principles of intracellular organization via active and passive phase transitions. *Rep. Prog. Phys.* **81**, 046601 (2018).
48. S. Alberti *et al.*, A user's guide for phase separation assays with purified proteins. *J. Mol. Biol.* **430**, 4806–4820 (2018).
49. A. Patel *et al.*, ATP as a biological hydrotrope. *Science* **356**, 753–756 (2017).
50. C. M. Romero, M. S. Páez, J. A. Miranda, D. J. Hernández, L. E. Oviedo, Effect of temperature on the surface tension of diluted aqueous solutions of 1,2-hexanediol, 1,5-hexanediol, 1,6-hexanediol and 2,5-hexanediol. *Fluid Phase Equilib.* **258**, 67–72 (2007).
51. A. Bremer *et al.*, Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* **14**, 196–207 (2021).
52. E. W. Martin *et al.*, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
53. M. E. Cates, T. A. Witten, Chain conformation and solubility of associating polymers. *Macromolecules* **19**, 732–739 (1986).
54. R. D. Groot, W. G. M. Agterof, Monte Carlo study of associative polymer networks. I. Equation of state. *J. Chem. Phys.* **100**, 1649–1656 (1994).
55. F. Tanaka, Theoretical study of molecular association and thermoreversible gelation in polymers. *Polym. J.* **34**, 479–509 (2002).
56. T. S. Harmon, A. S. Holehouse, R. V. Pappu, Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. *New J. Phys.* **20**, 045002 (2018).
57. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).
58. M. A. Winnik, A. Yekta, Associative polymers in aqueous solution. *Curr. Opin. Colloid Interface Sci.* **2**, 424–436 (1997).
59. T. J. Nott *et al.*, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
60. Y.-H. Lin, J. D. Forman-Kay, H. S. Chan, Sequence-specific polyampholyte phase separation in membraneless organelles. *Phys. Rev. Lett.* **117**, 178101 (2016).
61. S. Ranganathan, E. Shakhnovich, The physics of liquid-to-solid transitions in multi-domain protein condensates. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2021.11.15.468745. Accessed April 21, 2022.
62. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
63. S. Ranganathan, E. I. Shakhnovich, Dynamic metastable long-living droplets formed by sticker-spacer proteins. *eLife* **9**, e56159 (2020).
64. 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).
65. A. Klosin *et al.*, Phase separation provides a mechanism to reduce noise in cells. *Science* **367**, 464–468 (2020).
66. J. R. Simon, N. J. Carroll, M. Rubinstein, A. Chilkoti, G. P. López, Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nat. Chem.* **9**, 509–515 (2017).
67. J. A. Riback *et al.*, Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209–214 (2020).
68. A. W. Fritsch *et al.*, Local thermodynamics govern formation and dissolution of *Caenorhabditis* elegans P granule condensates. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2102772118 (2021).
69. D. Deviri, S. A. Safran, Equilibrium size distribution and phase separation of multivalent, molecular assemblies in dilute solution. *Soft Matter* **16**, 5458–5469 (2020).
70. H. Zhao *et al.*, Energetic and structural features of SARS-CoV-2 N-protein co-assemblies with nucleic acids. *iScience* **24**, 102523 (2021).
71. J. Cubuk *et al.*, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **12**, 1936 (2021).
72. C. Iserman *et al.*, Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol. Cell* **80**, 1078–1091.e6 (2020).
73. I. Seim *et al.*, Dilute phase oligomerization can oppose phase separation and modulate material properties of a ribonucleoprotein condensate. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2120799119 (2022).
74. O. Gliko *et al.*, Metastable liquid clusters in super- and undersaturated protein solutions. *J. Phys. Chem. B* **111**, 3106–3114 (2007).
75. M. A. Vorontsova, H. Y. Chan, V. Lubchenko, P. G. Vekilov, Lack of dependence of the sizes of the mesoscopic protein clusters on electrostatics. *Biophys. J.* **109**, 1959–1968 (2015).
76. P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
77. T. A. Vilgis, "Polymer networks" in *Comprehensive Polymer Science and Supplements*, G. Allen, J. C. Bevington, Eds. (Pergamon, Amsterdam, 1989), pp. 227–279.
78. S. Banjade, M. K. Rosen, Phase transitions of multivalent proteins can promote clustering of membrane receptors. *eLife* **3**, e04123 (2014).
79. S. F. Banani *et al.*, Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
80. X. Su *et al.*, Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **352**, 595–599 (2016).
81. L. B. Case, X. Zhang, J. A. Ditlev, M. K. Rosen, Stoichiometry controls activity of phase-separated clusters of actin signaling proteins. *Science* **363**, 1093–1097 (2019).
82. W. H. Stockmayer, Theory of molecular size distribution and gel formation in branched-chain polymers. *J. Chem. Phys.* **11**, 45–55 (1943).
83. P. J. Flory, Molecular size distribution in three dimensional polymers. I. Gelation1. *J. Am. Chem. Soc.* **63**, 3083–3090 (1941).
84. K. M. Ruff, F. Dar, R. V. Pappu, Ligand effects on phase separation of multivalent macromolecules. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2017184118 (2021).
85. X. Zeng, A. S. Holehouse, A. Chilkoti, T. Mittag, R. V. Pappu, Connecting coil-to-globule transitions to full phase diagrams for intrinsically disordered proteins. *Biophys. J.* **119**, 402–418 (2020).
86. J. A. Greig *et al.*, Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. *Mol. Cell* **77**, 1237–1250.e4 (2020).
87. I. Alshareedah, M. M. Moosa, M. Pham, D. A. Potoyan, P. R. Banerjee, Programmable viscoelasticity in protein-RNA condensates with disordered sticker-spacer polypeptides. *Nat. Commun.* **12**, 6620 (2021).
88. S. Roberts *et al.*, Injectable tissue integrating networks from recombinant polypeptides with tunable order. *Nat. Mater.* **17**, 1154–1163 (2018).
89. K. Lasker *et al.*, A modular platform for engineering function of natural and synthetic biomolecular condensates. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.02.03.429226. Accessed February 03, 2021.
90. A. S. Holehouse, G. M. Ginell, D. Griffith, E. Böke, Clustering of aromatic residues in prion-like domains can tune the formation, state, and organization of biomolecular condensates. *Biochemistry* **60**, 3566–3581 (2021).
91. M. Y. Hein *et al.*, A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
92. S. Chong *et al.*, Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
93. D. T. McSwiggen, M. Mir, X. Darzacq, R. Tjian, Evaluating phase separation in live cells: Diagnosis, caveats, and functional consequences. *Genes Dev.* **33**, 1619–1634 (2019).
94. N. H. Cho *et al.*, OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983 (2022).
95. C. Mathieu, R. V. Pappu, J. P. Taylor, Beyond aggregation: Pathological phase transitions in neurodegenerative disease. *Science* **370**, 56–60 (2020).
96. M. Kar *et al.*, Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. Zenodo. https://zenodo.org/record/6652106#.YqseeuxOnAw. Deposited on 16 June 2022.