A scale-invariant log-normal droplet size distribution below the critical concentration for protein phase separation

Tommaso Amico[1#], Andrea Lazzari[1#], Antonio Trovato[1#],
Michele Vendruscolo[2*], Monika Fuxreiter[1,3 *], Amos Maritan[1*]

[1]*Department of Physics and Astronomy, University of Padova, Italy*
[2]*Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, UK*
[3]*Department of Biomedical Sciences, University of Padova, Italy*

* Correspondence: mv245@cam.ac.uk, monika.fuxreiter@unipd.it, amos.maritan@unipd.it
[#] Equal contributions

Abstract

Many proteins undergo a process of phase separation that leads to the formation of biomolecular condensates. It has been observed that some of these proteins form dense droplets of sizeable dimensions already below their critical concentration for phase separation. To understand this phenomenon, which is not readily compatible with classical nucleation theory, we investigated the properties of the droplet size distributions as a function of protein concentration. We found that these distributions can be described by a scale-invariant log-normal function with an average that increases progressively as the concentration approaches the critical value from below. These results suggest the existence of a universal behaviour independent from the sequences and structures of the proteins undergoing phase separation, which is typically observed for second-order transitions. By determining the parameters of the scale-invariant distribution, we show that it is possible to use it to estimate the critical concentration for phase separation.

Introduction

Many proteins have been shown to undergo a phase separation process into a liquid-like condensed state (1-5). This process appears to be of physiological significance since it may lead to the formation of biomolecular condensates (3-7). As a consequence, it is closely controlled by the protein homeostasis system (8, 9), and its dysregulation has been associated with a broad range of human diseases (10, 11). It is therefore important to understand the fundamental nature of this process (5, 12, 13), to provide insights for the identification of ways of modulating it through pharmacological interventions.

Several models have been proposed to explain the phenomenon of protein phase separation. A commonly adopted theoretical framework is the Flory-Huggins theory of phase separation (14, 15). In its simplest form, this theory describes a first-order process of phase separation in a system of homopolymers, which later was adapted to polyampholites (16). The Flory-Huggins theory has been extended to associative polymers, with the aim of modelling the

sequence composition of proteins, by Flory and Stockmayer, who described the phase separation in terms of a third-order gelation process (17, 18). Semenov and Rubinstein modified the Flory-Stockmayer theory, reporting that the gelation process is not a phase transition, as all the derivatives of the free energy are analytical at the gelation point (19). It was also argued that another generalization of the Flory-Stockmayer theory (20) could describe phase transitions in cytoskeletal networks (21). More recently, it has been suggested that the phase separation process is coupled with percolation (22). In an older study, γ-crystallin was reported to undergo a second-order phase separation process (23), consistently with observations in lysozyme solutions (24).

To understand better the nature of the transition underlying protein phase separation on the basis of recent experimental observations (25, 26), here we study the distribution of the size of the droplets below the critical concentration. This question appears promising as a starting point to develop new insights since it has been reported that proteins can form droplets of sizeable dimensions already well below the critical concentrations (25, 26). This behaviour is not predicted by classical nucleation theory (27), and not clearly linked with the idea that protein phase separation can be described as a first-order transition. This is because in a first-order phase transition, nucleation takes place in a supersaturated system (28), while in a subsaturated system particles can still self-assemble, but with a probability that decreases exponentially with the size of the assemblies.

In the present study, we find that the distribution of the sizes of the droplets of the protein FUS follows a scale-invariant log-normal distribution (26). These findings indicate the presence of a universal behaviour compatible with a second-order transition. As an application of this observation, we address the challenge of estimating the critical concentration of phase separation. Without knowing this critical concentration, it is difficult to establish where proteins are located in their phase diagram, how close they are to the phase boundary between the native and droplet states, and how pharmacological interventions can alter their phase behaviour. Based on the properties of the scale-invariant log-normal distribution of droplet sizes, we find a correlation between the moments of the distribution and the distance from the phase boundary. Our results thus indicate that the moments of the distribution can be used as a scale-invariant gauge to estimate the phase boundary of proteins.

Results

Derivation of the scaling ansatz. Empirical evidence indicates that protein self-assembly into liquid-like condensates is characterized by: (i) a phase separation transition at a critical concentration ρ_c, (ii) a formation of droplets of sizeable dimensions already below ρ_c, and (iii) a droplet size distribution that, after an initial transient, does not change with the time interval of the experiment, although individual droplets can form, grow, shrink and dissolve. An initial analysis of droplet size distributions observed experimentally led us to ask whether the transition could be described in terms of a scaling theory, as commonly done for critical phenomena (29). We also note that this approach is analogous to analyzing the cluster size distribution in a percolation problem (30).

In our analysis, we called $P_>(s|\rho)$ the survival distribution function (SDF = 1-CDF, where CDF is the cumulative distribution function) corresponding to the probability to observe a droplet of

size greater than s, when the concentration is ρ. The distance from the critical concentration ρ_c is measured in terms of the dimensionless variable $\tilde{\rho} = (\rho - \rho_c)/\rho_c$, which also allows to compare data from different experiments, as explained below. In the vicinity of ρ_c, $|\tilde{\rho}|$ is small and according to the scaling hypothesis one can assume that the only quantity that matters in determining the form of $P_>(s|\rho)$ is the characteristic droplet size, s_c. Furthermore, s_c should have a singular behaviour

$$s_c = a\left|\frac{\rho-\rho_c}{\rho_c}\right|^{-\varphi} = a|\tilde{\rho}|^{-\varphi} \tag{1}$$

where $\varphi > 0$ is a critical exponent and a is a constant. In principle, $P_>(s|\rho)$ could depend separately on both s and ρ as well as on the particular system and the preparation conditions, but assuming that only s_c matters, we obtain the scaling ansatz

$$P_>(s|\rho) = s^{-\alpha} \cdot f(s/s_c), \tag{2}$$

where $\alpha \geq 0$ is another critical exponent. Thus, the scaling of the SDF is equivalent to saying that, apart from the singular behaviour $s^{-\alpha}$, the remaining s and ρ dependence occurs only through the ratio s/s_c. All extra dependencies are encapsulated in s_c and, possibly, on the specific form of the scaling function f. A consequence of scaling and singular behavior described by Eq. 1, i.e. the divergence of the characteristic size of the droplets, is that details of any specific system may not affect the value of the critical exponents, which are therefore expected to be universal.

To determine the exponents α and φ, we introduce the moments of $P_>(s|\rho)$

$$\langle s^k \rangle = \int_0^\infty s^k \cdot \left(-\frac{dP_>(s|\rho)}{ds}\right) ds = c_k \cdot s_c^{k-\alpha} \tag{3}$$

where c_k is a proportionality constant relative to the k-th moment, independent of $\tilde{\rho}$. From Eq. 3, we deduce that

$$\frac{\langle s^{k+1} \rangle}{\langle s^k \rangle} = \frac{c_{k+1}}{c_k} \cdot s_c = \frac{c_{k+1}}{c_k} \cdot a|\tilde{\rho}|^{-\varphi} \tag{4}$$

If the scaling ansatz of Eq. 2 is correct, by plotting the ratio of moments $\langle s^{k+1} \rangle / \langle s^k \rangle$ for various values of k as a function of $1/|\tilde{\rho}|$ in a *log-log* scale, we should obtain straight parallel lines with slope φ and intercept $-ln(ac_{k+1}/c_k)$.

Scaling behaviour of the droplet size distributions. We investigated the validity the ansatz in Eq. 2 using experimental data on the RNA-binding protein FUS, which are available for both the untagged and the SNAP-tagged protein (26). We calculated the survival distribution function (Eq. 2) and its moments (Eq. 3). We then plotted the moments versus the inverse distance from the critical concentration $1/|\tilde{\rho}|$ in log-log scale (**Figure 1**). We used the estimates for the critical concentration, ρ_c = 5.1 μM for FUS and ρ_c = 6.6 μM for SNAP FUS, obtained later, in a self-consistency check of the validity of the scaling ansatz. We observed that the

moment ratios at different distances from the critical concentration fall onto straight lines, as predicted by the scaling ansatz. In addition, the average slope of the lines for different moment ratios is 1.02 ± 0.01 for untagged FUS, and 1.07 ± 0.04 for SNAP FUS, which is consistent with φ=1 for the exponent in the scaling ansatz in Eq. 1 (**Figure 1**, left panel).

Having determined the *φ*-exponent, we can also determine the exponent *α* using Eq. 3

$$\left\langle s^k \right\rangle = c_k \cdot s_c^{k-\alpha} \quad \propto \quad \left| \frac{\rho - \rho_c}{\rho_c} \right|^{-\varphi \cdot (k-\alpha)} \tag{5}$$

The exponent *α* is then calculated from

$$m = \varphi(k-\alpha) \rightarrow \alpha = k - \frac{m}{\varphi} \tag{6}$$

where *m* is the slope of the linear fit of the log-log plot of <s> (the first moment, k=1) versus $1/|\widetilde{\rho}|$. We then plotted the mean droplet size, <s>, versus the distance from the critical concentration $|\widetilde{\rho}|$ on a natural logarithm scale, which could be fitted using a line with a slope m= 1.00 ± 0.02 for untagged FUS, and 1.04 ± 0.06 for SNAP FUS (**Figure 1**, right panel), which is consistent with m=1. Using the value of φ = 1, determined based on Eq. 4, we obtain α = 0.

Taken together, these data support the validity of the scaling ansatz, which suggests the existence of a second-order phase transition at some critical concentration.

The droplet size distribution is log-normal. The above analysis suggests that the droplet size distribution may follow a log-normal distribution. The scaling ansatz of Eq. 2 for the SDF is equivalent to the following scaling for the probability density distribution

$$P(s|\rho) = -\frac{dP_>(s|\rho)}{ds} = s^{-1-\alpha} \cdot F(s/s_c), \quad s_c \ \propto \ |\tilde{\rho}|^{-\varphi} \tag{7}$$

where *f,* the scaling function in Eq. 2, and *F* are related as follows

$$f(z) = z^{\alpha} \int_z^{\infty} x^{-\alpha-1} F(x) \mathrm{d}x \tag{8}$$

The log-normal droplet size distribution *P*(*s*|*ρ*) should be

$$\begin{aligned} P(s|\rho) &= s^{-1} \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{ -\frac{\left(lns - lns_0\right)^2}{2\sigma^2} \right\} = \\ &= s^{-1} \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{ -\frac{\left(ln(s/s_0)\right)^2}{2\sigma^2} \right\}. \end{aligned} \tag{9}$$

Consequently, the size survival distribution function should be

$$P_{>}(s|\rho) = \frac{1}{2} \cdot erfc\left(\frac{\ln(s/s_0)}{\sigma \cdot \sqrt{2}}\right) \tag{10}$$

where ln s_0 and σ can be determined as the average and the variance, of ln s obtained at each concentration:

$$\begin{aligned} \langle \ln s \rangle &= \ln s_0 \\ Var(s) &= \left\langle \left(\ln s\right)^2 \right\rangle - \langle \ln s \rangle^2 = \sigma^2 \end{aligned} \tag{11}$$

The droplet sizes follow a log-normal distribution only if the survival distribution functions or, equivalently, the size distribution functions collapse when plotted versus Eq. 9 or Eq. 10 with characteristic s_0 and σ values of each droplet size distributions obtained at different concentrations (Eq. 11). We determined ln s_0 and σ values for each distribution (**Figure 2**, right panels) and plotted the properly rescaled size distribution functions versus $s\left(1-\frac{\rho}{\rho_c}\right)$ (**Figure 2**, left panel). We observed that the size distribution functions collapsed for both FUS and SNAP FUS (**Figure 2**, left panel). Furthermore, the collapsed curve overlapped with the analytic log-normal distribution we computed with s_0=1, σ=1, the normal distribution in the rescaled variables.

The observed collapse confirms that droplet size distributions from different experiments follows a log-normal behaviour.

Independence of the variance of the distribution from the concentration. The results about the scaling and log-normal behaviour described above are consistent if the variance of the log-normal distribution, σ, is independent of $\tilde{\rho}$ (31). This is evident from Eqs. 9 and 10 if we set $s_c = s_0$. The k-th moment of the log-normal distribution is

$$\left\langle s^k \right\rangle = s_0^k e^{k^2\sigma^2/2}, \tag{12}$$

which compared with the scaling prediction Eq. 3 leads to $\alpha = 0$, appropriate for the log-normal, and

$$c_k = e^{k^2\sigma^2/2} \tag{13}$$

which is independent of $\tilde{\rho}$ if σ is independent of $\tilde{\rho}$ (see Eq. 3).

This prediction is verified in the bottom right panel of **Figure 2**, where the σ^2 values are shown to be nearly uniform at different concentrations, with the exception of the data at the lowest concentration values, i.e. those furthest away from the critical concentration. These results are consistent with the scaling ansatz in the vicinity of $\tilde{\rho} = 0$.

Taken together the results in the previous sections, we found that scaling of the droplet size distribution, Eq. 7, holds with universal exponent $\alpha = 0$, $\varphi = 1$ (**Figure 1**) and

$$F(x) = x^{-1}\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\left(\frac{\ln x}{\sigma}\right)^2/2\right\}, \tag{14}$$

with σ independent of $\tilde{\rho}$ (**Figure 2**).

Our analysis does not exclude the possibility that σ might depend on specific experimental conditions, even though the data that we have analysed are suggestive of at most a weak dependence.

Estimation of the critical concentration from the scaling invariance. The fact that the scaling ansatz is satisfied for different set of experiments offers the possibility to estimate the critical concentration. Our scaling predicts that $(\langle s_k \rangle)^{-1/k}$ versus ρ should be a straight line with a slope depending on k. It is important to note that this is the consequence only of the scaling ansatz and not of the log-normal distribution. The line should intersect the ρ-axis at the critical concentration providing an estimate of ρ_c.

We illustrate this process using the FUS data. We plotted $(\langle s_k \rangle)^{-1/k}$ versus ρ. As expected, for different values of k we obtained a straight line fit of the points near ρ_c (**Figure 3**). Due to experimental uncertainties, the various lines, one for each value of k, lead to a slightly different estimate of ρ_c. The average of the estimated ρ_c values is 5.1 ± 0.2 μM for FUS and 6.6 ± 0.8 μM for SNAP FUS (**Figure 3**). The different ρ_c. values predicted based on the scaling ansatz using different k values enable the estimation of error of the predicted critical concentration (**Figure 3**). Note that the two estimates of the critical concentration for FUS and SNAP FUS are compatible within 2 standard deviations, but also that both estimates are higher than the value estimated originally from the plateau of the absorbance of a spin-down assay (ρ_c = 2 μM of untagged and ρ_c = 3 μM of tagged FUS (26)). Thus, we used both the values of ρ_c to probe the collapse of the size distribution functions (**Figure 4**). We observed that both in cases of untagged and tagged FUS, the SDFs collapsed with our estimated value of ρ_c (**Figure 4**).

These results indicate that the scaling model can be used to estimate the critical concentration based on the distribution of droplet sizes.

Discussion

Growing experimental evidence indicates the presence of protein condensates both in vitro and in vivo (3-5, 32), but the mechanisms responsible for the formation of these condensates are not fully understood. In particular, there is still the need of a general framework capable to fully account for all the experimental observations of protein phase separation. Here, to help build this framework, we aimed at describing phase separation as a universal phenomenon using a scaling ansatz.

With this aim, we analyzed a series of droplet size distributions of FUS recently reported (26). We found that the droplet size distribution followed a log-normal behaviour and can be described by a scale-invariant model. The critical exponents of the scaling model are consistent for different systems. The compatibility of the droplet distributions with a scale-invariant model indicates its universality, so that the properties for a large class of systems are independent of the dynamical details. This is in contrast to the Flory-Huggins theory, which describes the droplets as the result of a nucleation process in the metastable state of supersaturated system. We also note the generality of the log-normal model for nucleation and grain size growth ranging from crystal seeding (33, 34) to the mass size distributions in organism growth of various sea organisms (31).

As a practical consequence of the scaling model, we found that the moments of the droplet size distribution correlate with the deviation from the critical concentration. This suggests that the analysis of the moments of the droplet size distribution can provide a universal gauge of the proximity to the phase boundary irrespective of the actual mechanism of phase transition.

The presence of a scale-invariant log-normal distribution for the droplet size below the critical concentration is not readily consistent with classical nucleation theory. More generally, the presence of scale invariance of this distribution imposes stringent constraints on possible theoretical models describing protein phase separation. It will be interesting to understand how to modify the Flory-Huggins theory to account for the presence of a second order transition. In this respect, we note that at the temperature at which the data that we used were acquired, the systems that we studied are not likely to be close to the critical point of the Flory-Huggins theory because there is a difference in concentration between the dilute and dense phases. It is interesting, however, to also note that near the critical point of the Flory-Huggins theory, the first-order transition line (the binodal line) and the line where the coexistence of the two phases disappears (the spinodal line) are close. In protein systems under these conditions, the correlation length may become larger than the size of the cell, and a first-order transition may become effectively indistinguishable from a second-order transition.

Conclusions

We have addressed the problem of understanding the nature of the transition that leads to protein phase separation. By analysing the functional form of the droplet size distribution, we found a scale-invariant log-normal behaviour, which suggests that protein phase separation may have features typical of second-order phase transitions.

As a prediction derived from this functional form of the droplet size distribution, we have presented a method of estimating the critical concentration for protein phase separation. This distribution can be determined in the test tube, cell cultures or human tissues. The droplet dimensions could be assessed in 1D (length), 2D (area) or 3D (volume).

We note that second-order phase transitions usually result from critical phenomena, and that these processes tend to be highly sensitive to environmental conditions. The modulation of the formation and dimensions of protein droplets therefore could be more tunable than what

expected from first-order transitions. This feature would appear to be favourable for the control of the formation of biomolecular condensates by the protein homeostasis system.

Methods

Calculations of survival probability distribution (SDF) of the droplet size

The fluorescence intensity of the FUS droplets was reported to be proportional to their diameter (26). The sizes of untagged FUS droplets formed at 0.125, 0.25, 0.5 and 1.0, 2.0 μM concentrations were determined by 51 measurements with three repetitions for each concentration, and the sizes of SNAP-tagged FUS at 0.125, 0.25, 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 μM concentrations were determined in 64 measurements with three repetitions for each concentration (26). For each data point, we determined the mean droplet size and standard deviation of the three experimental replicates. The survival function was computed as $P_{>x} = N_{>x}/N_{tot}$, where $N_{>x}$ is the number of droplets above size x, and N_{tot} is the number of droplets. The values of x were chosen according to the original data. Then we computed the k-th moment of each droplet size, determined the average for each value of k and computed the ratios of the subsequent moments (Eq. 4). Quadratic error propagation was used to estimate errors for the different moments and related functions. We tried several k values from 0.25 to 3 in 0.25 increments, determined the moment ratios in each case. We selected those four k values (k=0.5, k=1.0, k=1.5 and k=2.0) where the moment ratios at each concentration provided the best linear fit for both systems.

Determination of the critical exponents

We generated a log-log plot of the moments (Eq. 4) and $\tilde{\rho}$ the distance from the critical concentration (**Figure 1**). For each value of k, we performed a linear regression weighted by the inverse of the variance of the data points corresponding to the different concentrations (26). At each k value, we determined the slope of the linear regression, resulting in four φ_i values, with an associated error from the weighted linear regression.

Given the independent observations φ_i with variances σ_i^2, the value of the critical exponent was obtained as inverse-variance weighted average $\bar{\varphi} = \frac{\sum_i \varphi_i/\sigma_i^2}{\sum_i 1/\sigma_i^2}$. The error of the critical exponent φ was computed as $Var(\bar{\varphi}) = \frac{1}{\sum_i 1/\sigma_i^2}$.

The exponent α was determined based on the log-log plot of the mean of the droplet size distribution versus the distance from the critical concentration $1/\tilde{\rho}$ (**Figure 1**). We performed a linear regression weighted by the inverse of the variance of the data points corresponding to the different concentrations (Eq. 6) using φ =1.

Determination of the parameters of the log-normal behavior

At each concentration, from the distribution of ln(s) we determined s_0 as $\langle \ln(s) \rangle = \ln(s_0)$, and σ as the standard deviation of ln(s). At each concentration, we plotted the droplet size distribution function versus $\frac{ln(s/s_0)}{\sigma}$ (Eq. 10) using the s_0 and σ values determined from the corresponding droplet size distribution at the given concentration. The log-normal behaviour is demonstrated by the resulting collapse (**Figure 2**). Furthermore, the collapsed curves overlapped with a

theoretical log-normal curve computed with s_0=1, **σ** = 1, which is just the normal distribution in the rescaled log variables.

Determination of the critical concentration

We used the moments of the droplet sizes (Eqs. 1 and 2) and plotted them versus the concentration. We used k values (0.5, 1.0, 1.5, 2.0) to cover majority of the statistics, as in **Figure 1**. We performed the linear regression weighted by the inverse variance, determining the intercept with the x-axis. We had the regression errors associated to both intercept and slopes for each value of k. We estimated the critical concentration from each fit as the intercept calculated with $\rho_{c,i} = \frac{-q}{m} \equiv g$ where *q* and *m* are the y-axis intercept and the slope retrieved by the fit, respectively. The error for each estimate of $\rho_{c,i}$ is obtained propagating, on the expression *g* above, as

$$\sigma_{\rho_{c,i}} = \sqrt{\left(\frac{\partial g}{\partial m}\right)^2 \cdot \sigma_m^2 + \left(\frac{\partial g}{\partial q}\right)^2 \cdot \sigma_q^2} \tag{15}$$

where σ_m is error of the fitted slope *m* and σ_q is the error of the y-intercept of the fit *q*. The estimated ρ_c was obtained as inverse-variance weighted average

$$\overline{\rho_c} = \frac{\sum_i \rho_{c,i}/\sigma^2_{\rho_{c,i}}}{\sum_i 1/\sigma^2_{\rho_{c,i}}} \tag{16}$$

The variance (error) of the estimated critical concentration was calculated as

$$Var(\overline{\rho_c}) = \frac{1}{\sum_i 1/\sigma^2_{\rho_{c,i}}} \tag{17}$$

Estimation of the critical concentration

We determined the distance from criticality $\tilde{\rho}$ using the estimated ρ_c (**Figure 3**). We plotted the dropet size distribution functions at each concentration versus the product of droplet size s and $|\widetilde{\rho}|$. (**Figure 4**). Only when determining the distance from the estimated criticality, we observed the collapse of the droplet size distribution functions.

The *k*-th moment of the log-normal distribution (Eq. 12) and the scaling prediction with α = 0 (Eq. 3, main text) gives

$$(\langle s^k \rangle)^{-1/k} = \frac{1}{s_0} e^{-\frac{k\sigma^2}{2}} \tag{18}$$

with $s_0 = a \cdot \left|\frac{\rho - \rho_c}{\rho_c}\right|^{-1}$, where *a* is a proportionality constant independent of k, ln s_0 is the mean and σ is the standard deviation of the logarithm of the droplet probability size distribution. Therefore

$$(\langle s^k \rangle)^{-1/k} = \frac{1}{a} e^{-\frac{k\sigma^2}{2}} \left| \frac{\rho - \rho_c}{\rho_c} \right| \quad (19)$$

Acknowledgements
We acknowledge insightful discussions with Dr. Serena Carra, Dr. Jonathan Vinet, Dr. Alex Buell and Dr. Soumik Ray.

References

1. Brangwynne CP*, et al.* (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324(5935):1729-1732.
2. Hyman AA, Weber CA, & Julicher F (2014) Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* 30:39-58.
3. Banani SF, Lee HO, Hyman AA, & Rosen MK (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* 18(5):285-298.
4. Alberti S & Hyman AA (2021) Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat Rev Mol Cell Biol* 22(3):196-213.
5. Lyon AS, Peeples WB, & Rosen MK (2021) A framework for understanding the functions of biomolecular condensates across scales. *Nat Rev Mol Cell Biol* 22(3):215-235.
6. Buchan JR & Parker R (2009) Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* 36(6):932-941.
7. Andersen JS*, et al.* (2005) Nucleolar proteome dynamics. *Nature* 433(7021):77-83.
8. Mateju D*, et al.* (2017) An aberrant phase transition of stress granules triggered by misfolded protein and prevented by chaperone function. *EMBO J* 36(12):1669-1687.
9. Qamar S*, et al.* (2018) FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation-pi Interactions. *Cell* 173(3):720-734 e715.
10. Vendruscolo M & Fuxreiter M (2022) Protein condensation diseases: therapeutic opportunities. *Nature communications* 13(1):5550.
11. Mitrea DM, Mittasch M, Gomes BF, Klein IA, & Murcko MA (2022) Modulating biomolecular condensates: a novel approach to drug discovery. *Nat Rev Drug Discov* 21(11):841-862.
12. Fuxreiter M & Vendruscolo M (2021) Generic nature of the condensed states of proteins. *Nature cell biology* 23(6):587-594.
13. Jawerth L*, et al.* (2020) Protein condensates as aging Maxwell fluids. *Science* 370(6522):1317-1323.
14. Flory PJ (1942) Thermodynamics of high polymer solutions. *The Journal of chemical physics* 10:51-61.
15. Huggins ML (1942) Some Properties of Solutions of Long-chain Compounds *The Journal of Physical Chemistry* 46(151-158).
16. Overbeek JTG & Voorn MJ (1957) Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *Journal of Cellular and Comparative Physiology* 49:7-26.

17. Flory PJ (1941) Molecular Size Distribution in Three Dimensional Polymers I. Gelation. *J. Am. Chem. Soc.* 63:3083-3090.
18. Stockmayer WH (1944) Theory of Molecular Size Distribution and Gel Formation in Branched Polymers. *The Journal of chemical physics* 11:45-55.
19. Semenov AN & Rubinstein M (1998) Thermoreversible Gelation in Solutions of Associative Polymers.1. Statics. *Macromolecules* 31:1373-1385.
20. Tavares JM, Teixeira PI, Telo da Gama MM, & Sciortino F (2010) Equilibrium self-assembly of colloids with distinct interaction sites: thermodynamics, percolation, and cluster distribution functions. *The Journal of chemical physics* 132(23):234502.
21. Bueno C, Liman J, Schafer NP, Cheung MS, & Wolynes PG (2022) A generalized Flory-Stockmayer kinetic theory of connectivity percolation and rigidity percolation of cytoskeletal networks. *PLoS Comput Biol* 18(5):e1010105.
22. Mittag T & Pappu RV (2022) A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol Cell* 82(12):2201-2214.
23. Thomson JA, Schurtenberger P, Thurston GM, & Benedek GB (1987) Binary liquid phase separation and critical phenomena in a protein/water solution. *Proc Natl Acad Sci U S A* 84(20):7079-7083.
24. Ishimoto C & Tanaka T (1977) Critical Behavior of a Binary Mixture of Protein and Salt Water. *Phys. Rev. Lett.* 39:474-477.
25. Stender EGP*, et al.* (2021) Capillary flow experiments for thermodynamic and kinetic characterization of protein liquid-liquid phase separation. *Nature communications* 12(1):7289.
26. Kar M*, et al.* (2022) Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc Natl Acad Sci U S A* 119(28):e2202222119.
27. Li Y, Lubchenko V, Vorontsova MA, Filobelo L, & Vekilov PG (2012) Ostwald-like ripening of the anomalous mesoscopic clusters in protein solutions. *The journal of physical chemistry. B* 116(35):10657-10664.
28. Shimobayashi SF, Ronceray P, Sanders DW, Haataja MP, & Brangwynne CP (2021) Nucleation landscape of biomolecular condensates. *Nature* 599(7885):503-506.
29. Brankov JG, Danchev DM, & Tonchev NS (2000) *Theory of critical phenomena in finite-size systems: scaling and quantum effects* p 460.
30. Stauffer D & Aharony A (1992) *Introduction To Percolation Theory* (Taylor & Frances) p 192.
31. Giometto A, Altermatt F, Carrara F, Maritan A, & Rinaldo A (2013) Scaling body size fluctuations. *Proc Natl Acad Sci U S A* 110(12):4646-4650.
32. Boija A*, et al.* (2018) Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* 175(7):1842-1855 e1816.
33. Binder K & Virnau P (2016) Overview: Understanding nucleation phenomena from simulations of lattice gas models. *The Journal of chemical physics* 145: 211701-211711.
34. Bergmann RB & Bill A (2008) On the origin of logarithmic-normal distributions: An analytical derivation, and its application to nucleation and growth processes. . *Journal of Crystal Growth* 310:3135-3138.

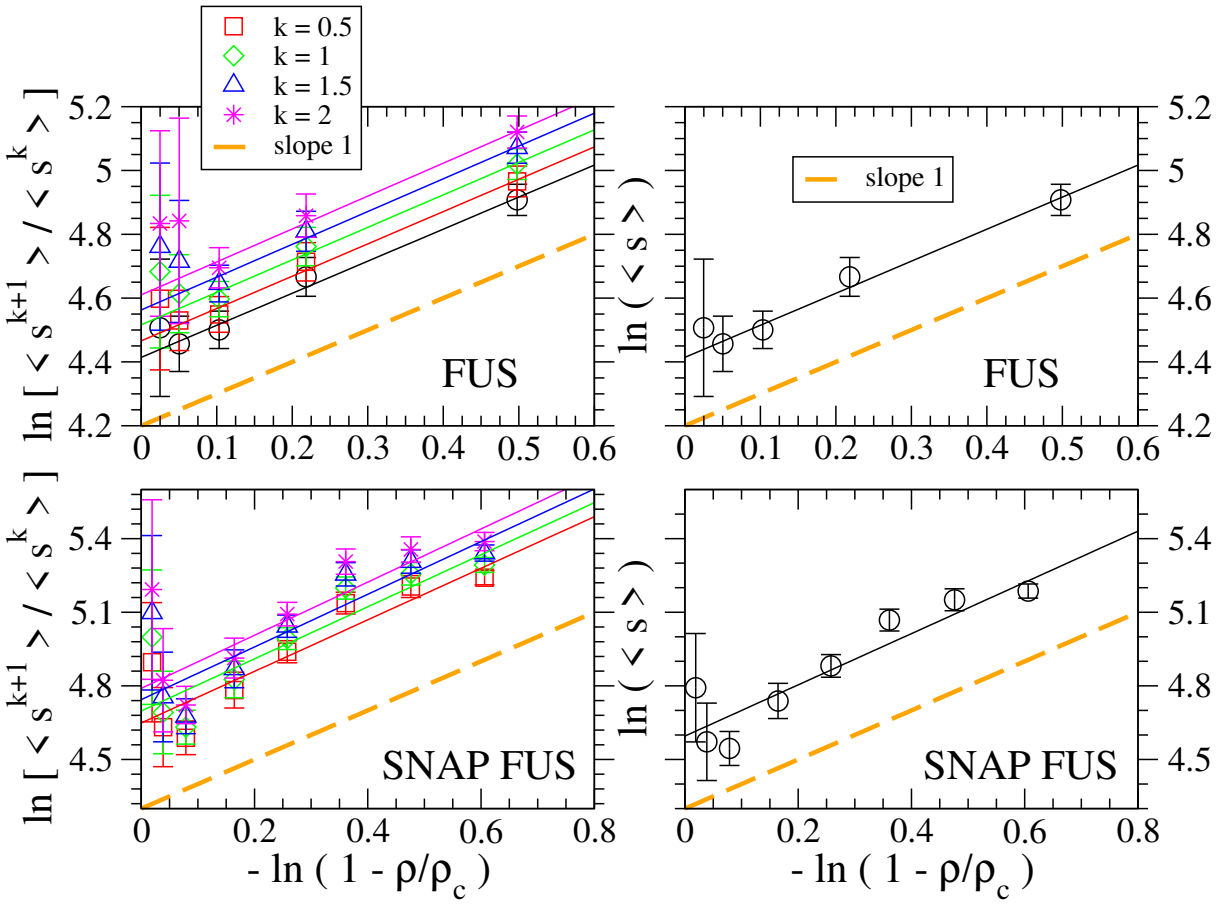

Figure 1. Determination of the critical exponents of the scaling ansatz. The ratios of the average moments of the droplet sizes ($<s^{k+1}>/<s^k>$, at k=0.5, 1, 1.5, 2) are represented at various distances from the critical concentration ($|\widetilde{\rho}|$). Top row: FUS. Bottom row: SNAP FUS. Error weighted linear regressions are performed in both cases excluding the data point at the lowest concentration ρ = 0.125 μM. The fit corresponding to the scaling ansatz, compatible with φ=1 and α=0, is represented by a line with a slope of 1.

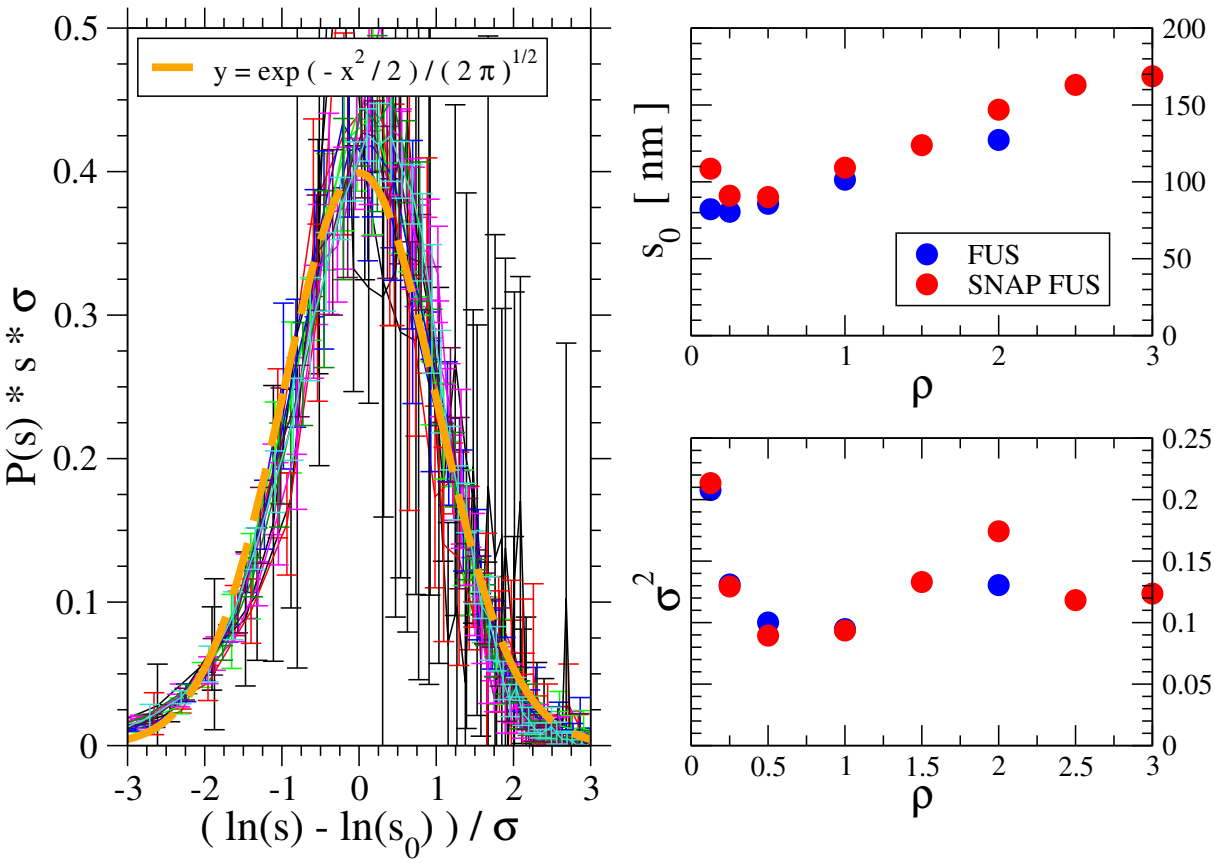

Figure 2. The collapse of the size distribution functions indicates a log-normal behavior. The droplet size distribution functions for both untagged FUS and SNAP-tagged FUS (colour codes are the same as in Figure 4) are plotted after rescaling sizes by means of the lns_0 and σ values, the first and second moment of the logarithm of the droplet size, which are characteristic of the distributions at each concentration. The variation with concentration of lns_0 and σ is shown in the right panels for both FUS and SNAP FUS. Rescaled curves for both the untagged and the tagged protein collapse in the left panel to the normal distribution (orange dashed line), as predicted if the non-rescaled droplet sizes follow a log-normal distribution.

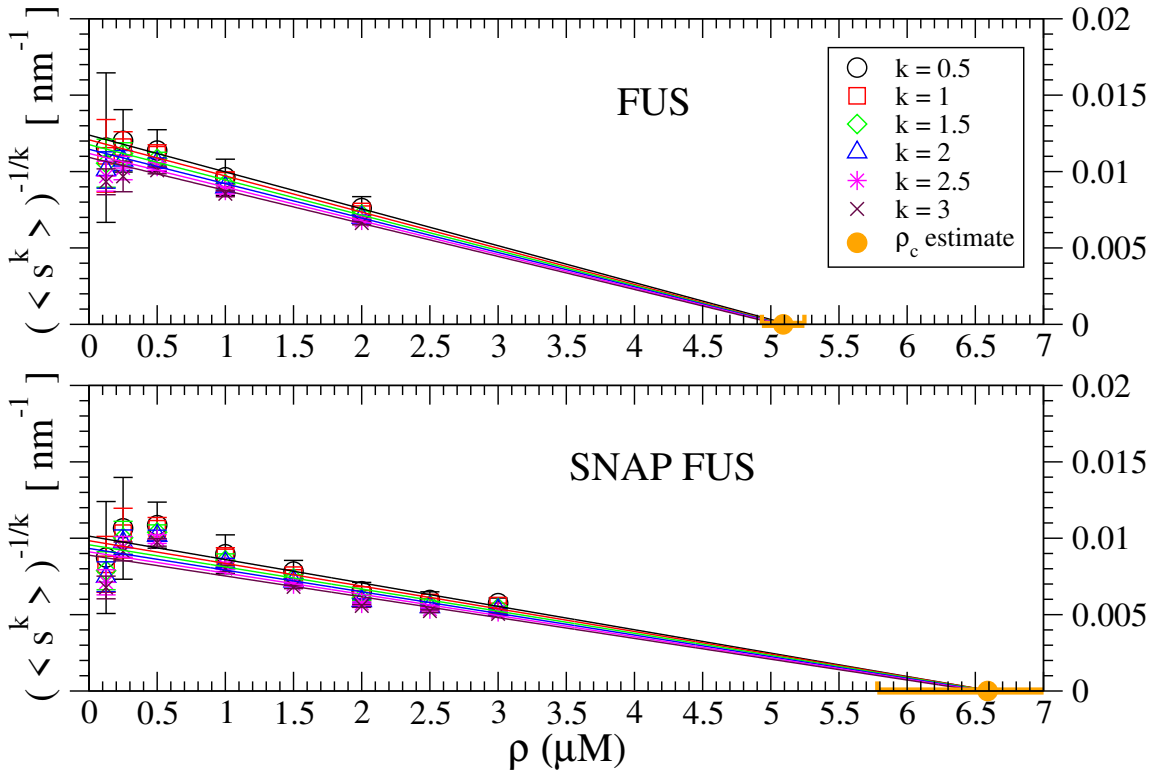

Figure 3: Determination of the critical concentration using the scaling model. The scaling model predicts that the function of the moments plotted versus the concentration ρ becomes a straight line near the critical concentration ρ_c and intersects the ρ-axis at ρ_c, independently of the value of k. Top row: FUS. Bottom row: SNAP FUS. Error weighted linear regressions are performed in both cases excluding the data point at the lowest concentration ρ = 0.125 μM. The resulting estimate of the critiical concentration is shown in orange along with the corresponding standard deviation.

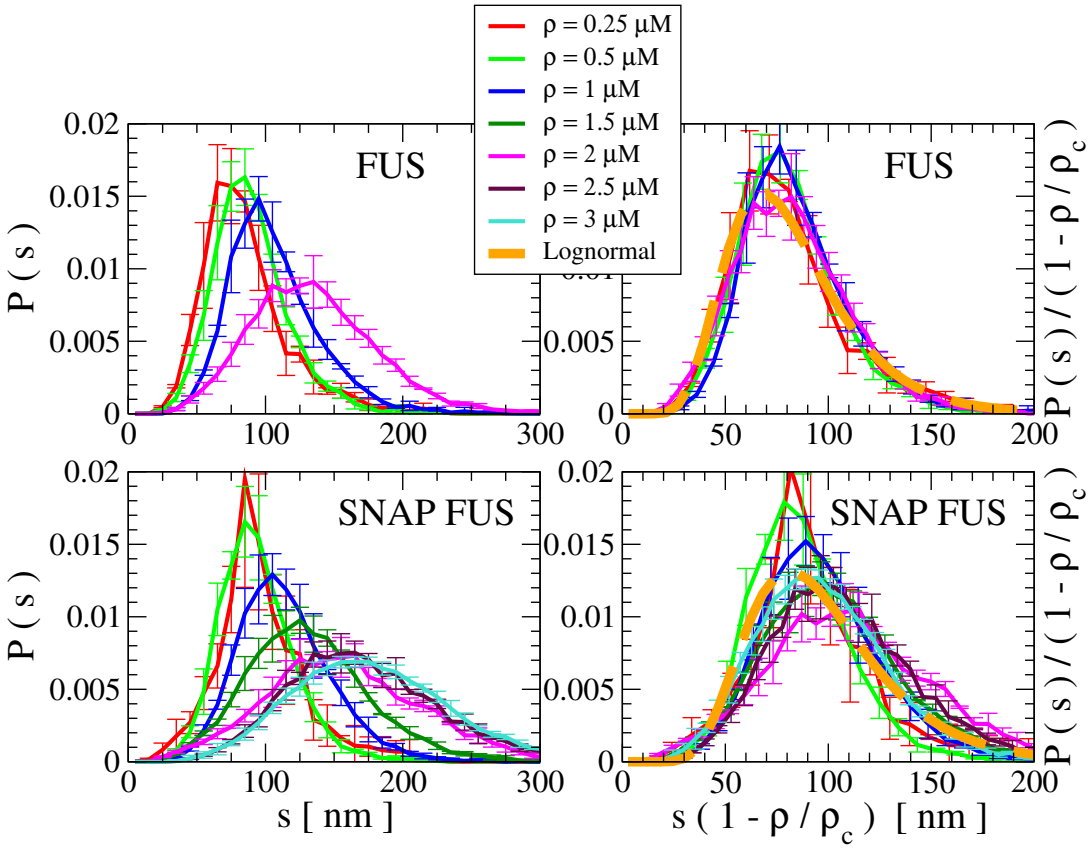

Figure 4. Log-normal droplet size distributions are independent from the distance from the criticality. If the scaling ansatz (Eq. 2) holds, the standard deviation σ of the log-normal distribution should not depend on the distance from the critical concentration, and collapse should be achieved by rescaling the size with the distance $|\widetilde{\rho}| = 1 - \frac{\rho}{\rho_c}$ from the critical concentration. We derived droplet sizes from FUS (26) experiments at 0.125, 0.25, 0.5, 1.0 and 2.0 μM concentrations of untagged FUS and 0.125, 0.25, 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 μM concentrations of SNAP-tagged FUS. Top row: FUS. Bottom row: SNAP FUS. Left column: non-rescaled droplet size distributions. Right column: the collapse of the rescaled droplet size distributions. For both FUS and SNAP FUS the log-normal distribution with the moments determined by the droplet size distribution at the highest concentration value is shown in dashed orange. The droplet size distributions at the lowest concentration value are not shown.