bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Quantification of surface tension effects and nucleation-and-growth rates during self-assembly of biological condensates

Zsuzsa Sárkány[a,b], Fernando Rocha[c], Anna Bratek-Skicki[d,e], Peter Tompa[d,e,f], Sandra Macedo-Ribeiro[a,b] and Pedro M. Martins[a,b,]*

[a]IBMC Instituto de Biologia Molecular e Celular, Universidade do Porto, 4150-180 Porto, Portugal

[b]i3S Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4150-180 Porto, Portugal

[c]LEPABE Laboratory for Process Engineering Environment Biotechnology and Energy, Faculdade de Engenharia da Universidade do Porto

[d]VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnology, Brussels, Belgium

[e]Structural Biology Brussels (SBB), Bioengineering Sciences Department, Vrije Universiteit Brussel (VUB), Brussels, Belgium

[f]Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary

[*]To whom correspondence may be addressed; E-mail: pmartins@icmc.up.pt

Abstract

Liquid-solid and liquid-liquid phase separation (PS) drives the formation of functional and disease-associated biological assemblies. Principles of phase equilibrium are here employed to derive a general kinetic solution that predicts the evolution of the mass and size of biological assemblies. Thermodynamically, protein PS is determined by two measurable concentration limits: the saturation concentration and the critical solubility. Due to surface tension effects, the critical solubility can be higher than the saturation concentration for small, curved nuclei. Kinetically, PS is characterized by the primary nucleation rate constant and a combined rate constant accounting for growth and secondary nucleation. It is demonstrated that the formation of a few number of large condensates is possible without active mechanisms of size control and in the absence of coalescence phenomena. Our exact analytical solution can be used to interrogate how the elementary steps of PS are affected by candidate drugs.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Multiple lines of evidence indicate that liquid-liquid phase separation (LLPS) of proteins has an important role in the spatiotemporal regulation of cell functioning and disease development [1]. These findings motivate a renewed interest in the detailed description of the nucleation-and-growth mechanisms by which new protein phases are generated. Given the tinctorial properties of amyloid fibrils, a large number of kinetic curves have been recorded to characterize protein aggregation *in vitro* using thioflavin-T and other specific dyes [2]. These studies are important to understand the nucleation-and-growth of amyloid fibrils and other insoluble deposits associated with neurodegenerative diseases. Biophysical models such as those proposed by Crespo et al. [3,4] and by Finke and Watzky [5,6], can describe, at least qualitatively, the hyperbolic, S-shaped and sigmoidal progress curves usually reported for peptides and proteins such as amyloid-β peptide (Aβ), α-synuclein, tau, prion, transthyretin (TTR), insulin, ataxin-3, β_2-microglobulin, huntingtin, etc. In the case of sigmoidal curves, the nucleation-and-growth mechanism is further validated by scaling laws of half-life coordinates and maximal aggregation rate vs. protein concentration [7,8], and by measurements of particle size over time [9]. Although less frequently observed, hyperbolic-shaped progress curves characterize the amyloid aggregation of TTR [10], as well as the aggregation of Aβ at high peptide concentrations [11] or in the presence of amyloid inhibitors [12].

Hyperbolic, S-shaped and lag-phased curves are also expected by the nucleation-and-growth model proposed by Zwicker et al. [13] intended to describe the sigmoidal growth of centrosomes as an LLPS process controlled by an enzymatic activity of two centrioles within the centrosome matrix. Other membraneless organelles, like nuclear bodies and cytoplasmic granules undergo coacervation through the recruitment of protein and RNA molecules in supersaturated solutions, giving rise to a power-law dependence of the average droplet size $\langle R \rangle$ on time t. If a fixed number of growing droplets is assumed, the time exponent n takes on values of $\sim 1/2$ in purely diffusional regimes and ~ 1 in kinetic regimes determined by surface attachment [14]. The $n = 1/2$ exponent is also predicted by the theory of diffusion-limited precipitation, which, in addition, rationalizes the application of the Johnson–Mehl–Avrami–Kolmogorov (JMAK) equation to the early stages of LLPS progress curves [14]. Both the JMAK- and logistic-type equations predict that the supersaturated phase becomes fully depleted of protein for long reaction times or, equivalently, that the cellular compartment becomes completely filled with the droplet phase [4,14]. In order to numerically fit experimental data, these equations can be modified by the inclusion of corrected asymptotic limits for $t \to \infty$ [14,15], whereas a power-law relationship between $\langle R \rangle(t)$ and the reaction conversion are admissible if the number of growing particles is approximately constant [6]. After the nucleation-and-growth period has passed, the initially

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

supersaturated phase becomes saturated and the volume fraction of the new phase increases no more. From this point onwards, the average size of dense droplets in saturated solutions can continue to increase due to Ostwald ripening and/or coalescence processes in which smaller droplets convert to fewer and larger droplets driven by surface tension effects. The coalescence period that follows nucleation-and-growth is well characterized by a dynamic scaling exponent n of $1/3$ [14].

The nucleation-and-growth of a few isolated particles without the continual formation of new nuclei is a challenge to the existing kinetic models with important implications for our understanding of pathogenic protein aggregation and cell organization. The time and extension of nucleation can be regulated by enzymatic activity [13], parallel oligomerization reactions [8] and ATP-driven reactions [14] that indirectly affect the supersaturation level. New techniques and methods for studying LLPS show, however, that the formation of a reduced number of non-coalesced condensates is possible in the absence of external sources of energy or matter [15–20]. To help understand these intriguing observations, we start by including surface tension effects (STEs) on established nucleation-and-growth models; the resulting closed-form kinetic equations are then tested against LLPS progress curves and are used to predict the variation with time of droplet size distributions. Finally, we discuss fundamental and biological implications arising from STEs-affected assembly and summarize the steps that should be taken to quantify STEs and nucleation-and-growth rates.

Thermodynamics and kinetics of nucleation-and-growth

The formation of liquid protein phases follows thermodynamic and kinetic principles that are common to other phase-transition phenomena. For example, the theoretical free energy of the mixing of binary polymer-solvent solutions can be used to estimate the equilibrium concentrations of the protein-depleted (or *light*) and protein-enriched (or *dense*) phases formed upon LLPS [21,22]. In temperature vs. composition phase diagrams, these concentrations are generally expressed in terms of the protein volume fractions φ_L and φ_D, and define the so-called *coexistence curve* or binodal (Fig. 1a). For attractive interactions between molecules, there is a critical temperature T_c above which phase separation does not occur, regardless of the concentration [23,24]. LLPS provoked by increasing temperatures can, however, occur in the case of predominantly hydrophobic interactions [25]. In liquid-solid phase diagrams, the equilibrium compositions of the light (liquid) and dense (solid) phases are given by the *solubility curve* (or liquidus) and by the solidus line, correspondingly (Fig. 1b). The volume fraction of the protein in the solid phase does not change significantly with temperature and, in the case of protein crystals, can be lower than 0.5. Liquid-liquid immiscibility can be metastable in relation to the solid phase, in which case LLPS is a precursor step of the more favourable liquid-solid

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transition [24,26,27]. According to the phase rule of thermodynamics, the number of possible microstates due to demixing increases with the number of non-reactive components present in the mixture [28]. Consequently, the complexity of the phase diagrams increases dramatically by considering the occurrence of interaction partners and metabolites, but also by including the effect of pH, ionic strength and additives [29], possible eutectic behaviours [30], the spontaneous formation of the new phase by spinodal decomposition [28], etc. In what follows, we will address the regions of protein phase diagrams where the formation of stable particles (protein crystals, aggregates or droplets) takes place by a nucleation-and-growth mechanism.

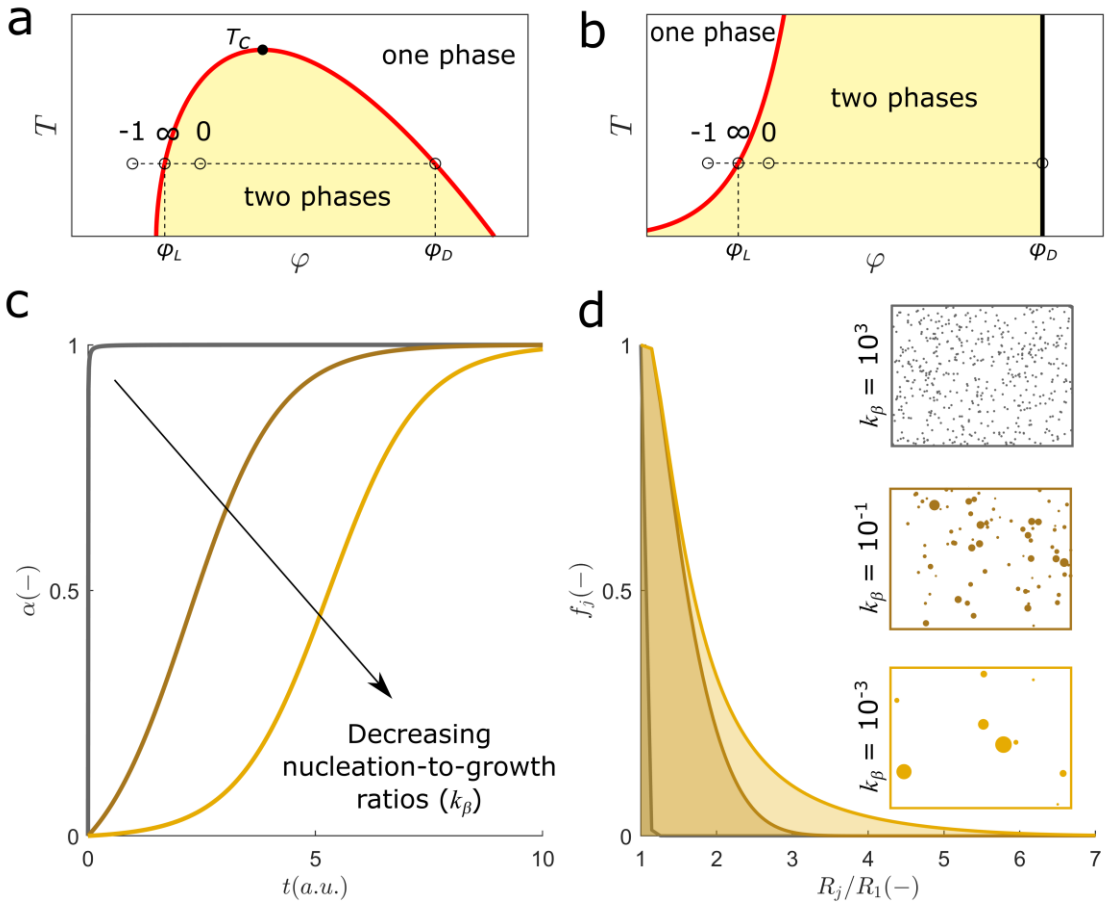

Figure 1. Thermodynamics and kinetics of nucleation-and-growth. (a and b) Schematic representation of phase diagrams of binary protein-water mixtures representing temperature (T) vs. volume fraction of the protein (φ). Points along the dashed lines represent undersaturated (-1), supersaturated (0) and saturated conditions (∞) at a fixed temperature. (a) liquid-liquid equilibrium: the volume fraction of the protein in the light (φ_L) and in the dense (φ_D) phase are given by the coexistence curve (solid line) that shows a maximum at the critical temperature T_c. (b) liquid-solid equilibrium: the values of φ_L and φ_D are given by solubility curve (red line) and by the solidus line (black line), respectively. (c) Theoretical progress curves for nucleation-to-growth ratios of (from left to right) $k_\beta = 10^3$, 10^{-1} and

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10^{-3} (*SI Additional Methods*). (d) Normalized frequency (f_j) of particles with size R_j calculated for long reaction time ($t \to \infty$) using the same parameters and colour code as in (c); R_1 is the size of primary nuclei. Cartoon boxes: the final population of particles is schematically represented as a distribution of circles. No secondary events such as secondary nucleation, breakage or coalescence are considered in (c) and (d).

Consider the change from the one-phase to the two-phase regime of the phase diagram by means of the isothermal increase of protein concentration from $\varphi_{-1} < \varphi_L$ to $\varphi_0 > \varphi_L$. If the temperature is kept constant, the system will naturally relax towards equilibrium through the nucleation and subsequent growth of a dense phase that has constant composition (φ_D) and decreases the protein volume fraction in the light phase φ until saturated conditions are reached ($\varphi_\infty = \varphi_L$). The individual rates of the nucleation and growth steps determine how fast the change from φ_0 to φ_∞ is, whereas the ratio of nucleation-to-growth rates (k_β) determines the shape of phase separation progress curves. In the absence of other events, increasing values of k_β lead to a transition from hyperbolic, to S-shaped to lag-phased progress curves (Fig. 1c) [3]. In concentration units, the amount of protein separated into the new phase is $M(t) = c_0 - c(t)$, where c_0 and $c(t)$ are the initial and instantaneous protein concentrations; $M(t)$ increases with time until reaching the final value of $M_\infty = c_0 - c_\infty$. In the phase diagram, c_0 and c_∞ correspond to volume fractions φ_0 and $\varphi_\infty = \varphi_L$, respectively. The instantaneous reaction conversion is $\alpha(t) = M(t)/M_\infty$. For large particles to be formed, primary nucleation should be much slower than the growth step ($k_\beta \ll 0.01$) so that early nucleated species can continue to grow during the whole duration of lag-phased progress curves. Because of the long period required for phase separation, large particles end up coexisting with recently formed ones that have much smaller sizes (Fig. 1d). If, instead, primary nucleation is fast compared with the growth step (high k_β), numerous small condensates will be produced until phase equilibrium is ultimately achieved (Fig. 1d). In this context, the sharp formation of a few number of large particles is an apparent paradox, whose solution is here sought using principles of phase equilibrium at curved interfaces.

Surface Tension Effects (STEs) on Nucleation-and-Growth

We investigated how STEs might affect the early steps of phase separation. Besides driving Ostwald ripening [31] and droplet coalescence in later stages of LLPS [32], surface tension is also expected to influence the initial assembly of critically-sized nuclei [33,34]. In fact, the condition of mechanical equilibrium between phases establishes that the difference in fluid pressure inside and outside a curved surface increases as the radius of curvature decreases. For a spherical surface of radius r, this positive

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

difference is given by the Young-Laplace equation $\Delta p = 2\gamma/r$, where γ is the surface tension. Consequently, the formation of primary nuclei with a radius of curvature $r \approx R_c$ is driven by a difference in chemical potential $\Delta\mu_c$ that is lower than the one corresponding to a flat interface ($r = \infty$). By combining the Young-Laplace equation with the condition of chemical equilibrium between phases ($\Delta\mu = 0$), an equation of the Ostwald–Freundlich type is obtained [35], relating the critical protein solubility c_c with saturation concentration expected for flat interfaces (c_∞):

$$c_c = c_\infty \exp\left(\frac{\upsilon\gamma}{\mathrm{R}TR_c}\right) \tag{1}$$

where υ is the molar volume of the protein and R is the universal gas constant. This equation estimates the protein solubility enhancement elicited by critically-sized nuclei relative to larger particles characterized by $r \approx \infty$. Taking as reference the values of $\gamma = 0.5$ mJ/m^2 estimated for lysozyme crystals [26], and of $\upsilon = 0.01$ m^3/mol, a 2% increase in protein solubility is predicted for values of $R_c = 100$ nm that are close to the initial sizes of amyloid aggregates and of liquid droplets measured by dynamic light scattering [8,16]. The enhancement of protein solubility will be as large as 21% if an estimate of $R_c = 10$ nm is used in Eq. 1.

The predicted magnitude of STEs is contingent on fixed parameters such as γ or R_c, but is also greatly dependent on particle geometry premises. It is likely, for example, that flatter secondary nuclei generated at the surface of the new phase would be less affected by STEs than sphere-like primary nuclei. On the other hand, even slight enhancements of protein solubility will significantly affect the variation in chemical potential during phase separation, which is generally approximated by the value of supersaturation [3]. Supersaturation can also be expressed in concentration units normalized to the initial conditions (subscript 0):

$$\frac{\Delta\mu(t)}{\Delta\mu_0} \approx \frac{(c(t) - c_\infty)/c_\infty}{(c_0 - c_\infty)/c_\infty} = \frac{c(t) - c_\infty}{c_0 - c_\infty} \tag{2}$$

When STEs affect primary nucleation, this typical definition has to be modified taking into account the critical solubility c_c corresponding to the protein concentration in equilibrium with critically-sized particles:

$$\frac{\Delta\mu_c(t)}{\Delta\mu_0} \approx \begin{cases} \dfrac{(c(t) - c_c)/c_c}{(c_0 - c_\infty)/c_\infty} = \dfrac{c_\infty}{c_c}\left(\dfrac{\Delta\mu(t)}{\Delta\mu_0} + \alpha_c - 1\right) & \text{for } c \geq c_c \\ 0 & \text{for } c < c_c \end{cases} \tag{3}$$

where $\Delta\mu(t)/\Delta\mu_0$ is given by Eq. 2, and the critical conversion α_c is:

$$\alpha_c = \frac{c_0 - c_c}{c_0 - c_\infty} \tag{4}$$

As implied from the above definitions, primary nucleation ceases to occur when the protein

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

concentration decreases below c_c or, equivalently, when the reaction conversion increases above α_c. Thus, STEs can be quantified as the reaction extent during which no primary nucleation is observed: $\text{STEs} = 1 - \alpha_c$.

Nucleated particles are the result of successive monomer addition events until the critical nucleus size is attained and intermolecular forces within the new phase start to prevail over the forces established with the surrounding solution [36,37]. Primary nucleation of macromolecules may also involve substantial structural rearrangements as in the case of amyloid fibril formation [38]. Although primary nucleation is a multistep process, the stochastic assembly of molecules that leads to the formation of a stable nucleus is generally considered rate-limiting owing to the high energy cost required for generating the new interface. In the presence of size-dependent STEs, the energetic cost will be even higher owing to the solubility enhancement caused by small, curved interfaces. It is expected that once this energetic barrier is surpassed, a period of very fast growth will follow catalyzed by the exponential decrease of solubility, and will continue until the radius of curvature is fully relaxed. In such cases, the timescale for primary nucleation is still determined by the initial formation of critically-sized nuclei, but the end-product of the nucleation process consists of particles with size R_1 larger than R_c. As these particles are sufficiently flat and/or large to be insensitive to STEs, their subsequent growth will be driven by the concentration difference $(c(t) - c_\infty)$, until the saturation concentration c_∞ is eventually reached. Therefore, in contrast to what happens for primary nucleation, the driving force for growth can still be defined using Eq. 2.

Departing from the Young-Laplace treatment of curved interfaces, we can conclude that the classical definition of driving force for primary nucleation has to be sometimes corrected to account for the enhancement of protein solubility caused by shape- and size-dependent surface tension effects. This enhancement can be pronounced in the case of small, spherical nuclei but rapidly vanishes as flatter geometries are considered. The existence of a critical solubility value c_c higher than the saturation concentration c_∞ implies that primary nucleation ceases to occur for sub-critical protein concentrations, i.e., only secondary processes take place for protein concentrations $< c_c$.

Exact Solutions of the General Nucleation-and-Growth Mechanism

To quantitatively evaluate the impact of STEs, previously established mass balance equations describing the nucleation-and-growth of crystals and amyloid fibrils [4,9] are now modified to include the generalized definition of the driving force for primary nucleation (Eq. 3 and Appendix: *General Nucleation-and-Growth Model*):

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\frac{1}{k_\alpha \Delta\mu_0}\frac{dP(t)}{dt} = \frac{M_\infty k_\beta}{N_1}\left(\frac{\Delta\mu(t)}{\Delta\mu_0} + \alpha_c - 1\right)^2 + \frac{k_2}{k_\alpha N_2}\frac{\Delta\mu(t)}{\Delta\mu_0}M(t) \quad (5a)$$

$$\frac{1}{k_\alpha \Delta\mu_0}\frac{dM(t)}{dt} = M_\infty k_\beta\left(\frac{\Delta\mu(t)}{\Delta\mu_0} + \alpha_c - 1\right)^2 + \frac{\Delta\mu(t)}{\Delta\mu_0}M(t) \quad (5b)$$

$$N(t) = \frac{M(t)}{P(t)} \quad (5c)$$

where $P(t)$ is the number concentration of particles; $N(t)$ is the mean particle size expressed in terms of number of condensed units; N_1 and N_2 are the number of condensed units constituting the primary- and secondary-nucleation particles, respectively; and k_α, k_β and k_2 are c_0-independent rate constants describing the autocatalytic, primary nucleation and secondary nucleation steps. Negligible secondary nucleation rates ($k_2 \approx 0$) will be assumed henceforth as one of our goals is to describe nucleation-and-growth processes producing large condensates, whilst numerous particles with mean size close to N_2 are expected in the presence of increasingly stronger secondary nucleation (Fig. S4) The nucleation-and-growth mechanism underlying the above differential equations is here called "general" because it applies to liquid-solid phase separation, but also to LLPS processes that, in principle, are more prone to STEs provoked by the curvilinear interface of liquid droplets. Despite that, the generalizations and simplifying hypotheses adopted to derive Eqs. 5a and 5b limit the applicability of this mathematical model to cases where: (i) primary nucleation is described by second-order rate equations on supersaturation, (ii) growth and secondary nucleation are autocatalytic processes whose rates increase with the mass concentration of particles and are first-order functions of supersaturation, (iii) $P(t)$ is not significantly affected by the occurrence of breakage/fragmentation events, nor by coalescence phenomena, and (iv) the volume of the supersaturated solution remains approximately constant [4,9]. After replacing the definitions of $\Delta\mu(t)/\Delta\mu_0 = 1 - \alpha(t)$ in Eq. 5b, a closed-form $\alpha(t)$ expression is obtained upon analytical integration (Fig. 2a) that is dependent on the time-variables t_1, t_c and τ (Fig. 2b). The critical instant t_c marks the transition from the nucleation-and-growth to the growth-only regime and coincides with the moment the protein concentration $c(t)$ equals c_c and the reaction conversion is $\alpha(t_c) = \alpha_c$; the characteristic time constant τ determines the timescale for phase separation and reduces to $2(k_\alpha \Delta\mu_0)^{-1}$ for $\alpha_c = 1$; t_1 is the instant of maximum $d\alpha(t)/dt$ slope as given by the split $\alpha(t)$ expression for $t \leq t_c$. Note that the products $k_\alpha \Delta\mu_0 t_c$, $k_\alpha \Delta\mu_0 t_1$ and $k_\alpha \Delta\mu_0 \tau$ are functions of two parameters only, α_c and k_β. The representation of the average particle size $\langle R \rangle$ as a function of $k_\alpha \Delta\mu_0$-normalized units of time shows that phase separation becomes slower as STEs get stronger (Fig. 2c). The particle growth curves confirm that the presence of STEs induces the formation of larger particles over more prolonged timescales (Fig. 2d). Similar effects are produced if

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

different nucleation-to-growth ratios are considered (see Figs. 3c and 3d below). Overall, the kinetics of phase separation is determined by the initial supersaturation $\Delta\mu_0$, by two kinetic parameters k_α and k_β that are independent of the initial protein concentration c_0, and by the c_0-dependent value of α_c.

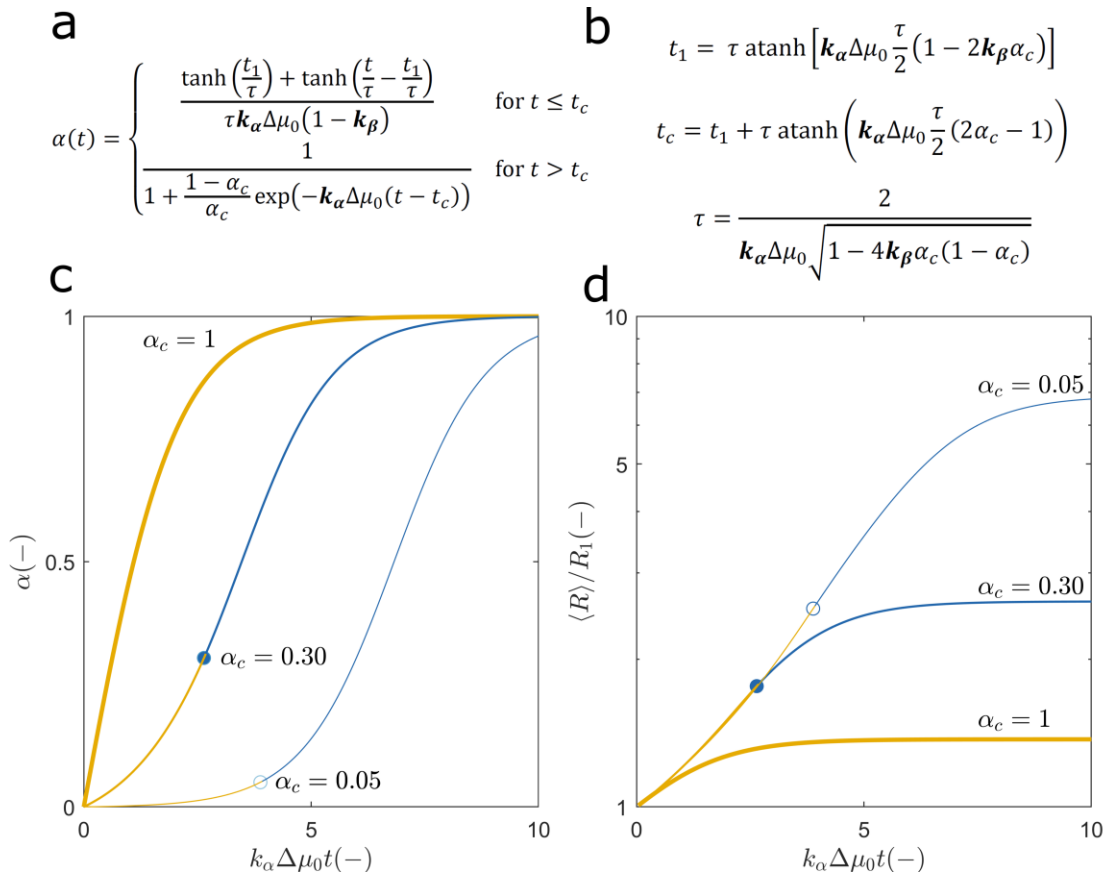

Figure 2. Exact solutions of Eq. 5 describing the nucleation-and-growth of particles in the presence of surface tension effects. (a) Analytical solution of Eq. 5b for the initial condition $\alpha(0) = 0$. (b) Definitions of time-variables t_1, t_c and τ in (a) as a function of two c_0-independent kinetic constants, k_α and k_β (in bold). (c) Representation of theoretical progress curves using (from left to right) $\alpha_c = 1$, 0.30 and 0.05 for a fixed nucleation-to-growth ratio ($k_\beta = 0.5$). Lines of different color represent the nucleation-and-growth period (orange) and the growth-only period (blue). Symbols indicate the location of the critical moment t_c. (d) Log-linear plot of the particle growth curves obtained by numeric integration of Eqs. 5a and 5c assuming the same model parameters as in (c). Symbols, line thicknesses and line colors are as in (c).

The occurrence of STEs may result in the rapid nucleation and subsequent growth of a reduced number of large particles; it can also explain unconventional effects of protein concentration on the measured $\alpha(t)$ curves. Before illustrating these possibilities with practical examples, approximate solutions of the $N(t)$ and $\langle R(t) \rangle$ curves are provided next towards a simple and general methodology for phase-

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

separation kinetic analysis.

Particle Growth Curves - Approximate Solution

Unlike the $\alpha(t)$ solution, the analytical $N(t)$ solution of Eqs. 5a and 5c is too complex to be used in routine analyses of growth curves as an alternative to the exact numerical solution. An approximate expression is now derived, assuming a constant number of growing particles $P(t) \approx P_\infty$, which implies a final mean size of $N_\infty = M_\infty/P_\infty$. On this basis, the exact $\alpha(t)$ equation presented in Fig. 2a for the growth-only period can be extended to the whole nucleation-and-growth period (Fig. 3). By combining $k_\alpha\Delta\mu_0$, k_β and α_c in a single variable β_∞, an approximate $\tilde{\alpha}(t)$ expression is obtained (Figs. 3a and 3b) containing two model parameters (β_∞ and $k_\alpha\Delta\mu_0$) that can be numerically fitted to later stages of $\alpha(t)$ curves (Fig. 3c). Direct replacement of this expression in Eq. 5c gives the equality $\tilde{N}(t)/N_\infty = \tilde{\alpha}(t)$, whose limit for $t \to 0$ leads to the result $N_1/N_\infty = \beta_\infty^{-1}$. The expression for $\tilde{\beta}(t) = \tilde{N}(t)/N_1$ is appropriate for estimating β_∞ and $k_\alpha\Delta\mu_0$ from droplet size measurements (Figs. 3a and 3d).

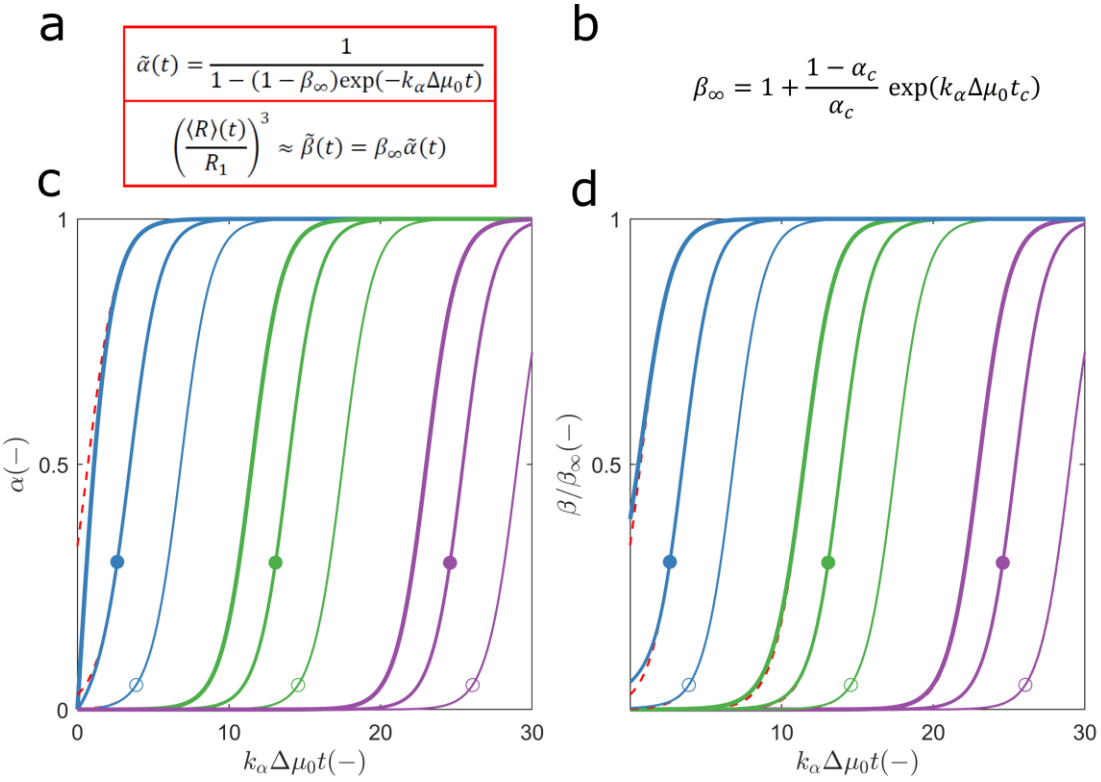

Figure 3. Exact and approximate solutions of Eq. 5. (a) The approximate $\tilde{\alpha}(t)$ and $\tilde{\beta}(t)$ expressions are derived by assuming a constant number of particles. (b) From this definition of the parameter β_∞, it follows that the expression for $\tilde{\alpha}(t)$ and $\tilde{\beta}(t)/\beta_\infty$ in (a) is equivalent to the exact $\alpha(t)$ solution derived for $t > t_c$ (Fig. 2a). (c) Solid lines: exact $\alpha(t)$ profiles computed as described in Fig. 2c using k_β values of (different colours from left to right) 0.5, 10^{-5}, and 10^{-10}, and α_c values of (decreasing order of line thickness) 1, 0.30 and 0.05. Dashed red lines (overlapped when not seen): $\tilde{\alpha}(t)$ profiles

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

calculated from the definitions in (a) and (b) using the same values of k_β and α_c as in the exact solution. Symbols: location of the critical coordinates t_c and α_c. (d) Comparison between the exact $\beta(t) = N(t)/N_1$ (solid lines) and approximate $\tilde{\beta}(t)$ (dashed red lines) solutions calculated using the same model parameters as in (c).

The differences between the exact and approximate solutions are more pronounced during the initial periods of fast nucleation processes and, as expected, are null during growth-only periods, independently of the considered value of k_β (Figs. 3c and 3d). The application of the $\tilde{\alpha}(t)$ expression to the early stages ($t < t_c$) of hyperbolic progress curves is disqualified by the high errors associated with this approximation (Fig. S5a). On the contrary, the $\tilde{\beta}(t)$ approximation adequately describes the generality of particle growth curves (Fig. S5b). In the worst possible scenario, the $\tilde{\beta}(0)/\beta_\infty$ estimates give approximate errors of -8% or $+8\%$ for values of $k_\beta \approx 10^{-1}$ or $k_\beta \approx 10$, respectively, if STEs are absent (Fig. S5b, inset). In all cases, initial differences between $\beta(t)$ and $\tilde{\beta}(t)$ vanish as phase separation progresses (Fig. 3d).

Kinetic Analysis of LLPS

The new insights provided by the general nucleation-and-growth model are first illustrated using LLPS kinetic data obtained for the low-complexity domain (LCD) of heterogeneous nuclear ribonucleoprotein A2 (hnRNPA2), the TAR DNA-binding protein 43 (TDP-43), the nuclear pore complex protein NUP98 and the Early Responsive to Dehydration 14 (ERD14) stress protein [16]. These aggregation-prone proteins can be kept stable in solution at a carefully selected extreme pH and then undergo LLPS by imposing native pH conditions through the addition of a small amount of concentrated buffer [16]. Particle growth curves measured by dynamic light scattering (DLS) reveal that droplets with initial sizes > 100 nm grow at different rates in the presence and the absence of 150 mM NaCl (Fig. 4). As an alternative to empirical $\langle R \rangle$ vs. t scaling laws, the $\tilde{\beta}(t)$ expression in Fig. 3a can be used to fit the measured data considering that $\langle R \rangle(t) = R_1 \tilde{\beta}(t)^{1/3}$ for invariant droplet shapes. Parameters β_∞ and $k_\alpha \Delta\mu_0$ are numerically fit to the experimental $\langle R \rangle(t)$ curves using the observed initial size as an estimate of $\langle R_1 \rangle$. When the final size $\langle R \rangle(\infty)$ is also known, the value of β_∞ can be estimated in advance from the relationship $\beta_\infty^{1/3} = \langle R \rangle(\infty)/R_1$. This combined parameter is affected both by α_c and k_β and represents the number of times the droplet volume is ultimately increased (Fig. 3b).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

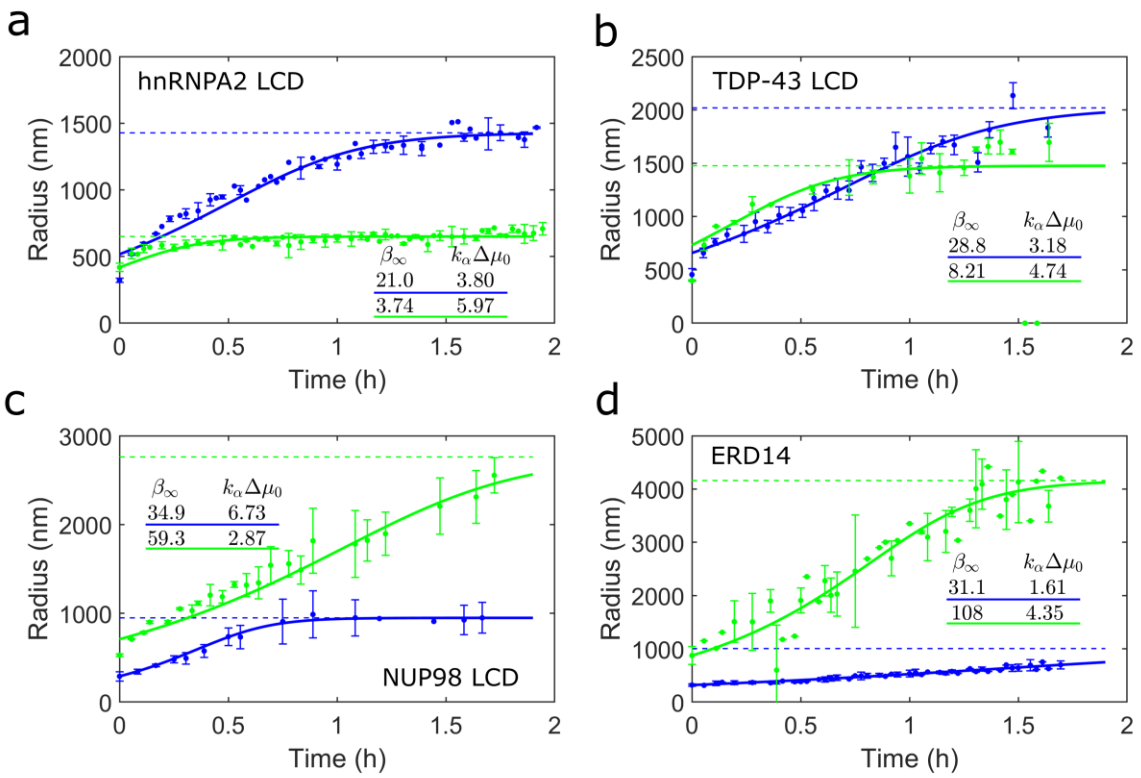

Figure 4. Kinetic analysis of LLPS of proteins (a) 20 µM hnRNPA2 LCD, (b) 80 µM TDP-43 LCD, (c) 10 µM NUP98 LCD and (d) 20 µM ERD14. Symbols: Droplet size evolution measured by Van Lindt et al. [16] in the absence (blue) and presence (green) of 150 mM NaCl [16]. Solid lines: fitted $\langle R\rangle(t)$ curve computed as the product $R_1\tilde{\beta}(t)^{1/3}$ using the values of β_∞ and $k_\alpha\Delta\mu_0$ listed in each panel's table. Dashed lines: predicted value of the final mean size $\langle R\rangle(\infty) = R_1\beta_\infty^{1/3}$.

To ascertain the reasons why larger or smaller droplets are produced in the different experiments it is necessary to know the specific values α_c and k_β that lead to large or small values of β_∞, respectively. In the present example, approximate values of α_c can be determined using estimates of the saturation concentration c_∞ and critical solubility c_c extracted from the 600 nm-turbidity increase that was induced by different protein concentrations [16]. Since LLPS does not spontaneously occur for $c_0 < c_c$, decreasing values of protein concentration eventually reach a point of $c_0 \approx c_c$ below which the initial turbidity signal changes no more. Accordingly, values of c_c could be identified (Table 1) from the protein titration results in Figure S9 of the source manuscript [16]. The same experimental data is used to estimate c_∞ values from the x-intercept of the maximum turbidity $\Delta F_{\max}$ vs. c_0 straight line given that the final amount of the dense phase is proportional to the concentration difference $\Delta c_0 = c_0 - c_\infty$ [8,9]. As expected for aggregation-prone proteins, very low values of the saturation concentration are obtained for all proteins under study ($c_\infty \approx 0$). Once c_∞ and c_c estimates are

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

available, STEs can be quantified as STEs $= 1 - \alpha_c$ using the definition of α_c (Eq. 4). The analysis of the $\tilde{\beta}(t)$ solution in Figs. 3d and S5 shows that $\tilde{\beta}(t_c) = \alpha_c$ is a good approximation for all combinations of α_c and k_β. Thus, our next step in the characterization of model parameters is the identification of the instant t_c at which the equivalence $\langle R\rangle(t_c)/R_1 = \alpha_c^{1/3}$ is verified (Fig. 4). Then, the definitions of time constants τ and t_1 in Fig. 2a can be expressed as a function of parameters k_β, α_c and $k_\alpha\Delta\mu_0$, and replaced in the definition of t_c to obtain the k_β estimate (Table 1). With the k_β results in hand, values of τ and t_1 are finally obtained using their mathematical definitions in Fig. 2b.

Table 1. Model parameters estimated from the LLPS data measured by Van Lindt et al. [16] and from the fitted $k_\alpha\Delta\mu_0$ and β_∞ values in Fig. 4 for the cases of no added NaCl.

Protein	c_0 (μM)	c_c (μM)	α_c (–)	t_c (h)	R_1 (nm)	$k_\alpha\Delta\mu_0$ (h^{-1})	β_∞ (–)	k_β (–)	$\frac{k_\alpha\Delta\mu_0\tau}{2}$ (–)
hnRNPA2 LCD	20	2	0.90	1.333	517.2	3.80	21.0	0.074	1.01
TDP-43 LCD	80	20	0.75	1.317	658.5	3.18	28.8	0.090	1.04
NUP98 LCD	10	3.8	0.62	0.598	290.2	6.73	34.9	0.086	1.04
ERD14	20	10	0.50	1.283	319.3	1.61	31.1	0.908	3.29

Of all proteins and conditions considered in Table 1, hnRNPA2 LCD is the one less affected by STEs because the adopted initial concentration is much higher than c_c. In the absence of NaCl, the mean size increase $\langle R\rangle(\infty)/R_1$ experimentally measured for this protein was higher than for ERD14, the protein more strongly affected by STEs – compare Figs. 4a and 4d. This observation is explained by lower nucleation-to-growth ratios (lower k_β values) exhibited by the first protein. However, if LLPS had been followed over a longer period, a ~31-fold increase in droplet volume would be expected for ERD14, a value that is larger than the value of β_∞ obtained for hnRNPA2 LCD (Table 1). The case of ERD14 illustrates the possibility of fast nucleating droplets ($k_\beta = 0.91$) that slowly grow into micrometer-sized particles due to the existence of a growth-only period after $t = t_c$. In contrast to what happens for the other proteins, the τ values listed in Table 1 for ERD14 are much larger than $2(k_\alpha\Delta\mu_0)^{-1}$ as a consequence of more pronounced STEs and of faster nucleation rates. The reason why this protein takes more time to complete LLPS is explained by a slow growth step and by the increased importance of this step in the presence of STEs. As illustrated below in the Discussion section, strong STEs (low α_c) and fast nucleation rates (high k_β) uniquely combine to allow for functional phase separation via sharply regulated responses to minute concentration changes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Interestingly, the combination of stronger STEs (lower α_c) and higher k_β is observed for ERD14 (Table 1), which is an intrinsically disordered chaperone with the function of helping plants to survive under dehydration stress [39,40], whereas hnRNPA2, TDP-43 and NUP98 occur in bimolecular condensates associated with disease [16,41].

Predictive Power

We now investigate whether the two-parameter physical model can quantitatively predict the evolution of size distributions during the phase separation of TDP-43 LCD and NUP98 LCD. For these proteins, the measured DLS autocorrelation functions (Fig. S6) yield reliable size distributions in all stages of LLPS (Figs. 5a and 5b). To compare the experimental data with the model predictions, light scattering intensities are simulated as a function of time and droplet size using the zeroth $P(t)$, first $M(t)$ and second $Q(t)$ moments of the size distribution, from which the mean size $N(t)$ (first cumulant) and the variance $\sigma^2(t)$ (second cumulant) are determined (Appendix: *Size Distributions*). This strategy is an alternative to the recursive integration of the master equation (Eq. S2) over the possible values of $j = N_1, N_1 + 1, \ldots, N_\infty$, a method that becomes inpractical for the droplet sizes in question.

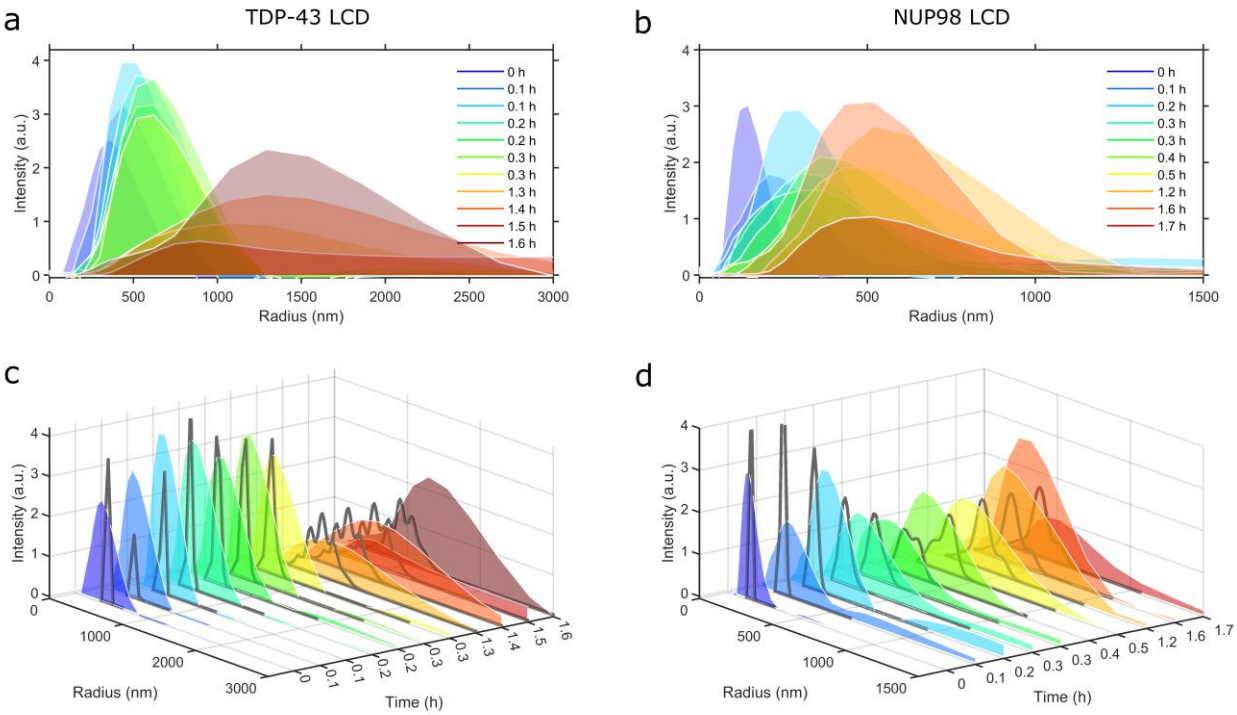

Figure 5. Time variation of the droplet size distributions. (a and b) DLS intensity distributions obtained for (a) TDP-43 LCD and (b) NUP98 LCD. (c and d) Gray lines: light scattering intensities predicted by the general nucleation-and-growth model for (c) TDP-43 LCD and (d) NUP98 LCD using the values of α_c, $k_\alpha \Delta\mu_0$ and k_β listed in Table 1. Further numerical details in *SI Additional Methods*.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Different colored curves: same experimental results as in (a) and (b).

The computed curves follow a similar trend to the one observed for the mean size, shape and scattering intensity of the DLS distributions (Figs. 5c and 5d). This fact is all the more remarkable when we consider that no numerical adjustments were required besides the choice of a fixed proportionality factor normalizing the scattering intensities during each experiment and that automatic Mie scattering corrections are applied to the simulated size distributions. Moreover, the droplet sizes are markedly distinct in the cases of TDP-43 LCD and NUP98 LCD and spread over distinct range values as LLPS proceeds. The combination of results in Figs. 4 and 5 illustrates that the general nucleation-and-growth model has the potential to characterize, and even predict intricate aspects of phase separation using the α_c measurable and two kinetic parameters.

Discussion

The sub-critical region of the phase diagram opens surprising avenues for deciphering the origins of intriguing biological processes, from pathogenic protein aggregation to the spatiotemporal control of condensate formation in the cell. Proteins can be stable and functional under supersaturated conditions without the immanent risk of undergoing primary nucleation and forming potentially toxic aggregates. In this light, the association of protein misfolding disorders with ageing would not result from the fact that critical macromolecules can be naturally supersaturated [42], but would rather reflect the increasing chances of the critical supersaturation barrier being crossed as the individual grows old. Such transgressions are likely to occur as the result of, for example, incidental cell stress, chemical modifications of the protein and altered proteostasis. If the nuclei resist cellular quality control mechanisms, a single species might be enough to elicit a cascade of events comprising the secondary nucleation of new aggregates and the cell-to-cell spreading of the disease in a prion-like fashion [43]. Another possibility opened by STEs is the formation of large and uniformly distributed particles without recourse to coalescence or external players such as enzymes and chaperones. This possibility was hard to conceive from a purely biophysical point of view: either primary nucleation is fast and steady-state conditions are rapidly achieved through the formation of numerous small particles, or primary nucleation is slow relative to the growth step and non-uniform size distributions are obtained with very large particles coexisting with very small ones (Figs. 1c and 1d, Movie S1). The occurrence of large and uniformly distributed condensates in the cell has been enigmatic and is often ascribed to some unknown or not fully outlined regulatory process(es). However, the observation of the same pattern in test-tube experiments demonstrates that STEs are essential for the understanding of biological phase separations. Narrow size distributions are expected in the presence of strong STEs

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

because the occurrence of primary nucleation becomes limited to the initial moments of phase separation (Figs. S3c and S3d and Movie S2). Splitting nucleation-and-growth into two consecutive steps facilitates the production of large condensates with precisely controlled size distributions.

By deliberately simplifying our analysis to phase separation of binary mixtures, we demonstrate that size heterogeneity is not the inexorable result of nucleation-and-growth processes which, on the contrary, can be spontaneously regulated according to basic physicochemical principles of phase equilibrium. This said, considering STEs at the quantitative level is particularly useful for understand how multicomponent LLPS can be accurately controlled in the cell through the modulation of interfacial properties. Being interfacially active, crucial proteins, such as G3BP1 in stress granules, MEG-3 in P granules, centrosomin and DSpd-2 in the centrosome, and TPX-2 in the mitotic spindle, can induce a generalized solubility drop in the lighter phase that provokes the recruitment of all molecule whose concentration is above the critical limit. Surface tension modulation, therefore, emerges as a general organizing principle regulating phase separations in health and disease.

We envisage a practical application of our biophysical model in drug discovery, namely on the quantification of the effects of STEs-modulators and kinetic inhibitors on the progress curves of biological self-assembly. These curves can be measured in the presence and absence of drug candidates and then fitted to the exact $\alpha(t)$ solution (Fig. 2a) or the approximate $\langle R\rangle(t)$ solution (Fig. 3a) depending if mass-based or size-based data are available. Deviations from the theoretical $\langle R\rangle(t)$ curve (Fig. 3a) are expected if significant secondary nucleation occurs. In that case, the exact numerical solution of Eq. 5 or the approximate $\langle R\rangle(t)$ solution assuming no STEs present [9] can be adopted in the alternative. In principle, the $\alpha(t)$ solution (Fig. 2a) is widely applicable independently of the magnitudes of STEs and secondary nucleation rates. Size distribution analysis may additionally reveal the continuum transition from the monomeric state, to mesoscale clusters, to the formation of a new liquid phase [44]; when this happens, sharp nucleation-and-growth models do not apply and should be replaced by oligomerization equilibrium analysis.

Quantifying the effect of kinetic inhibitors on the rates of growth, primary nucleation, and secondary nucleation is well-established in protein misfolding research [7,45], but it is almost absent from biological LLPS research where chemical kinetic analysis is generally based on semi-empirical scaling laws and exponents. In addition to the $\alpha(t)$ and $\langle R\rangle(t)$ solutions here derived, we also proposed novel methodologies to quantify STEs from the measurements of the saturation concentration c_∞ and critical solubility c_c (see analysis to Fig 4). We showed, moreover, that the discrete size distributions of biological condensates can be inferred at the quantitative level using kinetic parameters fitted to $\alpha(t)$ or $\langle R\rangle(t)$ progress curves (Fig. 5). This opens new possibilities for gauging the effect of phase

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

separation modulators from measured values of equilibrium concentration and size-distribution parameters.

Conclusions

Biological phase separation is associated with multiple aspects of cell physiology and may be conducive to neurodegenerative processes. How, when and where this phenomenon occurs remains difficult to predict even under test-tube experimental conditions. Here we propose a physical description of protein phase separation that explains intriguing observations such as the spontaneous formation of a few condensates with large and uniform sizes. We show that the caveats in the current nucleation-and-growth theory can be solved by addressing the role of surface tension during the formation of primary nuclei. Shape- and size-dependent STEs provoke an increase in the thermodynamic solubility, which is reflected in higher energetic barriers for primary nucleation. A "margin of safety" is therefore introduced so that the birth of the new phase becomes thermodynamically unfavored in metastable regions of the phase diagram. The difference between the saturation limit and the critical solubility can be interpreted in the light of the Young-Laplace equation for curved interface equilibria. These principles apply to liquid-solid and LLPS alike and are included in the master equation describing nucleation-and-growth phenomena (Eqs. 5 and S2), whose closed-form solution is here derived (Fig. 2a). In most situations, the exact solution simplifies to an exponential-type function describing how the average particle size changes over time (Fig. 3a). The model predictions are validated using kinetic data measured for different proteins during the phase separation of different-sized biological condensates. We found that the final size, as well as the moment of formation of biological condensates become highly regulated in the presence of strong STEs and that the evolution of droplet size distributions can be predicted based on the α_c measurable and two kinetic parameters (Figs. 5c and 5d). The generalized nucleation-and-growth model can be used for chemical kinetics analysis of isolated systems, but it could also be extended over complex, multicomponent systems for interrogating mechanisms of cell regulation and neurotoxicity involving phase separation.

Materials and Methods

Model Validation: The general nucleation-and-growth model and its validity are tested using kinetic data of LLPS (Fig. 4) and amyloid aggregation (Fig. S1), by rationalizing the formation of phase-separated condensates with large and uniform sizes under conditions where active biological processes are not present (Fig. S3d and Movies S1 and S2), and by comparing the predicted time evolution of droplet sizes with the measured one (Figs. 5c and 5d). The raw data for graphs presented in Fig. 4 are

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

available in the Supplementary Data files of reference [16]. The raw data presented in Fig. S1 were digitized by us from the original reference [10]. Details of the adopted numerical methods in *SI Additional Methods*.

DLS measurements: Measurements of DLS autocorrelation functions were carried out on a DynaPro NanoStar (Wyatt) instrument during LLPS of 80 μM TDP-43 LCD and 10 μM NUP98 LCD induced by pH jumps from 3.5 to 7.5 (TDP-43 LCD) and 3.0 to 7.5 (NUP98 LCD). A disposable cuvette (WYATT technology) was filled with 100 μl of protein solution at pH and concentration values at which LLPS occurred. The sides of the cuvette were filled with water and a cap was put on top. The samples were illuminated by a 120 mW air-launched laser of 658 nm wavelength intensity and the scattered light was recorded at a scattering angle of 90° at 25 °C, for a period of 6 h, collecting 10 acquisitions (8 s each). Each measurement was repeated at least 3 times. Hydrodynamic radii of the particles in solution were estimated from the diffusion coefficient delivered from the analysis of measured autocorrelation functions as described in Figs. S6a and S6b.

Supporting Information

- Appendix
- Additional Methods
- Additional Discussion
- Additional Figures
- SI References.
- Movies S1 and S2

Acknowledgements

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952334 (PhasAGE). This research was financed by (i) FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operational Programme for Competitiveness and Internationalization (POCI), Portugal 2020, and by Portuguese funds through FCT—Fundação para a Ciência e a Tecnologia/ Ministério da Ciência, Tecnologia e Ensino Superior (FCT/MCTES) in the framework of projects PTDC/QUI-COL/2444/2021, POCI-01-0145-FEDER-031323 (PTDC/MED-FAR/31323/2017), and POCI-01-0145-FEDER-007274 ("Institute for Research and Innovation in Health Sciences"), (ii) by Base Funding– UIDB/00511/2020 of the Laboratory for Process Engineering, Environment, Biotechnology and Energy (LEPABE) funded by national funds through the FCT/MCTES (PIDDAC).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

[1] Y. Shin, C. P. Brangwynne, *Science* **2017**, *357*, eaaf4382.

[2] P. M. Martins, S. Navarro, A. Silva, M. F. Pinto, Z. Sárkány, F. Figueiredo, P. J. B. Pereira, F. Pinheiro, Z. Bednarikova, M. Burdukiewicz, et al., *Front. Mol. Neurosci.* **2020**, *13*, 582488.

[3] R. Crespo, E. Villar-Alvarez, P. Taboada, F. A. Rocha, A. M. Damas, P. M. Martins, *J. Biol. Chem.* **2016**, *291*, 2018.

[4] R. Crespo, F. A. Rocha, A. M. Damas, P. M. Martins, *J. Biol. Chem.* **2012**, *287*, 30585.

[5] M. A. Watzky, R. G. Finke, *J. Am. Chem. Soc.* **1997**, *119*, 10382.

[6] M. A. Watzky, E. E. Finney, R. G. Finke, *J. Am. Chem. Soc.* **2008**, *130*, 11959.

[7] S. I. A. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, T. P. J. Knowles, *Proc. Natl. Acad. Sci.* **2013**, *110*, 9758.

[8] A. Silva, Z. Sárkány, J. Fraga, P. Taboada, S. Macedo-Ribeiro, P. Martins, *Biomolecules* **2018**, *8*, 108.

[9] A. Silva, B. Almeida, J. S. Fraga, P. Taboada, P. M. Martins, S. Macedo-Ribeiro, *Angew. Chem. Int. Ed.* **2017**, *56*, 14042.

[10] A. R. Hurshman, J. T. White, E. T. Powers, J. W. Kelly, *Biochemistry* **2004**, *43*, 7365.

[11] P. Hortschansky, V. Schroeckh, T. Christopeit, G. Zandomeneghi, M. Fändrich, *Protein Sci.* **2005**, *14*, 1753.

[12] A. Henning-Knechtel, S. Kumar, C. Wallin, S. Król, S. K. T. S. Wärmländer, J. Jarvet, G. Esposito, S. Kirmizialtin, A. Gräslund, A. D. Hamilton, et al., *Cell Rep. Phys. Sci.* **2020**, *1*, 100014.

[13] D. Zwicker, M. Decker, S. Jaensch, A. A. Hyman, F. Jülicher, *Proc. Natl. Acad. Sci.* **2014**, *111*, E2636.

[14] J. Berry, C. P. Brangwynne, M. Haataja, *Rep. Prog. Phys.* **2018**, *81*, 046601.

[15] Q. Bai, Q. Zhang, H. Jing, J. Chen, D. Liang, *J. Phys. Chem. B* **2021**, *125*, 49.

[16] J. Van Lindt, A. Bratek-Skicki, P. N. Nguyen, D. Pakravan, L. F. Durán-Armenta, A. Tantos, R. Pancsa, L. Van Den Bosch, D. Maes, P. Tompa, *Commun. Biol.* **2021**, *4*, 77.

[17] H. Le Ferrand, M. Duchamp, B. Gabryelczyk, H. Cai, A. Miserez, *J. Am. Chem. Soc.* **2019**, *141*, 7202.

[18] B. A. Gibson, L. K. Doolittle, M. W. G. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, M. K. Rosen, *Cell* **2019**, *179*, 470.

[19] H. Zhang, S. Elbaum-Garfinkle, E. M. Langdon, N. Taylor, P. Occhipinti, A. A. Bridges, C. P. Brangwynne, A. S. Gladfelter, *Mol. Cell* **2015**, *60*, 220.

[20] S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, M. K. Rosen, *Cell* **2016**, *166*, 651.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[21] S. Alberti, A. Gladfelter, T. Mittag, *Cell* **2019**, *176*, 419.

[22] J.-M. Choi, A. S. Holehouse, R. V. Pappu, *Annu. Rev. Biophys.* **2020**, *49*, 107.

[23] A. E. Posey, A. S. Holehouse, R. V. Pappu, in *Methods Enzymol.*, Elsevier, **2018**, pp. 1–30.

[24] C. Haas, J. Drenth, *J. Cryst. Growth* **1999**, *196*, 388.

[25] G. L. Dignon, W. Zheng, Y. C. Kim, J. Mittal, *ACS Cent. Sci.* **2019**, acscentsci.9b00102.

[26] P. G. Vekilov, *Soft Matter* **2010**, *6*, 5254.

[27] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, J. P. Taylor, *Cell* **2015**, *163*, 123.

[28] H. Falahati, A. Haji-Akbari, *Soft Matter* **2019**, *15*, 1135.

[29] A. Bratek-Skicki, R. Pancsa, B. Meszaros, J. Van Lindt, P. Tompa, *FEBS J.* **2020**, *287*, 1924.

[30] P. Bennema, J. Vaneupen, B. Vanderwolf, J. Los, H. Meekes, *Int. J. Pharm.* **2008**, *351*, 74.

[31] D. Zwicker, A. A. Hyman, F. Jülicher, *Phys. Rev. E* **2015**, *92*, 012317.

[32] Y. Shin, Y.-C. Chang, D. S. W. Lee, J. Berry, D. W. Sanders, P. Ronceray, N. S. Wingreen, M. Haataja, C. P. Brangwynne, *Cell* **2018**, *175*, 1481.

[33] J. D. Camino, P. Gracia, S. W. Chen, J. Sot, I. de la Arada, V. Sebastián, J. L. R. Arrondo, F. M. Goñi, C. M. Dobson, N. Cremades, *Chem. Sci.* **2020**, *11*, 11902.

[34] S. F. Shimobayashi, P. Ronceray, D. W. Sanders, M. P. Haataja, C. P. Brangwynne, *Nature* **2021**, *599*, 503.

[35] J. A. W. Elliott, *J. Phys. Chem. B* **2020**, *124*, 10859.

[36] J. Nývlt, *J. Cryst. Growth* **1968**, *3–4*, 377.

[37] D. Kashchiev, G. M. van Rosmalen, *Cryst. Res. Technol.* **2003**, *38*, 555.

[38] A. K. Buell, in *Int. Rev. Cell Mol. Biol.*, Elsevier, **2017**, pp. 187–226.

[39] N. Murvai, L. Kalmar, B. Szalaine Agoston, B. Szabo, A. Tantos, G. Csikos, A. Micsonai, J. Kardos, D. Vertommen, P. N. Nguyen, et al., *Cells* **2020**, *9*, 1856.

[40] P. N. Nguyen, M.-A. Tossounian, D. S. Kovacs, T. T. Thu, B. Stijlemans, D. Vertommen, J. Pauwels, K. Gevaert, G. Angenon, J. Messens, et al., *Biochim. Biophys. Acta BBA - Gen. Subj.* **2020**, *1864*, 129506.

[41] S. Terlecki-Zaniewicz, T. Humer, T. Eder, J. Schmoellerl, E. Heyes, G. Manhart, N. Kuchynka, K. Parapatics, F. G. Liberante, A. C. Müller, et al., *Nat. Struct. Mol. Biol.* **2021**, *28*, 190.

[42] G. Vecchi, P. Sormanni, B. Mannini, A. Vandelli, G. G. Tartaglia, C. M. Dobson, F. U. Hartl, M. Vendruscolo, *Proc. Natl. Acad. Sci.* **2020**, *117*, 1015.

[43] E. W. Martin, T. S. Harmon, J. B. Hopkins, S. Chakravarthy, J. J. Incicco, P. Schuck, A. Soranno, T. Mittag, *Nat. Commun.* **2021**, *12*, 4513.

[44] M. Kar, F. Dar, T. J. Welsh, L. Vogel, R. Kühnemuth, A. Majumdar, G. Krainer, T. M.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Franzmann, S. Alberti, C. A. M. Seidel, et al., *Phase Separating RNA Binding Proteins Form*
2 *Heterogeneous Distributions of Clusters in Subsaturated Solutions*, Biophysics, **2022**.
3 [45] Z. Sárkány, F. Rocha, A. M. Damas, S. Macedo-Ribeiro, P. M. Martins, *Chem. - Asian J.* **2019**,
4 DOI 10.1002/asia.201801703.
5

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

Quantification of surface tension effects and nucleation-and-growth rates during self-assembly of biological condensates

Zsuzsa Sárkány, Fernando Rocha, Anna Bratek-Skicki, Peter Tompa, Sandra Macedo-Ribeiro and Pedro M. Martins

- Appendix
    - General Nucleation-and-Growth Model
    - Size Distributions
- Additional Methods
    - Numerical Methods
- Additional Discussion
    - Kinetic Analysis of Amyloid Aggregation
    - STEs on phase diagrams and particle size distributions
- Additional Figures
    - Figure S4. The impact of secondary nucleation on the final particle size in the absence of STEs.
    - Figure S5. Maximum error of the $\tilde{\alpha}(t)$ and $\tilde{\beta}(t)$ approximations (Fig. 3a) relatively to the exact solution of Eq. 5.
    - Figure S6. DLS autocorrelation functions during LLPS of (a) TDP-43 LCD and (b) NUP98 LCD.
    - Figure S7. Using the gamma distribution to describe the exact solution of the master model equation (Eq. S2).
- SI References.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

Appendix

General Nucleation-and-Growth Model: In previous master equations describing nucleation-and-growth of amyloid fibrils, the concentration of filaments composed of $j \geq N_2$ monomers (C_j) is determined by the rate constants k_1, k_2, k_+ and k_- characterizing the processes of primary and secondary nucleation, growth, and fragmentation, respectively [1,2]:

$$\begin{aligned}\frac{\mathrm{d}C_j(t)}{\mathrm{d}t} &= \frac{1}{j}\left[k_1\Delta\mu^2(t)\delta_{j,N_1} + k_2\Delta\mu(t)M(t)\delta_{j,N_2}\right] + \\ &+k_+\Delta\mu(t)(j-1)C_{j-1}(t) - k_+\Delta\mu(t)jC_j(t) + \\ +\frac{1}{j}&\left[-k_-C_j(t)(j-N_2)\mathcal{H}(j-2N_1-1) + \sum_{i=j+1}^{\infty} k_-C_i(t)\mathcal{H}(j-2N_1-1)\right]\end{aligned} \tag{S1}$$

where the Kronecker delta function sets the sizes of the primary and secondary nuclei and the Heaviside function establishes a minimum size of $2N_1 + 1$ molecules above which fragmentation starts to occur [2]. We assume that STEs mainly affect the driving force for primary nucleation and thus only the k_1 term has to be expressed in relation to the critical supersaturation level ($\Delta\mu_c$):

$$\begin{aligned}\frac{\mathrm{d}C_j(t)}{\mathrm{d}t} &= \frac{1}{j}\left[k_1\Delta\mu_c^2(t)\delta_{j,N_1} + k_2\Delta\mu(t)M(t)\delta_{j,N_2}\right] + \\ &+k_+\Delta\mu(t)(j-1)C_{j-1}(t) - k_+\Delta\mu(t)jC_j(t) + \\ +\frac{1}{j}&\left[-k_-C_j(t)(j-N_2)\mathcal{H}(j-2N_1-1) + \sum_{i=j+1}^{\infty} k_-C_i(t)\mathcal{H}(j-2N_1-1)\right]\end{aligned} \tag{S2}$$

The number and mass concentrations correspond to the zeroth and first moments of the $C_j(t)$ distribution:

$$P(t) = \sum_j C_j(t) \tag{S3a}$$

$$M(t) = \sum_j jC_j(t) \tag{S3b}$$

and are used to obtain Eqs. 5a and 5b after differentiation, replacement of Eq. S2 and algebraic manipulation of the summations; to obtain Eq. 5a, negligible fragmentation rates are assumed, whereas in Eq. 5b the terms dependent on k_- ultimately cancel each other. Furthermore, from the definitions of $\alpha(t) = M(t)/M_\infty$, $M_\infty = c_0 - c_\infty$ and $\Delta\mu_c(t)/\Delta\mu_0$ (Eq. 3), constants k_α and k_β can be defined as a function of the elementary rate constants as follows:

$$k_\alpha = (k_+ + k_2) \tag{S4a}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$k_\beta = \frac{k_1}{c_\infty(k_+ + k_2)}\left(\frac{c_\infty}{c_c}\right)^2 \tag{S4b}$$

In the absence of STEs, $c_c = c_\infty$ and $\Delta\mu_c = \Delta\mu$, which simplifies Eqs. 5a and 5b to a system of ODEs of known analytical solution [1, 2]:

$$P(t) = \frac{M_\infty \alpha(t)}{N_1}\frac{k_b}{1-k_b}\left[\frac{\ln\big(1-\alpha(t)\big) + k_\alpha \Delta\mu_0 t}{\alpha(t)(1-k_b)}\left(1 - \frac{k_2}{k_\alpha}\frac{N_1}{N_2}\right) - \left(1 - \frac{k_2}{k_\alpha k_b}\frac{N_1}{N_2}\right)\right] \tag{S5a}$$

$$\alpha(t) = 1 - \frac{1}{k_b[\exp(k_\alpha \Delta\mu_0 t) - 1] + 1} \tag{S5b}$$

where $k_b = k_n/(k_+ + k_2)$, with $k_n = k_1/c_\infty$ [1, 2].

In the presence of STEs, Eq. 5b can be solved isolatedly to obtain the following analytical solution:

$$\alpha(t) = \begin{cases} \dfrac{\tanh\left(\frac{t_1}{\tau}\right) + \tanh\left(\frac{t}{\tau} - \frac{t_1}{\tau}\right)}{\tau \boldsymbol{k_\alpha} \Delta\mu_0 (1 - \boldsymbol{k_\beta})} & \text{for } t \le t_c \\ \dfrac{1}{1 + \frac{1-\alpha_c}{\alpha_c}\exp\big(-\boldsymbol{k_\alpha}\Delta\mu_0(t - t_c)\big)} & \text{for } t > t_c \end{cases} \tag{S6}$$

where

$$t_1 = \tau\, \mathrm{atanh}\left[\boldsymbol{k_\alpha}\Delta\mu_0\frac{\tau}{2}(1 - 2\boldsymbol{k_\beta}\alpha_c)\right] \tag{S7a}$$

$$t_c = t_1 + \tau\, \mathrm{atanh}\left(\boldsymbol{k_\alpha}\Delta\mu_0\frac{\tau}{2}(2\alpha_c - 1)\right) \tag{S7b}$$

$$\tau = \frac{2}{\boldsymbol{k_\alpha}\Delta\mu_0\sqrt{1 - 4\boldsymbol{k_\beta}\alpha_c(1-\alpha_c)}} \tag{S7c}$$

Therefore, Eq. S5b is the limit case of Eq. S6 when $\alpha_c = 1$ (no STEs present).

Size Distributions. The particle size distributions (PSDs) predicted by the general nucleation-and-growth model are obtained by recursive integration of Eq. S2 over the possible values of j. This method cannot be always adopted since heavy computational requirements are required when larger particles are considered: in the exemplary case of a 10 kDa globular protein, a small spherical droplet of ~100 nm radius would contain $j > 3 \times 10^5$ monomeric units. As an alternative, the mean size $N(t)$ and variance $\sigma^2(t)$ of the $C_j(t)$ distribution can be determined as a function of the principal moments using the definitions of the first and second cumulants:

$$N(t) = \frac{M(t)}{P(t)} \tag{S8a}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\sigma^2(t) = \frac{Q(t)}{P(t)} - \left[\frac{M(t)}{P(t)}\right]^2 \tag{S8b}$$

The definition of the second moment,

$$Q(t) = \sum_j j^2 C_j(t) \tag{S9}$$

is differentiated and used together with Eq. S2 and the $P(t)$, $M(t)$ sumations (Eq. S3) to obtain:

$$\frac{1}{k_\alpha \Delta\mu_0}\frac{dQ(t)}{dt} = N_1 M_\infty k_\beta \left(\frac{\Delta\mu(t)}{\Delta\mu_0} + \alpha_c - 1\right)^2 + \frac{k_2}{k_\alpha}\frac{\Delta\mu(t)}{\Delta\mu_0} N_2 M(t) + \frac{k_+}{k_\alpha}\frac{\Delta\mu(t)}{\Delta\mu_0}(2Q(t) + M(t) - 3P(t) + 1) \tag{S10}$$

Therefore, an exact solution for the zeroth, first and second moments is possible by solving Eqs. 5a, 5b and S10 simultaneously. As described in detail in *SI Additional Methods*, the mean and variance of $C_j(t)$ are used to predict the shape of droplet size distributions measured by DLS (Figs. 6c and 6d of the main text).

Additional Methods

Numerical Methods: The model equations are fitted to experimental data in Figs. 4 and S1 using the *lsqcurvefit* function of Mathworks MATLAB R2020b (Natick, MA, USA). The PSDs in Figs. 1d and S3d are generated by numerically solving Eq. S2 for long reaction times using the model parameters identified in each figure's caption assuming that $N_1 = 2$ monomers, that no condensates are initially present $C_j(0) = 0$, and a total of $M_\infty = 1000$ monomers separating into the new phase. Each particle size distribution is normalized by the maximum frequency value and represented as a function of $R_j/R_1 = \left(N_j/N_1\right)^{1/3}$. In Movies S1 and S2, the number of condensates represented in each timeframe coincides with the total number of particles calculated per unit volume; condensates composed of N_j monomers are schematically represented as circles with symbol size N_j.

To compute the $\alpha(t)$ progress curves (Figs. 1c and S3c), the system of ordinary differential equations (ODEs) comprising Eqs. 5 and Eq. S2 was numerically solved and the results expressed in $k_\alpha \Delta c_0$-normalized units of time.

The theoretical size distributions in Figs. 6c and 6d are generated from the solution of Eqs. 5a, 5b and S10 for the values of the α_c, $k_\alpha\Delta\mu_0$ and k_β parameters listed in Table 1 for TDP-43 LCD and

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

NUP98 LCD. Negligible secondary nucleation ($k_2 \approx 0$) and the absence of initial condensates are assumed; the used R_1 values are inferred from the initial DLS distribution (450 nm for TDP-43 LCD and 160 nm for NUP98 LCD). Replacing the $P(t)$, $M(t)$ and $Q(t)$ solutions in Eq. S8 gives the variation with time of the mean size and variance of the distributions. We verified that the gamma probability distribution adequately describes the gradual change from exponential to normal distribution predicted by the master model equation Eq. S2 (Fig. S7). The approximate $C_j(t)$ profiles are obtained from the gamma distribution function using as parameters $\theta(t) = \sigma^2(t)/N(t)$ and $k(t) = N(t)/\theta(t)$ [3]. Finally, each droplet is modelled as a Mie scatterer, whose scattering intensity is calculated from the complex scattering amplitudes for two orthogonal directions of incident polarization. The MATLAB code *MatScat* [4] was used for this purpose assuming: a spherical geometry, collection of scattered light between scattering angles of 80 and 100 degrees, a complex refractive index of $1.39 + 0.01i$, a outer-medium refractive index of 1.333, a wavelength of 658 nm, and a proportionality factor of 1.11×10^{16} (TDP-43 LCD) and 1.00×10^{16} (NUP98 LCD) to normalize scattering intensity units.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

Additional Discussion

Kinetic Analysis of Amyloid Aggregation: The kinetics of amyloid aggregation of monomeric transthyretin (mTTR) was fully characterized by Hurshman et al. from the increase of 400 nm-turbidity and thioflavin-T fluorescence over time [5]. Only the measurements using the amyloid-specific dye are here used to quantitatively analyze mass-based $\alpha(t)$ progress curves (Fig. S1a). In previous analyses of TTR aggregation kinetics, the experimental data was well fitted by values of $k_\beta \gg 0.01$, which, nevertheless, changed with protein concentration in a manner not expected by the Classical Nucleation Theory (CNT) [1]. The applied concentrations of mTTR between 50 and 400 µg/mL are much higher than the solubility value of $c_\infty = 0.8$ µg/mL that was estimated from the monomer concentration in solution after long reaction times (1 week) independently of the departing mTTR concentration [5]. Although the terms *protein solubility* and *protein critical concentration* are often adopted interchangeably in the literature, the occurrence of STEs demands a clear distinction between c_∞ and c_c. The data measured by Hurshman et al. does not provide a direct estimate of c_c because the thioflavin-T signal becomes insensitive to protein aggregation for mTTR concentrations below 5 µg/mL [5]. Values of c_c between $c_\infty = 0.8$ µg/mL and 5 µg /mL are therefore acceptable. Model parameters k_α and k_β, and the additional unknown c_c can be numerically fitted without overparameterization problems from the analysis of the effect of protein concentration on phase separation kinetics (Fig. S1a). The obtained value of $c_c = 4.06$ µg/mL implies that the magnitude of STEs is low: $\alpha_c = 0.93$ from Eq. 4 using as c_0 the lowest concentration tested (50 µg/mL). Although only mild STEs are present, their contribution allows to adopt a single pair of concentration-independent kinetic constants to globally fit the measured effects of mTTR concentration on the $\alpha(t)$ time-course curves (Fig. S1a).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

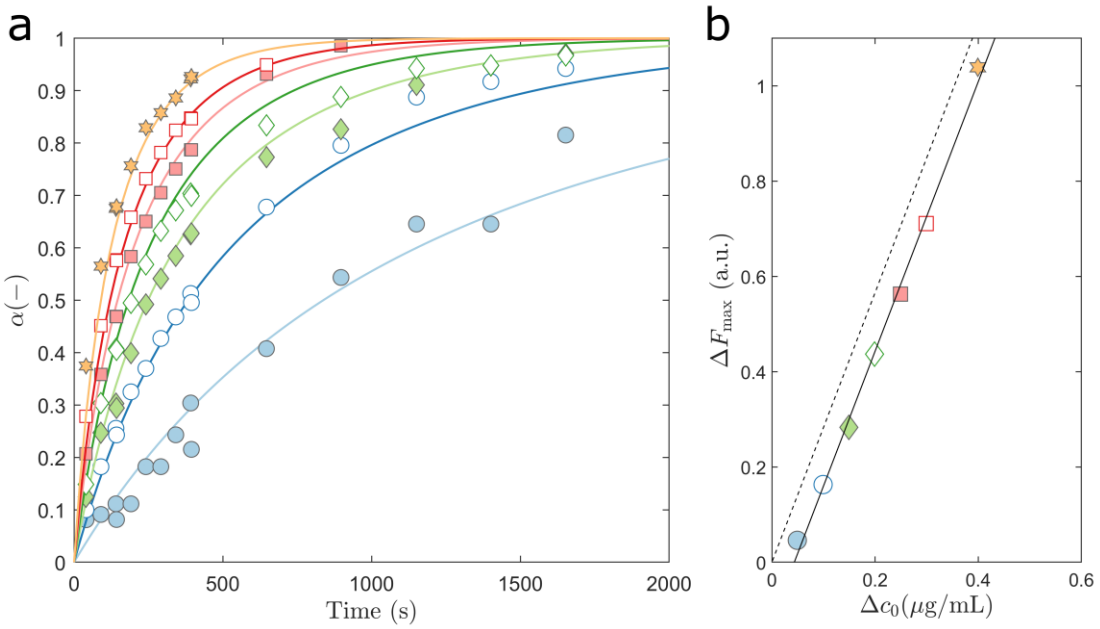

Figure S1. The general nucleation-and-growth model describes the (a) kinetics and (b) thermodynamics of mTTR aggregation measured by Hurshman et al. [5]. (a) Symbols: continuous fluorescence data digitized by us into periodic data and normalized by the end-point signal for mTTR concentrations of (from top to bottom) 400, 300, 250, 200, 150, 100 and 50 μg/mL. Solid lines: Progress curves predicted by Eq. S6 for the tested mTTR concentrations using the fitted parameters $k_\alpha = 1.05 \times 10^{-2}$ s^{-1}, $k_\beta = 1.71$ and $c_c = 4.06$ μg/mL. (b) Symbols: Values of maximum fluorescence measured as a function of $\Delta c_0 = (c_0 - c_\infty)$. Solid line: the experimental results follow a linear correlation. Dashed line: example of a straight line passing in the origin.

To cross-check the validity of protein-aggregation kinetic analyses, the maximum fluorescence signal obtained at the end of the reaction $\Delta F_{\max}$ should be evaluated against the concentration difference $\Delta c_0 = (c_0 - c_\infty)$. A direct proportion between these two quantities is expected in the absence of parallel aggregation pathways [1,6]. While a linear relationship between $\Delta F_{\max}$ and Δc_0 is confirmed, the positive X-intercept suggests the occurrence of TTR in other forms than monomers or amyloid fibrils (Fig. S1b). Confirming this indication, significant amounts of TTR dimers were identified by Hurshman et al. using analytical gel filtration and SDS-PAGE techniques [5]. Overall, the results in Fig. S1 confirm that the effect of protein concentration on mTTR aggregation is well described by standard nucleation theories modified to include STEs.

A distinguishing feature of the general nucleation-and-growth model is the predicted effect of protein concentration on the shape of reaction progress curves. Expectedly, in the case of fast-nucleating

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proteins such as mTTR, lowering the initial concentration not only slows down the rates of PS as it will also change the shape of the $\alpha(t)$ curves from hyperbolic to sigmoidal. This new possibility is admissible when the time interval during which primary nucleation occurs is significantly reduced by the occurrence of strong STEs. As we have seen for mTTR aggregation, only weak STEs are present in the range of protein concentrations studied by Hurshman et al. [5]. If the formation of amyloid fibrils of mTTR could be detected accurately for concentrations lower than 50 μg/mL, magnitudes of STEs higher than 0.07 ($\alpha_c < 0.93$) could be explored, and a lag period of apparently no aggregation would become visible as c_0 gradually approached the value of $c_c = 4.1$ μg/mL (Fig. S2, dashed lines). Based on this property, it is possible to gauge the intensity of STEs by graphically representing the reaction conversion as a function of Δc_0-normalized units of time: in principle, superimposing $\alpha(\Delta c_0 t)$ curves indicate that STEs are either weak or absent (Fig. S2).

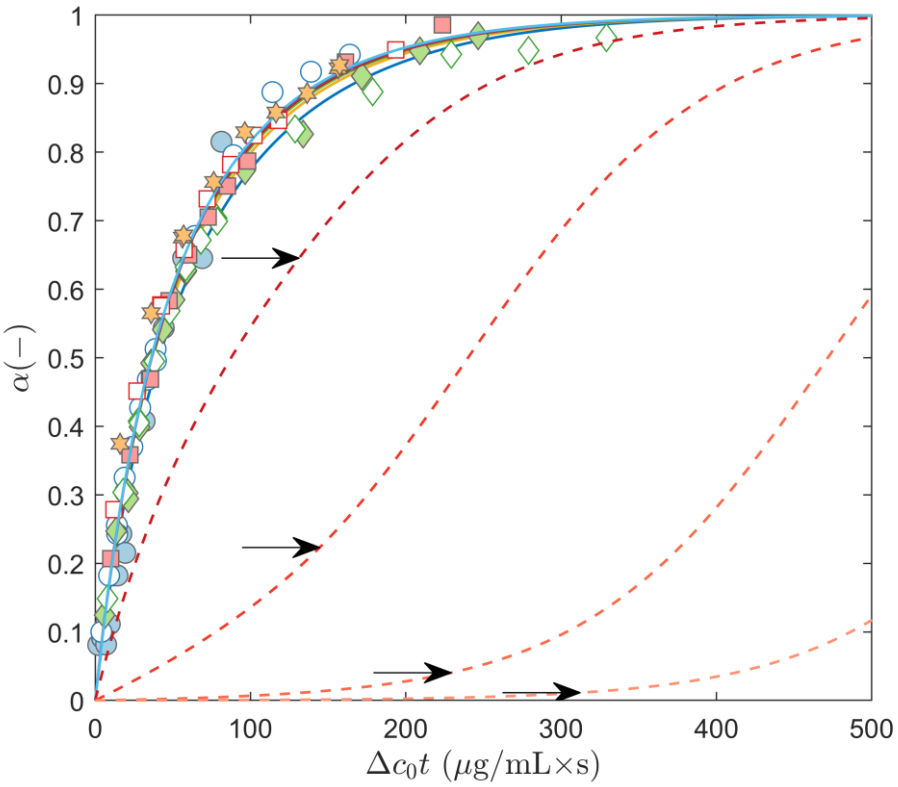

Figure S2. Identification of STEs from the representation of progress curves in Δc_0-normalized units of time. Symbols and solid lines: the results of Fig. S1a are now represented in normalized time units. Weak STEs influencing mTTR aggregation are confirmed by superimposing $\alpha(\Delta c_0 t)$ curves. Dashed lines: predicted $\alpha(\Delta c_0 t)$ curves for c_0 values of (from left to right) 10, 5, 4.2 and 4.1 μg/mL using the fitted parameters $k_\alpha = 1.05 \times 10^{-2}$ s^{-1}, $k_\beta = 1.71$ and $c_c = 4.06$ μg/mL. Arrows: critical reaction conversions (α_c).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

STEs on phase diagrams and particle size distributions: By including STEs in the theory of primary nucleation, a thermodynamic explanation emerges for the recurrent observation of a critical supersaturation limit below which phase separation does not occur at all unless seeding material is added. In terms of the CNT, this limit has been interpreted in relation to awaiting times arbitrarily chosen, below which nucleation is practically arrested [7]. In the presence of STEs, a thermodynamic (rather than kinetic) metastable zone exists where infinitely long times are required for primary nucleation to take place (in blue in Figs. S3a and S3b). In this sense, the width of the metastable zone is a parameter that can be objectively measured and is independent of the method applied to follow phase separation. Without the "margin of safety" created by STEs, protein self-assembly would be almost inevitable in the metastable environment of the cell [8].

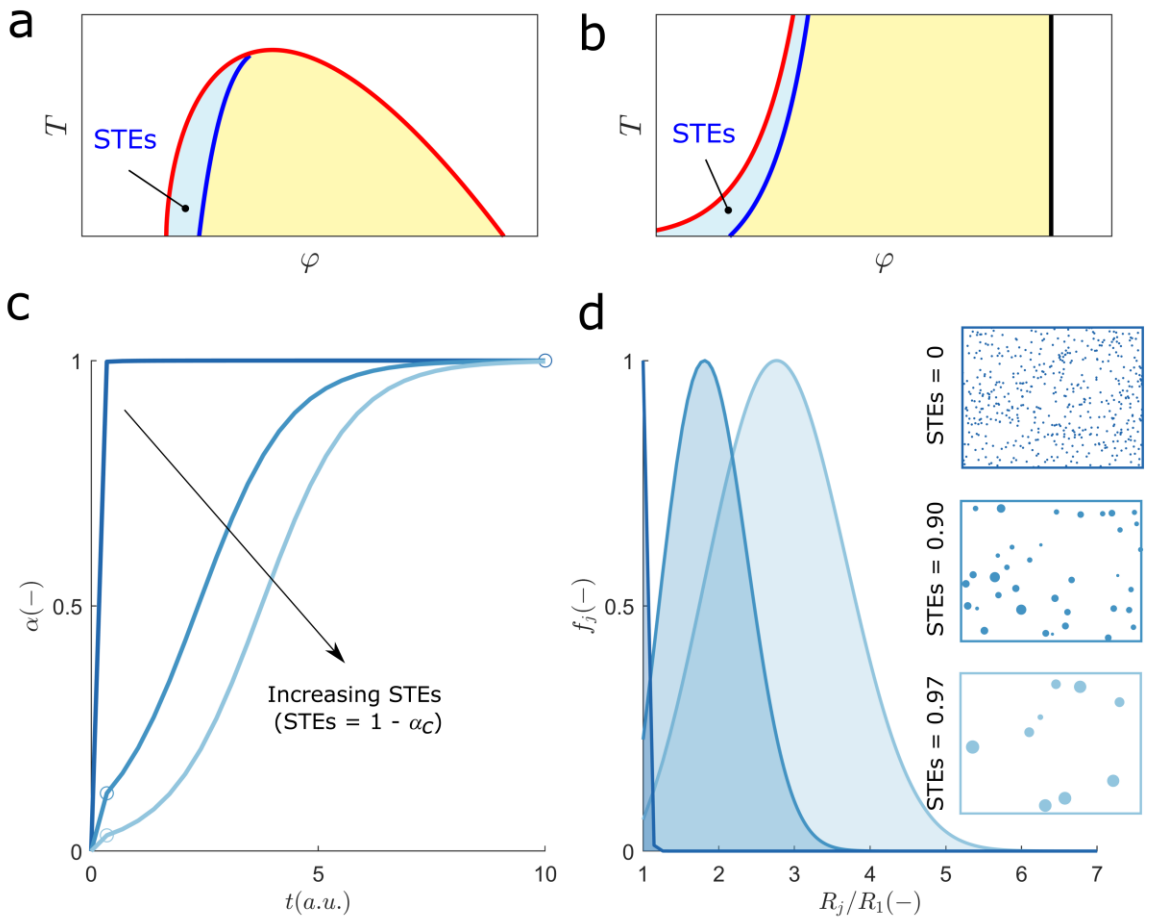

Figure S3. Thermodynamics and kinetics of nucleation-and-growth affected by STEs (refer to Fig. 1 of the main text for the case of no STEs present). (a and b) In the presence of STEs, primary nucleation becomes thermodynamically unfavored in the metastable region (blue area) of (a) liquid-liquid and (b) liquid-solid phase diagrams. (c) Theoretical progress curves calculated using $k_\beta = 10^3$ and (from left to right) STEs $= 0$, 0.90 and 0.97. Open circles: location of the critical coordinates t_c and α_c. (d) Final

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

f_j distribution calculated using the same model parameters and colour code as in (c). Large particle sizes uniformly distributed around the mean size can be produced in the presence of strong STEs (compare with Fig. 1d). Further numerical details in *SI Additional Methods*.

In the presence of strong STEs, a long growth-only period follows the short nucleation-and-growth period (Fig. S3c) giving rise to a small number of large particles (Fig. S3d). As we tend to the extreme of $\mathrm{STEs} = 1$, the initial formation of a single nucleus would be sufficient to lower the supersaturation level below the critical limit and phase transition would produce only one very large condensate. Movies S1 and S2 are time-lapse simulations showing how particles self-assemble until reaching the equilibrium size distributions in Fig. 1d and S3d, respectively. From comparing the two situations, we conclude that the combination of fast nucleation with strong STEs (movie S2, $\mathrm{STEs} = 0.97$) gives rise to a few large particles and, at the same time, leads to a sharp control of the moment new particles are formed. These features of spatiotemporal regulation of phase separation are not possible in the case of no STEs present: although large condensate bodies can still be produced under nucleation-limited conditions (Fig. 1d), new nuclei will keep arising erratically throughout the whole duration of phase separation (movie S1, $k_\beta = 5 \times 10^{-3}$).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

Additional Figures

a

$$\left(\frac{N_\infty}{N_1}\right)^{-1} = \frac{k_b}{1-k_b}\left[\frac{\ln(1/k_b)}{(1-k_b)}\left(1-\frac{k_2}{k_\alpha}\frac{N_1}{N_2}\right)-\left(1-\frac{k_2}{k_\alpha k_b}\frac{N_1}{N_2}\right)\right]$$

b

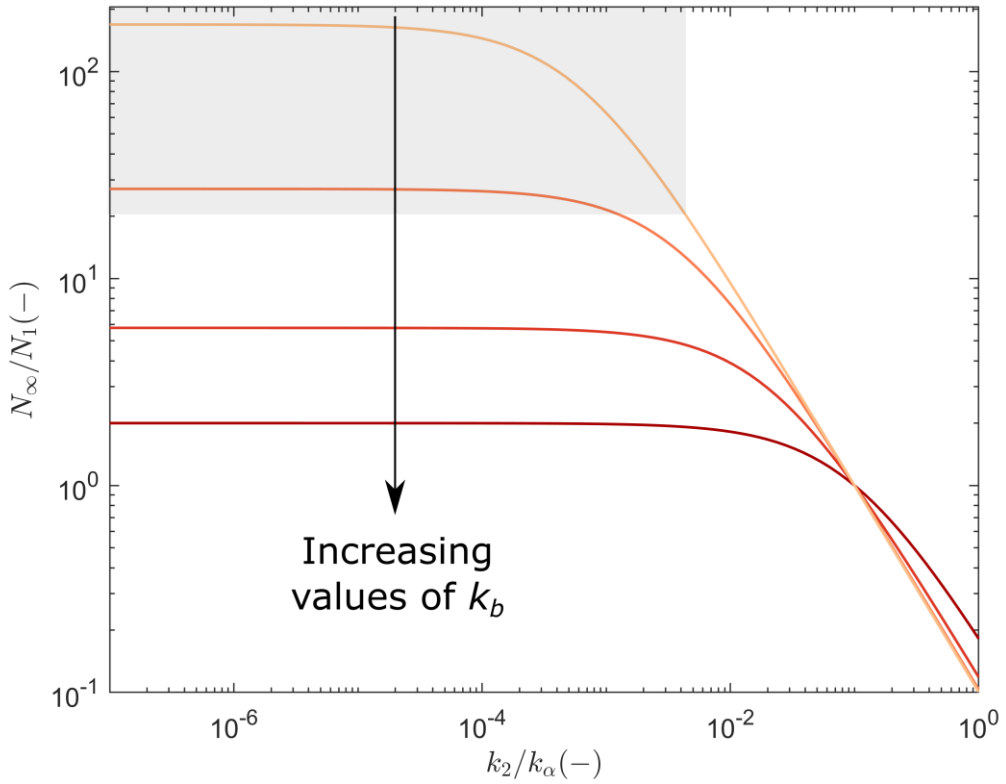

Figure S4. The impact of secondary nucleation on the final particle size in the absence of STEs. (a) The steady-state average particle size N_∞ can be calculated from Eqs. 5c and 5b extrapolated for long reaction times ($t \to \infty$). (b) The N_∞/N_1 ratio is calculated as a function of the k_α-normalized value of k_2 for values of k_b of (from top to bottom) 10^{-3}, 10^{-2}, 0.1, and 1 using $N_2/N_1 = 0.1$. Primary and secondary nucleation have to be much slower than growth for large particles to be formed (i.e., low values of both k_b and k_2/k_α are required to reach the shaded area).

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

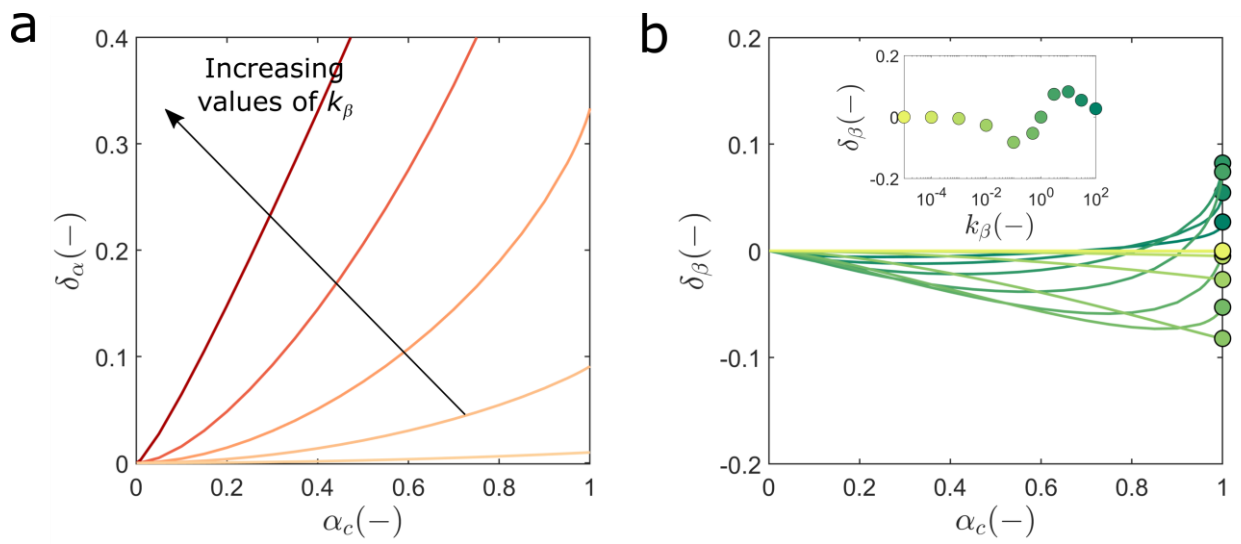

Figure S5. Maximum error of the $\tilde{\alpha}(t)$ and $\tilde{\beta}(t)$ approximations (Fig. 3a of the main text) relatively to the exact solution of Eq. 5. (a) Initial difference $\delta_\alpha = \tilde{\alpha}(0) - \alpha(0)$ for unseeded reactions ($\alpha(0) = 0$). The estimates of $\tilde{\alpha}(0) = \beta_\infty^{-1}$ are calculated from the definitions of β_∞ (Fig. 3b) and t_c (Fig. 2b) for values of k_β of (from lighter to darker shades of red) 10^{-2}, 10^{-1}, 0.5, 3 and 10^2. (b) Initial difference $\delta_\beta = \tilde{\beta}(0)/\beta_\infty - N_1/N(\infty)$, where $\tilde{\beta}(0) = 1$ (Fig. 3a) and $N_1/N(\infty)$ is numerically calculated for values of k_β from (lines from lighter to darker shades of green) 10^{-5} to 10^2. Symbols: higher values of δ_β are obtained for α_c values close to 1. Inset: the selected δ_β values are represented as a function of k_β.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

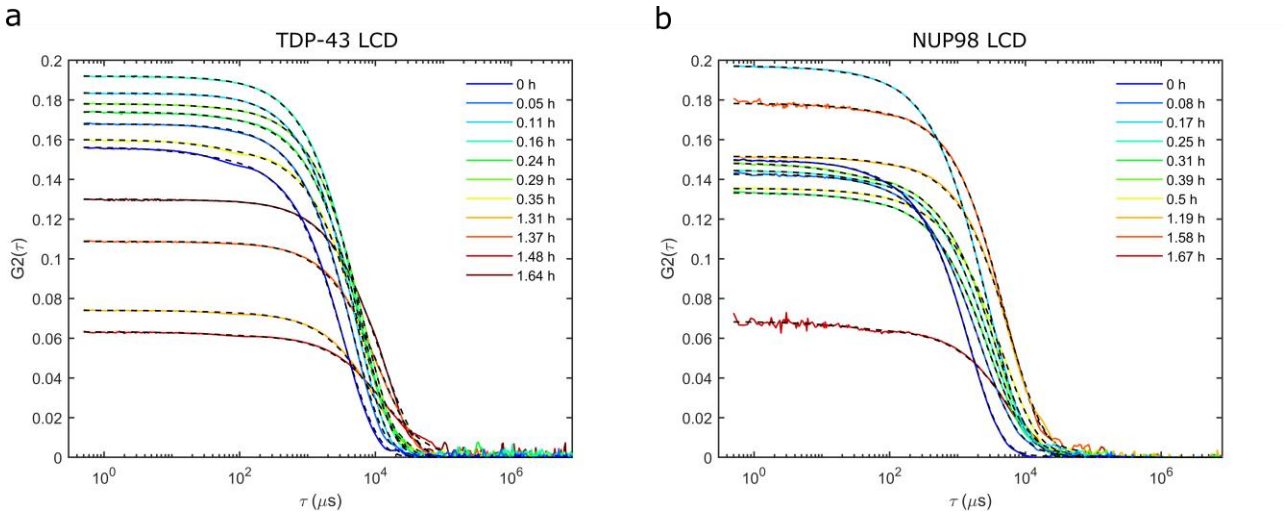

Figure S6. DLS autocorrelation functions during LLPS of (a) TDP-43 LCD and (b) NUP98 LCD. Solid lines: measured data from which the size distributions in Figs. 6a and 6b are inferred. Dashed lines: CONTIN analysis of the autocorrelation functions using the MATLAB code *rilt* [9,10].

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supporting Information (SI)

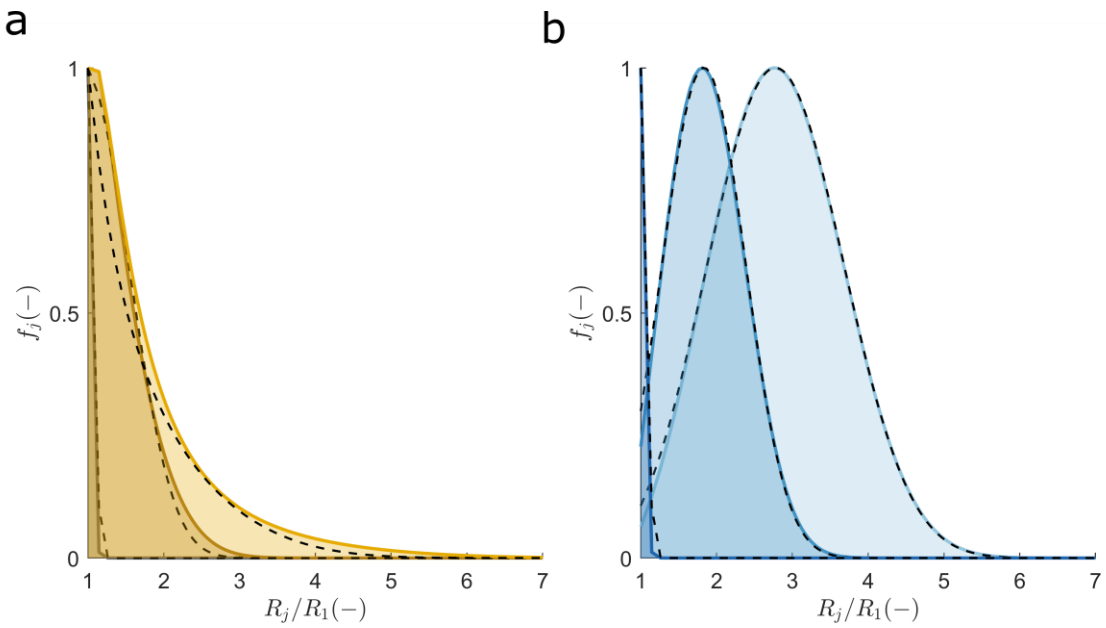

Figure S7. Using the gamma distribution to describe the exact solution of the master model equation (Eq. S2). The endpoint $C_j(t \to \infty)$ distributions obtained as in (a) Fig. 1d and (b) Fig. S3d are numerically fitted (dashed lines) by the two-parameter gamma distribution and represented in normalized frequency units.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.23.517626; this version posted November 24, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SI References

[1] R. Crespo, E. Villar-Alvarez, P. Taboada, F. A. Rocha, A. M. Damas, P. M. Martins, *J. Biol. Chem.* **2016**, *291*, 2018.

[2] A. Silva, B. Almeida, J. S. Fraga, P. Taboada, P. M. Martins, S. Macedo-Ribeiro, *Angew. Chem. Int. Ed.* **2017**, *56*, 14042.

[3] E. W. Weisstein, "Gamma Distribution." From MathWorld--A Wolfram Web Resource. https://mathworld.wolfram.com/GammaDistribution.html. **2022.**

[4] J. Schäfer, *MatScat* (https://www.mathworks.com/matlabcentral/fileexchange/36831-matscat), MATLAB Central File Exchange. **2022**.

[5] A. R. Hurshman, J. T. White, E. T. Powers, J. W. Kelly, *Biochemistry* **2004**, *43*, 7365.

[6] A. Silva, Z. Sárkány, J. Fraga, P. Taboada, S. Macedo-Ribeiro, P. Martins, *Biomolecules* **2018**, *8*, 108.

[7] D. Kashchiev, G. M. van Rosmalen, *Cryst. Res. Technol.* **2003**, *38*, 555.

[8] G. G. Tartaglia, S. Pechmann, C. M. Dobson, M. Vendruscolo, *Trends Biochem. Sci.* **2007**, *32*, 204.

[9] I.-G. Marino, *rilt* (https://www.mathworks.com/matlabcentral/fileexchange/6523-rilt), MATLAB Central File Exchange. **2022**.

[10] S. W. Provencher, *Comput. Phys. Commun.* **1982**, *27*, 229.