eLife

Biochemistry and Chemical Biology
Computational and Systems Biology

# The interplay between biomolecular assembly and phase separation

**Reviewed Preprint**
**v2** • October 30, 2024
Revised by authors

**Reviewed Preprint**
**v1** • May 2, 2024

**Giacomo Bartolucci**, **Ivar S Haugerud**, **Thomas CT Michaels**, **Christoph A Weber**

Max Planck Institute for the Physics of Complex Systems, Dresden, Germany • Faculty of Mathematics, Natural Sciences, and Materials Engineering: Institute of Physics, University of Augsburg, Universitätsstraße 1, Augsburg, Germany • ETH Zurich, Institute of Biochemistry, Zurich, Switzerland



**eLife Assessment**

The authors present an **important** theoretical framework that describes the interplay between liquid-liquid phase separation and protein aggregation within a mean-field model. This work will be of high interest to the biophysics and molecular biology communities, as it will help understand and analyse assembly within biomolecular condensates in cells or in-vitro. Major strengths of this **convincing** work are the consideration of aggregates with various dimensionality and the possibility for protein gelation.

https://doi.org/10.7554/eLife.93003.2.sa2

## Abstract

Many biological functions and dysfunctions rely on two fundamental processes, molecular assembly and the formation of condensed phases such as biomolecular condensates. Condensed phases generally form via phase separation, while molecular assemblies are clusters of molecules of various sizes, shapes, and functionality. We developed a theory that relies on thermodynamic principles to understand the interplay between molecular assembly and phase separation. We propose two prototypical classes of protein interactions and characterize their different equilibrium states and relaxation dynamics. We obtain results consistent with recent in vitro experimental observations of reconstituted proteins, including anomalous size distribution of assemblies, the gelation of condensed phases, and the change in condensate volume during ageing. Our theory provides the framework to unravel the mechanisms underlying physiological assemblies essential for cellular function, and aberrant assemblies which are associated with several neurodegenerative disorders.

## I. Introduction

Due to their structural complexity, proteins can interact in different ways, leading to coexisting phases or assemblies such as fibers and aggregates. Long-lived assemblies are often kept together by strong adhesive forces, with binding free energies ranging from 9 $k_{\mathrm{B}}T$ in the case of insulin

dimers [1], over 2.5 $k_BT$ per beta-sheet in amyloid fibers, to the 0.9 $k_BT$ per beta-sheet in the formation of assemblies of specific FUS segments called low-complexity aromatic-rich kinked segments [2]. Weak interactions are often responsible for the separation into liquid phases, each of distinct molecular compositions. The interaction free energies associated with the formation of P granules via phase separation in living cells are about 0.5 $k_BT$ per molecule [3]. The biological function of both assemblies and phase-separated compartments relies on the recruitment of specific biomolecules such as proteins, RNA or DNA [4–7]. Since assemblies and condensed phases can adhere to membrane surfaces, both not only mediate mechanisms for sorting and transport of molecules [8] but also affect the composition, shape and properties of intra-cellular surfaces [9–12].

Despite these similarities, molecular assemblies and coexisting phases also exhibit crucial differences. While the size of a condensed phase at equilibrium increases with the size of the system [13], this is not necessarily the case for molecular assemblies [14–17]. Moreover, the assembly kinetics tends to an equilibrium between assemblies of different sizes [14–16, 18], while condensed phases equilibrate the physico-chemical properties such as temperature, pressure and chemical potential between the spatially separated phases [13]. These differences suggest a rich interplay in a system where the molecular constituents can both oligomerise forming assemblies and give rise to coexisting phases [19–24].

In the last years, the interplay between phase separation and assembly formation has been the focus of many experimental efforts. Different proteins capable of forming condensed phases were shown to form oligomers below the saturation concentration [25, 26]. The authors proposed that such oligomers affect the phase separation propensity, however, the detailed mechanism remains elusive. Moreover, several experimental studies indicate that proteins in the protein-rich phase are linked, reminiscent of a physical gel [27–29]. Molecular simulations were performed that aimed at the sequence-specific origin of such phenomena [30–33]. However, even in elegantly coarse-grained simulation approaches, the large number of parameters makes it difficult to extract general mechanisms across different proteins. To develop an understanding of such general mechanisms that underlie the interplay between phase separation and molecular assembly, a theoretical framework that relies on thermo-dynamic principles is lacking.

While the theory of phase separation of a low number of different components [13, 34], as well as the formation of molecular assemblies in dilute environments [14, 35, 36], are well developed, only a few works addressed assembly formation beyond the dilute limit, where assemblies can form and also phase separate. For example, it has been shown that, in the presence of co-existing phases, the assembly size distributions at equilibrium can vary in the two phases [37, 38] and that the protein-rich phase can gelate [39–42]. These studies account for the scaling of the internal free energies of assemblies with their size but neglect the size dependence of the interaction propensities. Moreover, a discussion of the coupled phase separation and assembly kinetics is lacking. Other authors focused on systems composed of a scaffold component, that drives phase separation, and studied the dilute assembly kinetics of a second component that can interact with the scaffold [43–46]. In these works, the assemblies are considered to be dilute and the feedback of the assembly kinetics on the phase-separated compartment is neglected.

In this work, we introduce a framework that unifies the thermodynamic theories for phase separation with the theories developed for the formation of micelles and molecular assemblies at dilute conditions. This multiscale framework bridges assembly, a phenomenon occurring at the molecular scale, with phase separation occurring in the macroscopic realm. We present two classes of size-dependent interactions that are inspired by biologically relevant proteins. Our theory is able to reproduce results observed in recent experimental studies, such as the emergence of anomalous size distribution below saturation and the gelation of condensed phases above saturation, and characterise for which class and parameter values these phenomena manifest. Furthermore, we propose a non-equilibrium thermodynamic theory for the kinetics of molecular

assembly at non-dilute conditions which can lead to macroscopic, condensed phases above the saturation concentration. The complexity of our theory is reflected in a high dimensional phase space that is set by the number of differently sized assemblies. We developed efficient numerical schemes to investigate the kinetics of such systems for the case where diffusion is fast compared to assembly kinetics. In particular, we study how condensates, initially formed via the phase separation of monomers from the solvent change in response to the formation of assemblies. Our unified theory provides the answer to crucial biological questions, such as under which conditions the presence of coexisting phases affects the formation of assemblies, and could be key to interpreting and understanding recent observations of protein condensation in vitro [47], and in the cell cytoplasm [26, 29, 48, 49].

## II. Assembly and phase equilibria

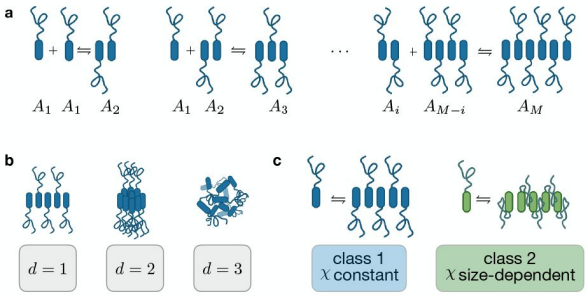
We begin by reviewing the equilibrium theory of multi-component mixtures composed of solvent (s) and monomers ($i = 1$) that can form assemblies composed of $i$ monomers, with $i < M$ see **Fig. 1a**. We consider a maximum assembly size $M$, but, as we will see, this assumption must be relaxed when monomers tend to form an assembly of infinite size. In the case when monomers and assemblies are dissolved in the solvent, the free energy density of the solution can be written as [37, 41, 50, 51]:

$$f_{\rm sol} = \frac{k_{\rm B}T}{\nu_{\rm s}}\left(\sum_{i=1}^{M}\frac{\phi_i}{\rho_i}\ln\left(\frac{\phi_i}{\rho_i}\right) + \frac{\omega_i}{k_{\rm B}T}\phi_i + \sum_{i\neq j=1}^{M}\frac{\chi_{ij}}{2k_{\rm B}T}\phi_i\phi_j + \phi_{\rm s}\ln\phi_{\rm s} + \frac{\omega_{\rm s}}{k_{\rm B}T}\phi_{\rm s} + \sum_{i=1}^{M}\frac{\chi_{is}}{k_{\rm B}T}\phi_i\phi_{\rm s}\right), \qquad (1)$$

where $\rho_i = \nu_i/\nu_{\rm s}$ denotes the relative molecular volume with $\nu_i$ as the molecular volume of assembly of size $i$, and $\nu_{\rm s}$ denotes the solvent molecular volume. The solvent volume fraction can be expressed as a function of the assembly volume fractions via $\phi_{\rm s} = 1 - \sum_{i=1}^{M}\phi_i$. The first and fourth terms in Eq. (1) are the mixing entropies.

The second and fifth contributions of $f_{\rm sol}$ account for the internal free energies $\omega$. Here, $\omega_{\rm s}$ denotes the internal free energy of the solvent, and $\omega_i$ are the internal free energies per monomer of an assembly of size $i$, stemming from the free energy of internal bonds that lead to assembly formation. Note that we chose to keep $\Phi_i/\rho_i$ in the logarithm argument instead of reabsorbing the linear term $-\Phi_i \ln(\rho_i)/\rho_i$ in the internal free energies $\omega_i$. With this choice, $\omega_i$, depends only on the free energies of the bonds, see Appendix B, Ref. [50], and the recent overview in the SI of Ref. [45]. The third and last terms in Eq. (1) capture the interactions of monomers belonging to different assemblies and with the solvent, where $\chi_{ij}$ is the corresponding interaction parameter. We note that varying the temperature $T$ affects the contributions of both the interaction and the internal free energy terms. The exchange chemical potentials of monomers belonging to an assembly of size $i$ reads

$$\mu_i = \nu_1\frac{\partial f_{\rm sol}}{\partial \phi_i}. \qquad (2)$$




## Figure 1

### Illustration of assembly reaction scheme and classification.

**a** Illustration of the chemical reaction network associated with the formation of assemblies $A_i$ with size $i$. **b** Identification of three classes based on assembly dimension: d=1,2,3. **c** Classification of assemblies based on the scaling of their Flory-Huggins interaction propensity.

## Assembly equilibrium

Assemblies can grow and shrink via association and dissociation. Such transitions among assemblies of different sizes are reminiscent of chemical transitions, see **Fig. 1a**. The condition of chemical equilibrium reads [16]:

$$\mu_i = \mu_1 = \text{const.} \quad \forall\, i = 2 \dots M\,, \tag{3}$$

where $\mu_i$ is the exchange chemical potential of monomers belonging to an assembly of size $i$; see Eq. (2). Using the free energy Eq. (1) and the equilibrium conditions Eq. (3), we can express the volume fraction of the assembly of size $i$ as a function of the monomer volume fraction $\Phi_1$ in the form:

$$\phi_i = \rho_i \left(\frac{\phi_1}{\rho_1}\right)^{\rho_i/\rho_1} \exp\left[\rho_i\left(1 - \frac{\omega_i - \omega_1}{k_\mathrm{B}T} - \frac{\chi_{is} - \chi_{1s}}{k_\mathrm{B}T}\phi_\mathrm{s} - \sum_j \frac{\chi_{ij} - \chi_{1j}}{k_\mathrm{B}T}\phi_j\right) - 1\right]. \tag{4}$$

The equation above together with the conservation of monomers

$$\phi_\mathrm{tot} = \sum_{i=1}^{M} \phi_i\,, \tag{5}$$

allows us to rewrite the volume fraction $\Phi_i$ of each assembly of size $i$, as a function of the conserved quantity $\Phi_\mathrm{tot}$. This relation $\Phi_i = \Phi_i(\Phi_\mathrm{tot})$ has an analytical expression in the case $d = 1$, see Eq. (C1) and Eq. (C3) in Appendix C.

## Phase equilibrium

Two phases in an incompressible, multi-component system are at phase equilibrium when the chemical potentials $\mu_i$ and the osmotic pressure $\Pi = -f_\mathrm{sol} + \sum_{i=1}^{M} \phi_i \partial f_\mathrm{sol}/\partial\phi_i$ balance in each phase [13, 17]:

$$\mu_i^\mathrm{I} = \mu_i^\mathrm{II}\,, \tag{6a}$$

$$\Pi^\mathrm{I} = \Pi^\mathrm{II}\,, \tag{6b}$$

where the superscripts I and II indicate the $\Phi_\mathrm{tot}$-rich and II the $\Phi_\mathrm{tot}$-poor phase, respectively.

## Thermodynamic equilibrium

Our system is at thermodynamic equilibrium when assembly and phase equilibrium hold simultaneously. The conditions above for phase equilibrium can thus be rewritten using $\Phi_i$ ($\Phi_\mathrm{tot}$) (Eq. (4)). In particular, the free energy density Eq. (1) can be recast in terms of the conserved variable, $\Phi_\mathrm{tot}$ [52, 53]. The phase diagram of the system can be then obtained via the common

tangent construction (i.e., Maxwell construction). This construction corresponds to the balance between the exchange chemical potentials and the osmotic pressure in both phases, see Chapter 2 in Ref. [52, 53]:

$$\mu(\phi_{\rm tot}^{\rm I}) = \mu(\phi_{\rm tot}^{\rm II})\,, \tag{7a}$$

$$\mu(\phi_{\rm tot}^{\rm I}) = \frac{f_{\rm sol}(\phi_{\rm tot}^{\rm II}) - f_{\rm sol}(\phi_{\rm tot}^{\rm I})}{\phi_{\rm tot}^{\rm II} - \phi_{\rm tot}^{\rm I}}\,. \tag{7b}$$

Eq. (3) and Eq. (7) establish how the behaviour of the mixture at equilibrium is affected by the parameters of the free energy in Eq. (1) such as internal free energies $\omega_i$ or interaction parameters $\chi_{ij}$. In the next section, we introduce classes based on the scaling of such parameters with assembly size $i$.

# III. Scaling of molecular volumes, internal free energies and interaction energies with assembly size

The composition of the phase-separated compartments and the size distributions of the assemblies in each phase will depend on the scaling form of the key parameters of the model with the assembly size $i$: the relative molecular volumes ($\rho_i$), the internal free energy of assemblies ($\omega_i$), and the interaction energies of assemblies among themselves ($\chi_{ij}$), and with the solvent ($\chi_{is}$).

In this work, we choose $\rho_i = i$. This choice reflects the fact that no solvent is present in assemblies and that the chemical reaction network in **Fig 1a** conserves the sum of molecular volumes. The assumption $\rho_i = 1$, leads to a phase diagram that is symmetric about $\Phi_1 = 1/2$, if only monomers and solvent are present. This might seem an oversimplification but, once assemblies of different sizes form, the phase diagram becomes asymmetric, as expected in most biological applications. Thus, this assumption simplifies the framework while the equilibrium states retain the essential qualitative features of realistic systems. In our model, assemblies form as a consequence of internal bonds among monomers. Each bond is associated with a free energy $e_{\rm int}$–$s_{\rm int}T$, with $e_{\rm int}$ and $s_{\rm int}$ the enthalpic and an entropic contribution, respectively. In Appendix B, we derive the scaling relationships for the internal free energies of linear ($d = 1$), planar ($d = 2$) and three-dimensional ($d = 3$) assemblies:

$$\omega_i \simeq \omega_\infty - \frac{e_{\rm int} - s_{\rm int}T}{i^{1/d}}\,. \tag{8}$$

The physical origin of the dependency $i^{1/d}$ is the scaling of the number of internal bonds in an assembly of dimension $d$. In **Eq. (8)**, $\omega_\infty = \lim_{i\to\infty}\omega_i$ is a constant that does not affect chemical nor phase equilibrium, except in the limit $M \to \infty$, which will be discussed later. In Appendix E, we discuss how variations of bond energy affect phase separation.

For the scaling of interaction energies $\chi_{ij}$ and $\chi_{is}$, we introduce two classes inspired by biologically relevant classes of proteins that can form assemblies and phase separate:

## 1. Class 1: Constant assembly-solvent interactions

This class corresponds to the case where each monomer, independently of the assembly it is part of, interacts equally with the solvent $\chi_{is} = \chi$.

Moreover, monomers in assemblies of different sizes interact equally with each other, implying that the corresponding Flory-Huggins parameter $\chi_{ij}$ vanishes:

$$\chi_{is} = \chi\,, \qquad \chi_{ij} = 0\,, \tag{9}$$

for a derivation of this relation starting from a lattice model see Appendix B. This class is inspired by biologically relevant proteins for which the oligomerization domains are well separated along the protein from hydrophobic phase separation domains. In this case, when monomers form an assembly, their phase separation domains remain exposed, leading to a monomer-solvent interaction that does not depend on assembly size. Examples belonging to this class include synthetic constructs like the so-called 'Corelets' [54], realised tethering intrinsically disordered protein fragments to oligomerizing domains [54], and proteins like NPM1, whose N-terminal oligomerization domain (that allows for the formation of pentamers) is considered to be separated from the disordered region (responsible for phase separation) and the RNA binding domain [55, 56].

### 2. Class 2: Size-dependent assembly-solvent interactions

This class describes the case where monomers in the assembly bulk and monomers at the assembly boundary have different interaction propensities with the solvent ($\chi'$ and $\chi$ respectively, see Appendix B for details). Similar to class 1, monomers in assemblies of different sizes interact equally with each other, leading to

$$\chi_{is} = \chi' + \frac{\chi - \chi'}{i^{1/d}}\,, \quad \chi_{ij} = 0\,. \tag{10}$$

The dependency $i^{1/d}$ originates from the scaling of the number of monomers in the bulk and in the boundary of assemblies, in different spatial dimensions $d$. This class corresponds to the general case in which the oligomerization domains of protein overlap with the phase separation domains. This case applies to segments of the intrinsically disordered region of the protein FUS, for example. In fact, recent experiments have shown the formation of assemblies in solutions containing specific FUS domains, called low-complexity aromatic-rich kinked segments (LARKS) [2, 57]. Strikingly, it was shown that hydrophobic domains along LARKS were buried in the formation of these assemblies and the author could quantify the hydrophobic area buried upon assembly formation. Another example could be Whi3, since it has been recently found that mutation that enhances oligomerization strength, lowers the density of Whi3 in the RNP condensates [49], suggesting that the formation of assemblies could screen Whi3 phase separation propensity. Finally, the formation of DNA nanostars has been recently shown to inhibit phase separation in DNA liquids, [58]

We consider the relevant interaction parameters, like internal free energies $\omega_i$ and interaction propensities $\chi$ and $\chi'$ as control parameters and vary them independently. This is a simplification since, in biology, they might be coupled, e.g. the swelling of the intrinsically disordered regions could lead to variations in monomer binding strength [59]. In the next sections, we characterize the equilibrium behaviour of systems belonging to these classes.

## IV. Assembly size distributions below and above saturation

Here, we discuss the differences between assembly equilibrium in homogeneous and phase-separating systems and outline the implications for biological mixtures. We first consider systems that are spatially homogeneous and composed of linear assemblies ($d$ = 1). Homogeneity can be

realized in dilute solutions if the total protein volume fraction $\Phi_{tot}$ is below the saturation volume fraction of phase separation $\phi_{tot}^{II}(T)$ (for a definition see Sec. II). Homogeneous systems governed by Eq. (4) at equilibrium, obeying the conservation Eq. (5), exhibit two limiting behaviours depending on the value of the conserved variable $\Phi_{tot}$. We define the *assembly threshold* $\Phi^*(T)$, that separates these two behaviours, as the value of $\Phi_{tot}$ for which there is a maximum of $\Phi_i$ for monomers ($i = 1$) with zero slope:

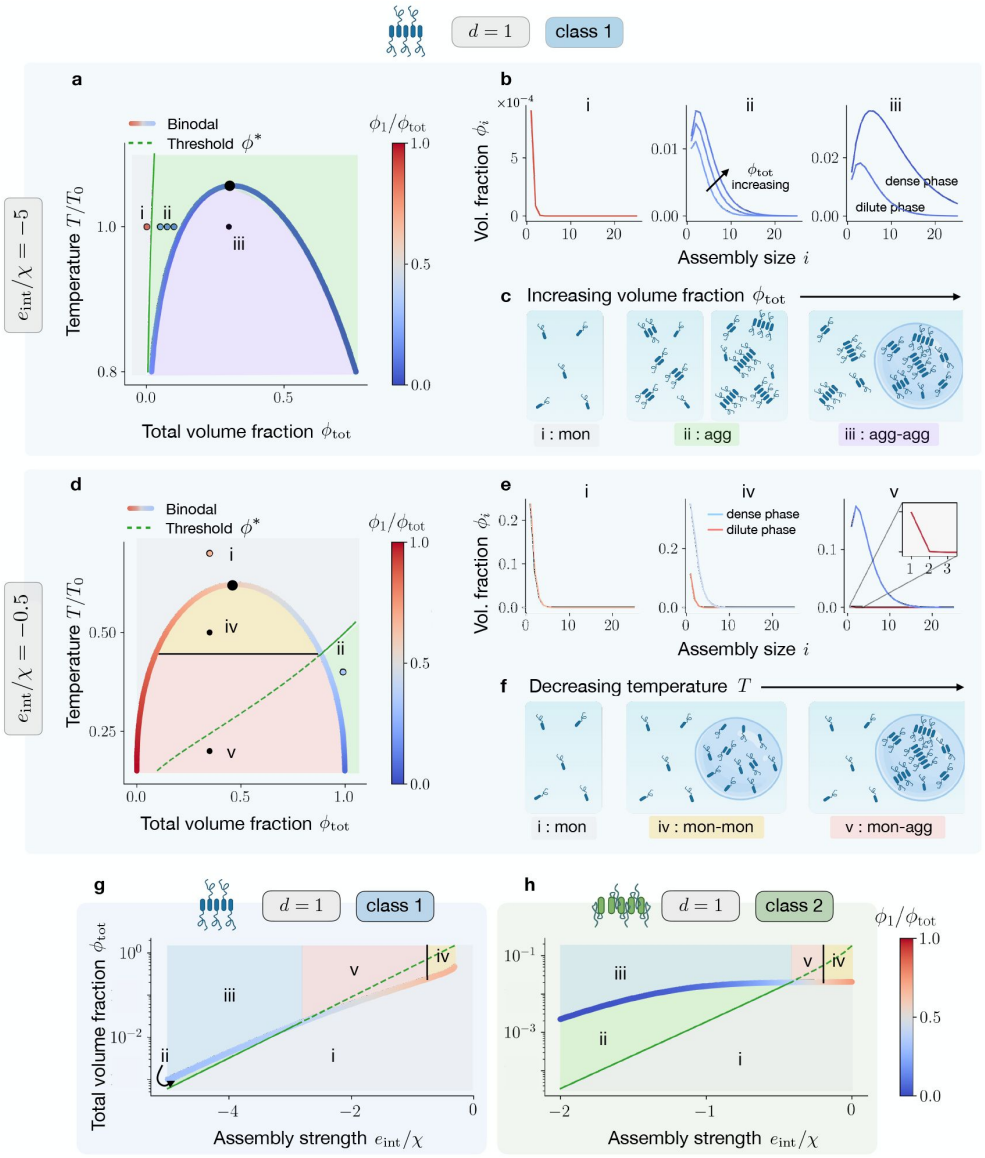
$$\left.\frac{\partial \phi_i(\phi_{tot})}{\partial i}\right|_{i=1,\phi^*} = 0. \qquad (11)$$

Indeed, for $\Phi_{tot} \ll \Phi^*$ the size distribution of linear assemblies ($d = 1$) is dominated by monomers ($\Phi_1 \simeq \Phi_{tot}$) while larger assemblies have vanishing volume fraction. For higher total volume fractions ($\Phi_{tot} \gtrsim \Phi^*$), the monomer volume fraction saturates at $\Phi_1 \lesssim \Phi^*$ and bigger assemblies start to populate the mixture. Above $\Phi^*$, the distribution becomes peaked at a value $i_{max} > 1$ and then exponentially decays for larger $i$; see **Fig. 6** in Appendix B. Both the maximum and the average of the distribution $\Phi_i$ scale with $\sqrt{\phi_{tot}}$ indicating that as $\Phi_{tot}$ is increased, larger assembly populate the system; see Appendix C for a detailed discussion for Class 1.

Now we consider systems that can phase separate. As outlined in Sec. II, at assembly equilibrium, we can recast the free energy as a function of the conserved variable $\Phi_{tot}$ by using **Eq. (4)**. For sufficiently large assembly-solvent interaction parameters $\chi$ and $\chi'$, the system can demix into two phases with different total volume fractions $\phi_{tot}^{I}$ and $\phi_{tot}^{II}$, which are the solutions of **Eq. (7)**. By means of $\phi_{tot}^{I/II}$, we can calculate the whole assembly size distribution in the two phases, i.e., $\phi_i^{I/II}$, via **Eq. (4)** and **Eq. (5)**.

We first discuss linear assemblies belonging to class 1, in the regime of high assembly strength $-e_{int}/\chi \gg 1$; see **Fig. 2a-c**. In **Fig. 2a**, we show the corresponding phase diagram as a function of $\Phi_{tot}$ and the rescaled temperature $T/T_0$ with $T_0 = \chi/k_B$. The domain enclosed by the binodal corresponds to phase separation. As indicated by the colour code (depicting the monomer fraction $\Phi_1/\Phi_{tot}$) each point in the diagram can have different assembly composition. In green we plot the assembly threshold $\Phi^*(T)$, at which intermediate-sized assemblies start to appear. Note that, with this choice of parameters, the assembly threshold precedes in $\Phi_{tot}$ the dilute branch of the binodal. We stress that, for $d = 1$, crossing the assembly threshold does not lead to a phase transition since, in contrast to crossing the binodal, it is not accompanied by a jump in the free energy or its derivatives. We can now define regions corresponding to qualitatively different phase and assembly behaviour. In particular, starting from a homogeneous system composed of monomers only (region "i"), increasing $\Phi_{tot}$ leads to the emergence of intermediate-sized assemblies (region "ii"). Increasing $\Phi_{tot}$ further, the system demixes into two phases both of which are rich in intermediate assemblies (region "iii"). Representative size distributions and illustrations of the state of the systems in the different regions are shown in **Fig. 2b** and **Fig. 2c**, respectively. For parameter values see **Table I** in Appendix A. This analysis showcases the potential of this framework to describe the appearance of mesoscopic clusters below the saturation, as recently observed experimentally in Ref. [26].

Remaining within class 1, we now discuss the case of low assembly strength $-e_{int}/\chi \sim 1$; see **Fig. 2d-f**. The interception between the binodal and the assembly threshold $\Phi^*$ defines two new regions, "iv" and "v", see **Fig 2 d**. In particular, in region "iv", both binodal branches lie below the assembly threshold, resulting in monomers dominating both coexisting phases, see **Fig 2e**, centre. On the other hand, in region "v" the $\Phi_{tot}$-rich phase exceeds the assembly threshold, resulting in phases with dramatically different compositions: the $\Phi_{tot}$-poor pase is populated only by monomers while intermediate-sized assemblies develop in the $\Phi_{tot}$-rich phase, see **Fig 2e** right. The spatial separation of assemblies into the $\Phi_{tot}$-rich phase has likely far-reaching biological implications. For example, it may reduce the toxic effects of aggregates in




## Figure 2

### Phase diagram and assembly size distributions for different classes and assembly strengths.

**a** Phase diagram as a function of $\Phi_{tot}$ and rescaled temperature $T/T_0$ (with $T_0 = \chi/k_B$) in the regime of high assembly strength, i.e. $-e_{int}/\chi \gg 1$. The green line is the volume fraction threshold $\Phi*(T)$ at which intermediate-sized assemblies start to appear, which in this regime precedes the binodal (coloured curve). As indicated by the colour code, the monomer fraction $\Phi/\Phi_{tot}$ mildly varies in the two phases. **b** Size distributions and **c** pictorial representations corresponding to different regions of the phase diagram, defined by the relative position of the binodal and the assembly threshold. In region "i", the system is homogeneous and composed of monomers only. Increasing the total volume fraction of assemblies $\Phi_{tot}$ beyond the assembly threshold $\Phi^*$, the system enters region "ii" where intermediate assemblies appear. Here, the sizes corresponding to the maximum and the average of the distribution $\Phi_i$ scale with $\sqrt{\phi_{tot}}$, see Appendix C. Finally, once $\Phi_{tot}$ exceeds the binodal, the system enters region "v" and demixes in two phases, both rich in intermediate assemblies. In **d-f** we focus on the low assembly strength regime, i.e. $-e_{int}/\chi \sim 1$. In phase diagram **d**, the binodal now precedes in $\Phi_{tot}$ the assembly threshold. **e** In region "iv", the system phase separates but in both phases monomers dominate the size distribution, while in region "v" the $\Phi_{tot}$-rich phase becomes populated by intermediate-sized assemblies. Progressively lowering the temperature allows switching between these regions, as depicted in **f. g**,**h** Behaviour of dilute mixtures as a function of assembly strength, for the two different classes. Notably, assembly below saturation becomes much more accessible for class 2, as can be seen by comparing the green regions "ii" in **g** and **h**.

neurodegenerative diseases [6]. In **Fig 2f**, we illustrate states corresponding to fixed $\Phi_{tot}$ and decreasing temperature $T$. Starting from a homogeneous monomeric state, region "i", the system transitions into a demixed state with monomers dominating both phases, region "iv", and finally to a demixed state with larger assemblies abundant in the $\Phi_{tot}$-rich phase, region "v".

We now highlight the differences between the two classes defined in Sec. III. In particular, we characterise how mixtures of monomers prone to assembly and phase separation behave with increasing $\Phi_{tot}$, varying the assembly strength $e_{int}/\chi$ but keeping the temperature $T$ fixed. In particular, for class 1, the emergence of assemblies before saturation typically occurs for a very narrow interval of volume fractions, see the green region labelled with "ii" in **Fig 2g**. Strikingly, for class 2, assembly below saturation are more favoured; see again region "ii" in **Fig 2h**. This difference arises because, within class 2, monomers in the bulk of an assembly have reduced interaction propensity with respect to the boundary ones. As a consequence, the formation of large clusters shifts the onset of phase separation to higher $\Phi_{tot}$ values. Summing up, phase separation controls the onset (see also Appendix E) and localization of assemblies. For the considered parameters, the $\Phi_{tot}$-rich phase contains larger assemblies compared to the $\Phi_{tot}$-poor phase. In the next section, we will see that, for planar and spherical assemblies ($d > 1$), this difference can become even more extreme with the protein-rich phase becoming one giant assembly, also referred to as the gel phase [15].

## V. Gelation of the protein-rich phase

In this section, we discuss the case of planar ($d = 2$) and three-dimensional assemblies, ($d = 3$), referring for simplicity to systems belonging to Class 1. In this case, as shown in Appendix D, even when neglecting protein solvent interactions ($\chi = 0$), the system can undergo a transition where the protein-rich phase becomes a large (macroscopic) assembly. In the thermodynamic limit $M \to \infty$, this transition corresponds to a phase transition, i.e., allowing for the emergence of an infinitely large assembly. In fact, above the threshold volume fraction $\Phi^{sg}$, we observe the emergence of such a macroscopic assembly occupying a finite fraction of the system volume that contains a macroscopic fraction of all monomers in the system – a behaviour reminiscent of Bose-Einstein condensation; see for example Chapter 7.3 of Ref. [41] for an interesting discussion on this analogy. The threshold volume fraction $\Phi^{sg}$ is affected by temperature and the free energy of internal bonds $\Delta\omega$ Eq. (D2); the definition of $\Delta\omega$ is given in Eq. (B2). We call this macroscopic assembly the gel phase, in agreement with previous literature [39–42]. Please note that, since we do not explicitly include the solvent in assembly formation (see reaction scheme in **Fig. 1a**), in our model the gel corresponds to a phase without solvent, $\Phi_{tot} = 1$. To account for biological gels that can be rich in water, our theory can be straightforwardly extended by incorporating the solvent into the reaction scheme.

We now focus on systems that phase separate as the result of interactions with the solvent ($\chi \neq 0$ in **Eq. (9)**) and discuss the interplay between phase separation and gelation. Volume fractions in the coexisting phases are determined by **Eq. (7)** and assembly equilibrium requires that **Eq. (3)** to be satisfied. As pointed out in Sec. II, we aim to find an expression for $\Phi_i(\Phi_{tot})$ via **Eq. (3)** and **Eq. (5)**, and then substitute it into the free energy **Eq. (1)**. However, for planar ($d = 2$) and three-dimensional assemblies, ($d = 3$), performing the thermo-dynamic limit $M \to \infty$ leads to a free energy composed of series that diverges in the thermodynamic limit. We know that this divergence is physical, and is caused by the gelation transition. The divergence can be resolved by introducing explicitly a term in the free energy that accounts for an infinite-sized a ssembly – t he g el. Thus, we write the system free energy as a composition of the solution free energy $f_{sol}$ and the gel free energy $f_{gel}$:

$$f = f_{sol} + f_{gel}\,, \qquad (12)$$

where $f_{sol}$ is defined in **Eq. (1)**. The gel free energy reads

$$f_{gel} = \frac{\omega_\infty}{\nu_1}\,\delta\left(1-\phi_{tot}\right)\,, \qquad (13)$$

with $\delta(\cdot)$ denoting the delta distribution. The gel free energy $f_{gel}$ is the free energy of a state with no solvent, where all monomers belong to an assembly of size $i\to\infty$. In fact, in the limit $\Phi_i = 0$ for all finite $i$ and $\Phi_{tot} = 1$, the free energy in **Eq. (1)** simplifies to the single contribution $\omega_\infty/\nu_1$. This observation sheds light on $\omega_\infty$, which has the physical interpretation of free energy associated with each bond among monomers belonging to the gel. For this reason, we chose $\omega_\infty$ to be proportional to the bond free energy among monomers in solution ($e_{int} - Ts_{int}$); see Appendix D for more details.

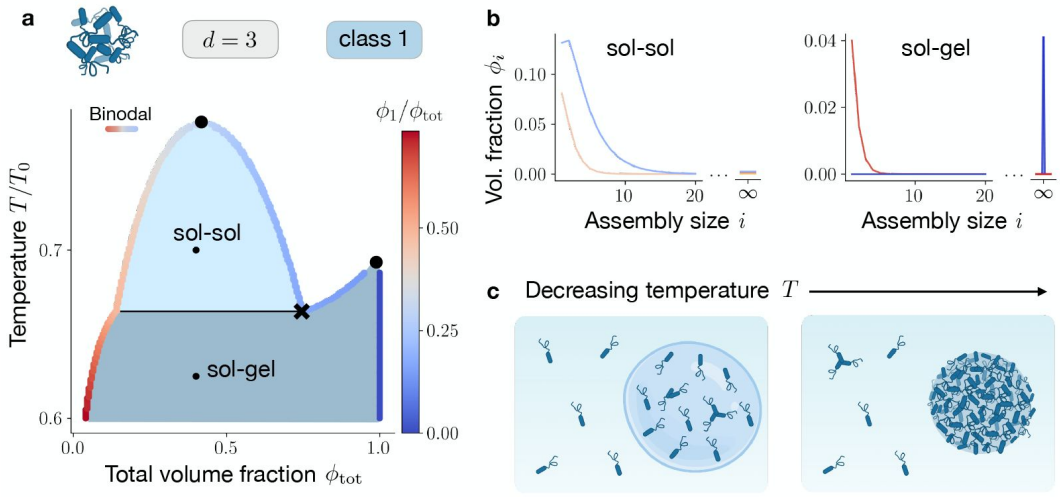
We can now perform a Maxwell construction by using **Eq. (12)** in **Eq. (7)**. The resulting phase diagram is displayed in **Fig. 3a**, where the binodal is coloured by the monomer fraction $\Phi_1/\Phi_{tot}$ in the coexisting phases. In phase-separated systems, gelation can be considered as a special case of phase coexistence between a protein-poor phase ("sol"), in which $\Phi^{sol} < 1$, and the gel phase, corresponding to $\Phi^{gel} = 1$. The domain in the phase diagram where a gel phase coexists with a soluble phase is shaded in blue and labelled as "sol-gel" in **Fig. 3a**. In the same panel, we show that lowering the tem[1] perature for large $\Phi_{tot}$ leads to a transition from the homogeneous state to the sol-gel coexistence. By contrast, for intermediate volume fractions, the system transits first through a domain corresponding to two-phase coexistence; see light blue domain labelled as "sol-sol" in **Fig. 3a**, where $\Phi_{tot} < 1$ in both phases. At the triple point (marked with the black cross) the gel phase of volume fraction $\Phi_{tot} = 1$ coexists with two "sol" phases, for which $\Phi_{tot} < 1$. In **Fig. 3b**, we show assembly size distributions representative of the "sol-sol" and "sol-gel" regions. The transition from the "sol-sol" to the "sol-gel" region is accompanied by a jump in the total volume fraction of the proteinrich phase $\phi_{tot}^{I}$, while the value in the protein-poor phase $\phi_{tot}^{II}$ changes smoothly. This finding confirms that it is the protein-rich phase that gelates; see **Fig. 3c** for an illustration. Having characterised the equilibrium of the mixtures belonging to different classes, we continue with the kinetics of assembly and phase separation in the next section.

## VI. Kinetic theory of assembly at phase equilibrium

Building upon the thermodynamic framework discussed in the previous sections, we devise a non-equilibrium kinetic theory for molecular assembly at non-dilute conditions, where the interactions can give rise to coexisting phases. Here, we restrict ourselves to the case where each phase is homogeneous and at phase equilibrium but not at assembly equilibrium [61], i.e., Eq. (6) is fulfilled during the kinetics while Eq. (3) is not satisfied in general. This partial equilibrium holds when the molecular transitions among assemblies are slow compared to phase separation. This case is often referred to reaction-limited [62, 63] and applies particularly well to molecular assemblies involving biological enzymes [64]. For simplicity, we present the kinetic theory and discuss the results for two coexisting phases.

We tailor the concepts developed in Ref. [61] to the case of incompressible systems, $dv_i/dt = 0$ and $dv_s/dt = 0$, and volume conserving assembly kinetics, $\sum_{i=1}^{M} r_i^{I/II} = 0$, where $r_i^{I/II}$ denotes the assembly rate of assembly $i$ in each phase. In this case, the total system volume $V = V^{I} + V^{II}$ is constant, i.e., $dV/dt = 0$, and the volume fractions of the assembly of size $i$, $\phi_i^{I/II}$, is governed by:

$$\frac{d}{dt}\phi_i^{I/II} = r_i^{I/II} - j_i^{I/II} - \frac{\phi_i^{I/II}}{V^{I/II}}\frac{d}{dt}V^{I/II}\,, \qquad (14)$$




## Figure 3

### Gelation transition in phase-separating systems.

**a** Phase diagram for planar (d=2) and three-dimensional (d=3) assemblies in the limit $M \to \infty$, as a function of $\Phi_{\rm tot}$ and the rescaled temperature $T/T_0$ (with $T_0 = \chi/k_B$). The coloured curve represents the binodal associated with the free energy $f$, which accounts for the emergence of an infinite assembly. The colour code of the binodal line depicts the monomer fraction $\Phi_1/\Phi_{\rm tot}$ in the phases, and the black cross 'x' indicates the triple point. In the region labelled as "sol-sol", the system demixes into two phases both populated mainly by monomers, see panel **b**, with $\phi_{\rm tot}^{\rm I/II} < 1$. In the region labelled as "sol-gel", on the other hand, a phase (the "sol"), obeying $\phi_{\rm tot}^{\rm sol} < 1$, coexists with a phase (the "gel") that is a macroscopic assembly, containing no solvent $(\phi_{\rm tot}^{\rm sol} = 1)$. The latter scenario is represented in panel **b**, right side. **c** Lowering the temperature allows transitions from the "sol-sol" to the "sol-gel" region, which manifest with a jump in the total volume fraction of the protein-rich phase.

while the solvent volume fraction in each phase is given as $\phi_{\rm s}^{\rm I/II} = 1 - \phi_{\rm tot}^{\rm I/II}$ with $\phi_{\rm tot}^{\rm I/II} = \sum_{i=1}^{M} \phi_i^{\rm I/II}$. Eq. (14) states that the volume fraction of assemblies in each phase $\phi_i^{\rm I/II}$ can vary due to three factors: the formation or dissolution of assemblies within the same phase (first term on the r.h.s), diffusion through the phase boundary (second term on the r.h.s), where $j_i^{\rm I/II}$ denote the diffusive exchange rates between the phases, and changes of the respective phase volumes $V^{\rm I/II}$ (last term on the r.h.s.). For more information, we refer the reader to Appendix E. The kinetics of phase volumes follows

$$\frac{d}{dt}\left(\ln\frac{V^{\rm I/II}}{V}\right) = -j_{\rm s}^{\rm I/II} - \sum_{i=1}^{M} j_i^{\rm I/II}\,. \qquad (15)$$

Moreover, mass conservation at the interface implies that the diffusive exchange rates of assemblies in the two phases are related via

$$j_i^{\rm I} = -j_i^{\rm II}\frac{V^{\rm II}}{V^{\rm I}} \qquad (16)$$

and analogously for the solvent $j_{\rm s}^{\rm I} = -j_{\rm s}^{\rm II}V^{\rm II}/V^{\rm I}$. Thus, the assembly kinetics conserves the total volume fraction defined as $\phi_{\rm tot} = \left(\phi_{\rm tot}^{\rm I}V^{\rm I} + \phi_{\rm tot}^{\rm II}V^{\rm II}\right)/V = 1 - \phi_{\rm s}$.

The exchange rates are determined by the conditions that maintain phase equilibrium, $d\mu_i^{\rm I}/dt = d\mu_i^{\rm II}/dt$ and $d\Pi^{\rm I}/dt = d\Pi^{\rm II}/dt$, where $\mu_i^{\rm I/II}$ are the exchange chemical potentials of the monomers in an assembly of size $i$ (Eq. (2)), and $\Pi^{\rm I/II}$ are the osmotic pressures in each phase; for more information, see Appendix E.

Using our kinetic theory, we can study the relaxation toward thermodynamic equilibrium which corresponds to simultaneous phase and assembly equilibrium. To account for association and dissociation processes associated with the reaction scheme in **Fig. 1**, the phase-dependent net reaction rate for the formation of a ($i$+$j$)-mer starting from a $i$-mer and a $j$-mer and vice versa are set by the exchange chemical potentials via the following law of mass action [65]:

$$\Delta r_{ij}^{\rm I/II} = k_{ij}\left[\exp\left(\frac{i\mu_i + j\mu_j}{k_{\rm B}T}\right) - \exp\left(\frac{(i+j)\mu_{i+j}}{k_{\rm B}T}\right)\right], \qquad (17)$$

where $k_{ij}$ is a size-dependent kinetic rate coefficient, see Appendix E for details. The assembly rates $r_i^{\rm I/II}$ entering Eq. (14) can finally be expressed as a function of the ($i$ + $j$)-mer exchange rate $\Delta r_{ij}^{\rm I/II}$, by

$$r_i^{\rm I/II} = \frac{i}{2}\sum_{j,k:j+k=i}\Delta r_{jk}^{\rm I/II} - i\sum_j \Delta r_{ij}^{\rm I/II}\,. \qquad (18)$$

In the next section, we compare the kinetics of systems belonging to Class 1 and 2, for $d$ = 1.

## VII. Assembly kinetics in coexisting phases

By integrating Eq. (14) numerically, we obtain the time evolution of $\phi_i^{\rm I/II}(t)$ and $V^{\rm I}(t)$, provided their initial values at $t$ = 0, $V^{\rm I}(t=0)$, and $\phi_i^{\rm I/II}(t=0)$, at phase equilibrium. Specifically, we consider an initial state solely composed of solvent and monomers demixed into a monomer-rich

and a monomer-poor phase (labelled with I and II respectively, see the illustration in **Fig. 4a**). For simplicity, we focus on linear assemblies ($d = 1$) and highlight differences between Class 1 and 2; for parameters see caption of **Fig. 4**. We note that the kinetics for $d > 1$, where gelation can occur, would require removing the upper bound in assembly size, i.e., studying trajectories in an $M$-dimensional space, where $M \to \infty$ in the thermo-dynamic limit. This case is numerically challenging and we leave its investigation for future work.

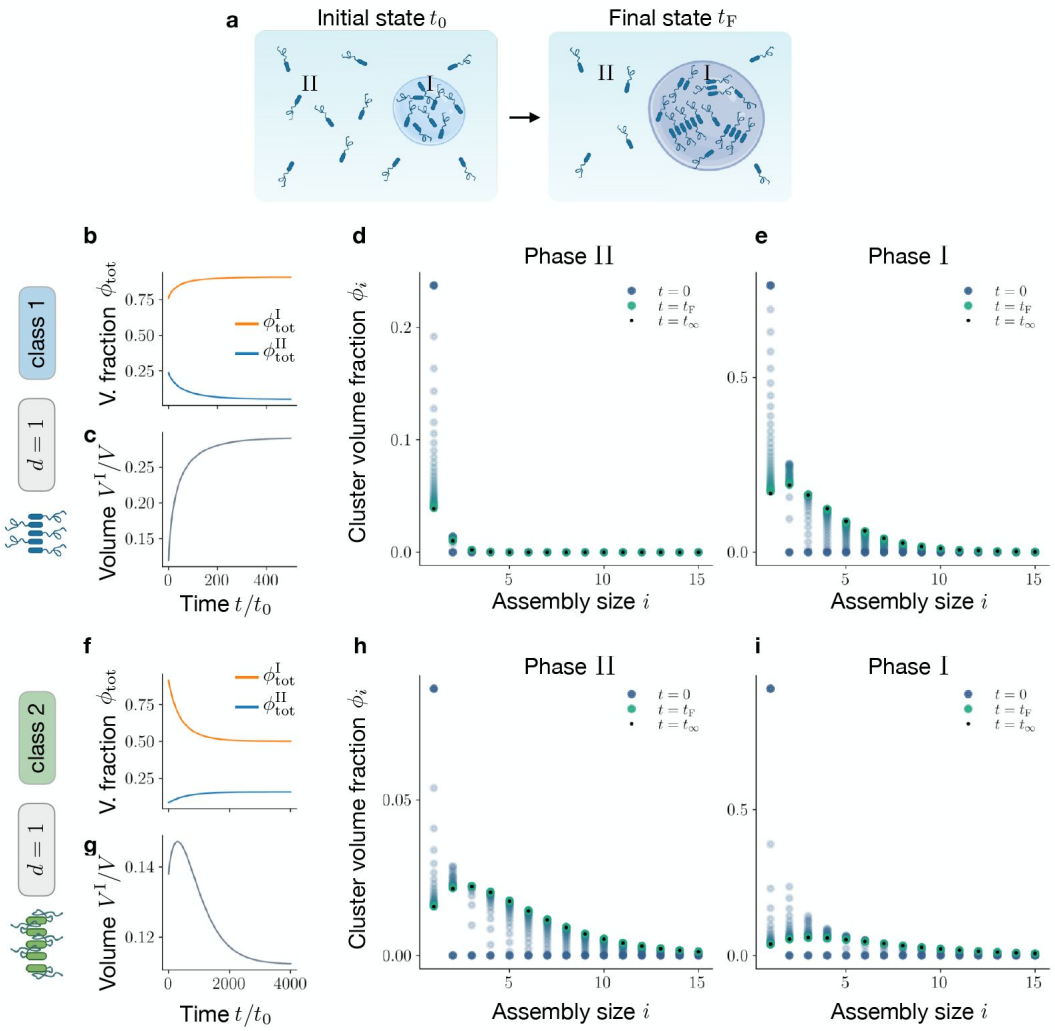
For Class 1, as monomers start forming assemblies, the mixing entropy decreases. As a result, the total amount of protein in the monomer-rich phase, $\phi_{\mathrm{tot}}^{\mathrm{I}}$, increases while $\phi_{\mathrm{tot}}^{\mathrm{II}}$ decreases (**Fig. 4b**). Such changes in total protein volume fractions induce phase volume variations (**Fig. 4c**). In particular, remaining within Class 1, since the monomer enrichment of phase I is less pronounced than the monomer depletion of phase II, the volume of the protein-rich phase $V^{\mathrm{I}}$ increases. An important finding of our work is that the distribution of assembly size evolves differently in each phase (**Fig. 4d,e**; and SI Movie 1). In phase II, which is initially poor in monomers, assemblies grow slowly toward an equilibrium distribution where the volume fractions monotonously decrease with assembly size, following an exponential decay. The kinetics in the initially monomer-rich phase I is fundamentally different. First, a very pronounced peak of intermediate-sized assemblies develops quickly. The faster kinetics compared to phase II is caused by monomer diffusion from II to I, which leads to negative feedback for assembly in II and positive feedback in I. This observation is reminiscent of studies on dilute, irreversible aggregation in coexisting phases [43]. The most abundant populations of intermediate-sized assemblies shrink slowly in time feeding the growth of larger assemblies. The resulting equilibrium distribution shows a notable peak of intermediate-sized assemblies followed by an exponential decay. Thus, the difference in the kinetics between the phases is dominantly a consequence of the fact that each phase strives towards a significantly different equilibrium distribution.

Assemblies belonging to Class 2, exhibit a different behaviour. In this class, as monomers assemble, their interaction propensity decreases. As a result, depending on the values of $\chi$ and $\chi'$, the total amount of protein in the protein-rich phase, $\phi_{\mathrm{tot}}^{\mathrm{I}}$, can decrease, as in the case of **Fig. 4f**. For this choice of parameters, the total amount of protein in the protein-poor phase, $\phi_{\mathrm{tot}}^{\mathrm{II}}$, increases, see again **Fig. 4f**. Furthermore, in this case, the initial rapid decrease in $\phi_{\mathrm{tot}}^{\mathrm{I}}$, followed by more moderate changes at later times, induces a non-monotonic phase volume variation (**Fig. 4g**), leading to a net shrinkage of phase I. Further investigation is needed to shed light on the physical mechanisms underlying this non-monotonic volume variation. In **Fig. 4h,i**, and SI Movie 2, we show how the volume fractions $\Phi_i(t)^{\mathrm{I/II}}$ for each assembly size $i$ evolve in both phases I and II. In the next section, we explore under which conditions phases grow or shrink during the relaxation to equilibrium.

## VIII. Assembly formation can increase or decrease condensate volume

Here, we discuss changes in phase volumes caused by the assembly kinetics introduced in Sec. VI. In particular, we focus on mixtures initially demixed in two phases, both composed of only monomers, and let the system relax to thermodynamic equilibrium. We then assess for which values of the control parameters $\Phi_{\mathrm{tot}}$ and $T$, the formation of assemblies in both phases leads to a growth of the $\Phi_{\mathrm{tot}}$-rich phase (phase I) and vice versa. More-over, we distinguish the two protein classes introduced in Sec. III.

To this end, we compare the phase diagram corresponding to the initial system, composed of monomers only, with the equilibrium phase diagram in which large assemblies populate the mixture. In **Fig. 5a**, we show the initial and final equilibrium binodals (black and coloured




## Figure 4

### Assembly kinetics at phase equilibrium.

Assuming that the relaxation to phase equilibrium is fast compared to assembly kinetics, we study the slow relaxation to assembly equilibrium in a compartmentalized system. **a** In the sketch, starting from an initial state composed of monomers and solvent only, assemblies selectively appear in phase I, increasing its volume $V^I$ and total volume fraction $\phi^I_{tot}$. **b, c** For Class 1, as time proceeds, the total macromolecule volume fraction in the two phases, $\phi^{I/II}_{tot}$, changes inducing the growth of phase I. In **d** and **e** we show the time evolution of the full-size distribution in phase II and I, respectively. **f, g** For Class 2, as time proceeds, changes in total macromolecule volume fraction in the two phases cause a shrinkage of phase I. This is reminiscent of recent experimental findings that quantify droplet volume changes along with droplet ageing [60]. **h, i** time evolution of assembly volume fractions $\Phi_i(t)$ in phase II and I, respectively. Time is measured in units of the discretization time step $t_0 = 0.01/\tilde{k}$, where the rate $\tilde{k}$ is introduced in Eq. (F5).

curve, respectively), for the case of linear assemblies ($d$ = 1) belonging to class 1. In this case, the domain corresponding to demixing enlarges once the system reaches its equilibrium state, i.e., assembly facilitates phase separation. We focus on the $\Phi_{\mathrm{tot}}$-$T$ domain enclosed by the black curve, where the system is phase separated at all times, and compute the initial and final $\Phi_{\mathrm{tot}}$-rich phase volumes via the total volume fraction conservation $V^I(t)/V = (\phi_{\mathrm{tot}} - \phi^{\mathrm{II}}_{\mathrm{tot}}(t))/(\phi^{\mathrm{I}}_{\mathrm{tot}}(t) - \phi^{\mathrm{II}}_{\mathrm{tot}}(t))$. As displayed in **Fig. 5a**, this allows us to identify two parameter regimes: at low $\Phi_{\mathrm{tot}}$ (orange area), the protein-rich phase grows as assemblies form, while above the dashed grey line (light blue area), it shrinks. Remarkably, linear assemblies ($d$ = 1) belonging to class 2 exhibit a completely different behaviour, see **Fig. 5b**. In this case, assembly formation shrinks the domain corresponding to demixing, thereby suppressing phase separation. In the domain enclosing the coloured curve, we can compute the initial and final volumes of the protein-rich phase for each value of $\Phi_{\mathrm{tot}}$ and $T$. In contrast to the previous case, we find that at low $\Phi_{\mathrm{tot}}$ (light blue area), the protein-rich phase shrinks as assemblies are formed, while for higher $\Phi_{\mathrm{tot}}$ values (orange area) condensate volume grows, as illustrated in **Fig. 5b**.

## IX. Conclusion

We discuss an extension of the classical theory of molecular assembly [14–16] to non-dilute conditions and study it for case where assemblies can phase-separate from the solvent and gelate. This extension relies on a thermodynamic free energy governing the interactions among all assemblies of different sizes and the solvent. We propose two classes to account for protein interactions relevant to biological systems that can phase separate and form assemblies. Classes differ in the way how energetic parameters for interactions and internal free energies depend on assembly size.

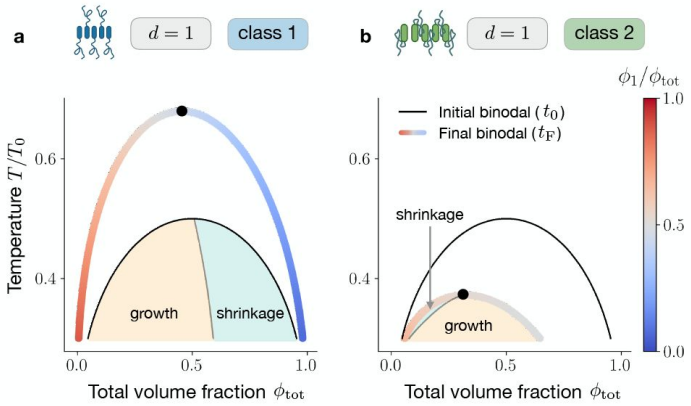
Using our theory, we report several key findings that arise from non-dilute conditions and the ability of assemblies to form a condensed phase. First, size distributions, in general, differ between the phases. In particular, monomers are not necessarily the most abundant species, and distribution tails can significantly deviate from the exponential decay known for classical assembly at dilute conditions [15]. Interestingly, this statement also applies to conditions below the saturation volume fraction beyond which phase separation can occur. Second, we showed that by lowering the temperature, the protein-rich phase can gelate, i.e., it consists of a single connected assembly of volume equal to the protein-rich phase (a gel). Upon gelation, the composition of the protein-poor phase changes continuously, while the protein-rich liquid phase discontinuously transits to the gel phase. Third, when monomers start assembling in the respective phases, the volume of the protein-rich phase can grow or shrink depending on the molecular interactions among the constituents.

Our key findings are consistent with recent experimental observations in living cells and in vitro assays using purified proteins. A decrease in droplet volume has been observed in phase-separated condensates composed of purified FUS proteins [60]. Up to now, it has remained unclear whether this kinetics relies on a glass transition as suggested in the discussion of Ref. [60], or on the formation of FUS oligomers in the protein-rich phase. However, a potential hint comes from independent studies, which indicate that FUS can form amyloid-like assemblies, that are associated with neurodegenerative disorders [6], at similar conditions [66, 67]. Moreover, the gelation of dense protein condensates upon temperature and heat stress was suggested in several in vivo studies in living cells [28]. The transition to a gelated condensate is believed to provide a protection mechanism for the protein expression machinery in the case of intracellular stress. Recently, in vitro experiments using purified proteins indicate anomalous size distributions of phase-separating proteins below saturation [26]. Our theoretically predicted size distributions could be compared to systematic experimental studies using single molecule techniques such as FRET. From this comparison, protein interactions of assembly-prone and

## Figure 5

### Identification of shrinkage and growth regions for different classes.

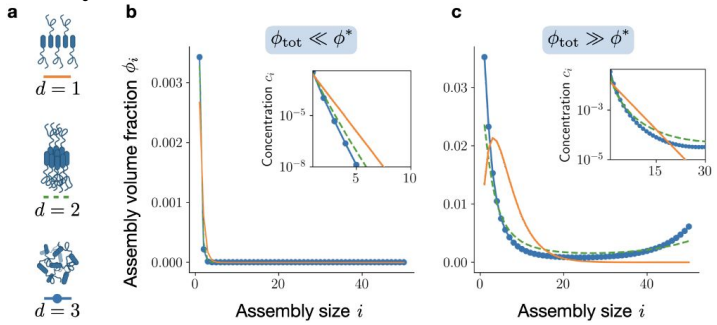
Here, we study phase-separating systems initially composed of monomers only and we monitor phase volume changes as they relax to thermodynamic equilibrium. **a** For linear assemblies (d=1) belonging to class 1 the final binodal line (coloured curve) is wider than the initial one (black curve), corresponding to monomers and solvent only (black curve). Areas in orange and light blue correspond to growth and shrinkage of the $\Phi_{tot}$-rich phase (phase I), respectively. **b** The behaviour of linear assemblies (d=1) belonging belonging to class 2 is remarkably different. Since, in this class, the interaction with the solvent is screened, the final binodal is shrunk compared to the initial one. As a consequence of the shrinkage, the domain corresponding to phase I growth (light blue area) precedes in $\Phi_{tot}$ the shrinkage domain (orange area), for class 2.




## Figure 6

### A volume fraction threshold separates two assembly regimes in homogeneous systems.

**a** Illustration of assemblies belonging to Class 1 with different spatial dimension. **b** Assembly size distribution at low total macromolecular volume fraction: $\Phi_{tot} = 0.2\Phi^*$. Disregarding assembly dimension, $d$, the macromolecules are mainly in the monomer state, i.e., $\Phi_1 \simeq \Phi_{tot}$. **c** For $\Phi_{tot} = 10\Phi^*$, the monomer volume fraction saturates at $\Phi_1 \simeq \Phi^*$ and big assemblys begin to populate the system. For linear assemblies (corresponding to $d = 1$ in **Eq. (8)**), the distribution becomes peaked at an intermediate value $i_{max} > 1$ and then exponentially cut off. For planar and three-dimensional assemblies, $d = 2, 3$, the distribution becomes bimodal, with peaks at $i = 1$ and $i = M$, the maximum assembly size ($M = 50$). This bimodal behaviour hints at the emergence of a gelation transition in the limit $M \to \infty$. In the insets, we show the scaling of concentrations $c_i = \Phi_i/\nu_i$ with assembly size. For $d = 2, 3$ and above the $\Phi^*$ threshold, deviations from the classical exponential decay are present. Here $e_{int}/\chi = -1$, $s_{int}/k_B = -1$, $M = 50$, $T/T_0 = 0.25$.



phase-separating proteins can be characterized using our proposed classes. Though many biologically relevant assembly processes are reversible and governed by thermodynamic principles, there are also a large number of assemblies that are persistently maintained away from equilibrium. For example, the formation or disassembly of assemblies can depend on the hydrolysis of ATP [68] while it can also act as cosolute [69, 70]. Since fuel levels are approximately kept constant in living cells, fuel-driven assembly processes are maintained away from equilibrium and thus cannot relax to thermodynamic equilibrium. It is an exciting extension of our work to consider fuel and waste components and how distributions of assembly sizes and the gelation of condensates are affected when maintained away from equilibrium.

## Acknowledgements

We thank J. Bauermann, K. Alameh, P. McCall, T. Harmon, L. Hubatsch, L. Jawerth, and F. Jülicher for fruitful discussions about the topic. We thank C. Seidel and T. Franzmann for pointing out the relevance of your theory for protein aggregation in biomolecular condensates. We acknowledge J.-F. Joanny for pointing out the references [37, 41]. We thank S. Safran for the very in-sightful feedback on the manuscript. We thank J. Bauer-mann, S. Horvát and C. Duclut for help improving the Mathematica code. G. Bartolucci and C. Weber acknowledge the SPP 2191 “Molecular Mechanisms of Functional Phase Separation” of the German Science Foundation for financial support. C. Weber acknowledges the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Fuelled Life, Grant Number 949021) for financial support. Figures created with BioRender.com.

## Appendix A: Suplementary information and parameters used

Parameters used in each figure is shown in **table I**.

**Movie 1** shows the time evolution of the assembly volume fractions in both phases, for Class 1. Parameters are the same of 4b-e.

**Movie 2** shows the time evolution of the assembly volume fractions in both phases, for Class 2. Parameters are the same of 4f-i.

| Figure | $e_{\rm int}/\chi$ | $s_{\rm int}/k_{\rm B}$ | $\chi'/\chi$ | $M$ | $\phi_{\rm tot}$ | $T/T_0$ |
|---|---|---|---|---|---|---|
| 2a | -5 | -2 | - | $\infty$ | - | - |
| 2b | -0.5 | -2 | - | $\infty$ | - | - |
| 2c | - | -2 | - | $\infty$ | - | 0.6 |
| 2d | - | -2 | 0.2 | $\infty$ | - | 0.25 |
| 3d | -2 | -2 | - | $\infty$ | - | - |
| 4b-e | -1 | -2 | - | 15 | 0.3 | 0.45 |
| 4f-i | -1 | -2 | 0.2 | 15 | 0.2 | 0.35 |
| 5a | -1 | -2 | - | 15 | - | - |
| 5b | -1 | -2 | 0.2 | 15 | - | - |

**Table I**

**Parameters corresponding to the figures in the main text. We made use of the temperature scale $T_0 = \chi/k_B$.**

# Appendix B: Scaling laws for internal and interaction energies

Here we provide a physical interpretation of the internal free energy $\omega_i$. For simplicity, we consider a homogeneous system solely composed of assemblies of size $i$, characterized by the volume fraction vector $\Phi^{(i)}$, with $\phi_i^{(i)} = 1$ and $\phi_j^{(i)} = 0$ for $i \neq j$. Making use of **Eq. (1)**, the internal free energy of such systems can be written as

$$\omega_i = f(\phi^{(i)})\nu_1 - k_B T\, i^{-1} \ln i^{-1}\,, \qquad \text{(B1)}$$

with $f(\Phi^{(i)})\nu_1$ being the free energy associated with each monomer belonging to the $i$-th assembly. The second term in the equation above is the conformational entropy that stems from having more accessible states with increasing assembly size. Thus, **Eq. (B1)** allows interpreting $\omega_i$ as the free energy of monomers inside an assembly of size $i$, coming only from bonds between monomers. To quantify it, we introduce the number of binding sites for each monomer, $z$. Following Ref. [16], we distinguish between $n_b$ monomers at the boundaries of the assembly, and $(i - n_b)$ in the assembly bulk. Monomers in the bulk can saturate all their $z$ binding sites while, in general, monomers at the boundaries are able to saturate only $z_b < z$. Thus, we get

$$\begin{aligned}\omega_i &= \frac{z(i-n_b) + z_b n_b}{2i}\,\Delta\omega \\ &= \frac{z}{2}\,\Delta\omega - (z - z_b)\frac{n_b}{2i}\,\Delta\omega\,,\end{aligned} \qquad \text{(B2)}$$

where $\Delta\omega$ is the free energy associated with the formation of a single bond, that is composed of an energetic and an entropic part. The factor two avoids double counting.

We describe three species of assemblies: linear, disc-like and three-dimensional. These can be realised by varying the number of binding sites and their orientation [71]. Linear assemblies ($d = 1$) are defined to have only two binding sites. They can be pictured as one-dimensional semi-flexible assemblies with no loops, leading to $n_b = 2$, $z = 2$ and $z_b = 1$. Planar assemblies ($d = 2$) are defined to have $z > 2$ co-planar binding sites, for which $n_b \simeq \sqrt{i}$. Three-dimensional assemblies ($d = 3$) are characterized by $z > 2$ binding sites with no precise orientation leading to $n_b \simeq i^{\frac{2}{3}}$. Summing up, we get

$$n_b \simeq i^{\frac{d-1}{d}}\,, \qquad \text{(B3)}$$

that inserted in Eq. (B2), decomposing $\Delta\omega$ in its ener-getic and entropic contribution, gives

$$\omega_i = \alpha\,(e_{int} - T s_{int}) - \left(\frac{e_{int} - T s_{int}}{i^{1/d}}\right)\,, \qquad \text{(B4)}$$

where $\alpha$ is a constant that depends on number and geometry of the binding sites. Identifying $\omega_\infty = \alpha\,(e_{int} - T s_{int})$, Eq. (B4) leads to Eq. (8), in the main text. The constant terms $\omega_\infty$, does not affect chemical nor phase equilibrium. However, in the case of $d = 2$, 3 and $M \to \infty$, $\omega_\infty$ it becomes important to study the gelation of the protein-rich phase, see Appendix D. In Eq. (8), the second term represents a boundary interaction penalty, accounting for the fact that monomers at the assembly boundary can realise fewer internal bonds than monomers at the assembly bulk, in analogy with the physical origin of surface tension.

We now discuss the size dependence of the interaction parameters $\chi_{ij}$. Starting from a lattice model, these parameters can be expressed in terms of the energetic parameters $e_{ij}$ corresponding to having two neighbouring monomers belonging to $i$ and $j$. In particular, $\chi_{ij} = 2e_{ij} - e_{ii} - e_{jj}$. Assuming that the energies associated with monomer-monomer interactions do not vary within assemblies, i.e., $e_{ij} = e_{11}$ is constant, we get $\chi_{ij} = 0$. Moreover, we now discuss the scaling of $\chi_{is} = 2e_{is} - e_{ii} - e_{ss}$. If the monomer-solvent interactions are also chosen to be size-independent, i.e., $e_{is} = e_{1s}$, we get $\chi_{is} = 2e_{1s} - e_{11} - e_{ss} = \chi$. This explains the scaling in Class 1 (see Eq. (9)).

However, many proteins of interest screen their hydrophobic interaction when forming assemblies [2, 49, 57] implying that the interactions between monomers in assembly $i$ with solvent (s) $e_{is}$ varies with assembly size $i$. In each assembly, this energy per monomer comes from two contributions. The first corresponds to monomers in the bulk which are $(n - n_b)$ and have interaction with solvent $e'_{1s}$. The second one corresponds to the $n_b$ monomers at the assembly boundary, characterised by interaction with solvent $e_{1s}$. We get

$$e_{is} = \frac{e'_{1s}(i - n_b) + e_{1s}n_b}{i}. \qquad \text{(B5)}$$

Using the scaling of $n_b/i$ already introduced above in the discussion of the internal free energy scaling, see B3, we obtain Eq. (10). By abbreviating $\chi'_{is} = 2e'_{1s} - e_{11} - e_{ss} = \chi'$; this case corresponds to Class 2.

# Appendix C: Linear assemblies belonging to class 1

For class 1 and $d = 1$, Eq. (4) reads

$$\phi_i = i\left(\frac{\phi_1}{\tilde{\phi}}\right)^i \tilde{\phi}, \qquad \text{(C1)}$$

where we have introduced the characteristic volume fraction

$$\tilde{\phi} = \exp\left(\frac{e_{\text{int}} - s_{\text{int}}T}{k_B T} - 1\right). \qquad \text{(C2)}$$

It is straightforward to verify that the latter volume fraction is proportional to the assembly threshold defined in Eq. (11), i.e. $\tilde{\phi} = \mathrm{e}/(1-\mathrm{e})^2\phi^* \sim \phi^*$.

In **Fig. 6**, we show the assembly size distribution in homogeneous mixtures obtained by numerically solving **Eq. (4)** together with **Eq. (5)**, with a cut-off $M = 50$. We characterise the behaviour of assemblies with different spatial dimensions $d = 1, 2, 3$, see **Fig. 6a**. For dilute solutions, corresponding to $\Phi_{\text{tot}} \ll \Phi^*$, the size distribution is dominated by monomers while larger assemblies have vanishing volume fraction, i.e., $\Phi_1 \simeq \Phi_{\text{tot}}$, see **Fig. 6b**. For $\Phi_{\text{tot}} \gg \Phi^*$, the monomer volume fraction saturates at $\Phi_1 \simeq \Phi^*$ and assemblies begin to populate the system. As depicted in **Fig. 6b**, above this threshold the size distribution depends crucially on assembly dimension $d$. For linear assemblies ($d = 1$ in **Eq. (8)**), the distribution becomes peaked at a value $M > 1$ and then exponentially decays. For planar and three-dimensional assemblies, $d = 2, 3$ in **Eq. (8)**, the distribution becomes bimodal peaked at $i = 1$ and $i = M$, the maximum assembly size ($M = 50$ in **Fig. 6c**). The behaviour of the system at high density can be quantitatively studied by performing the thermodynamic limit, i.e., $M \to \infty$. Within this limit, the series defined in the conservation law, **Eq. (5)** can be explicitly solved, leading to

$$\frac{\phi_1}{\phi} = \frac{1 + 2\frac{\phi_{\text{tot}}}{\phi} - \sqrt{1 + 4\frac{\phi_{\text{tot}}}{\phi}}}{2\frac{\phi_{\text{tot}}}{\phi}}. \qquad \text{(C3)}$$

Recalling that $\tilde{\phi} = \mathrm{e}/(1-\mathrm{e})^2\phi^*$, this leads to $\Phi_1 \simeq \Phi_{\mathrm{tot}}$, in the regime $\Phi_{\mathrm{tot}} \ll \Phi^*$, while for $\Phi_{\mathrm{tot}} \gg \Phi^*$, we get $\Phi_1 \simeq \Phi^*$.

The maximum of the volume fraction distribution in **Eq. (C1)** can be obtained imposing $\partial_i \Phi_i = 0$, leading to

$$i_{\max} = \frac{1}{\ln\left(\tilde{\phi}/\phi_1\right)} \simeq \sqrt{\frac{\phi_{\mathrm{tot}}}{\tilde{\phi}}}\,. \qquad \text{(C4)}$$

The approximate expression on the right hand is obtained using **Eq. (C3)** and expanding for $\phi_{\mathrm{tot}}/\tilde{\phi} \gg 1$.

The average $\langle i \rangle = \Sigma i\Phi_i / \Sigma\Phi_i$ is given by

$$\langle i \rangle = \frac{\phi_1}{\phi_{\mathrm{tot}}} \frac{3-\phi_1/\tilde{\phi}}{(1-\phi_1/\tilde{\phi})^3} \simeq 2\, i_{\max}\,, \qquad \text{(C5)}$$

where we expanded for $\phi_{\mathrm{tot}}/\tilde{\phi} \gg 1$ to obtain the approximate expression.

We can also derive an explicit expression for the free energy as a function of the conserved quantity $\Phi_{\mathrm{tot}}$. This is achieved by plugging $\Phi_i$ given in **Eq. (C1)** into **Eq. (1)**. Omitting linear terms, which do not influence phase equilibrium, we get

$$f_{\mathrm{sol}} = \frac{k_{\mathrm{B}}T}{\nu_1}\left[(1-\phi_{\mathrm{tot}})\ln(1-\phi_{\mathrm{tot}}) + \phi_{\mathrm{tot}}\ln\frac{\phi_1}{\tilde{\phi}} - \frac{\phi_1}{1-\phi_1/\tilde{\phi}} + \frac{\chi}{k_{\mathrm{B}}T}\phi_{\mathrm{tot}}(1-\phi_{\mathrm{tot}})\right], \qquad \text{(C6)}$$

where the dependence of $\Phi_1$ on $\Phi_{\mathrm{tot}}$ is expressed in **Eq. (C3)**.

## Appendix D: Gelation transition for two and three-dimensional assemblies

Parameters are the same of **Fig. 3**, see **Table I**

As outlined in **Fig. 6,for** $d = 2, 3$, at high $\Phi_{\mathrm{tot}}$ for $M$ finite, the size distribution shows a bimodal behaviour. This suggests for the limit $M \to \infty$ that the system undergoes a gelation transition, which is defined as the emergence of an assembly that is comparable with the system size [16, 41, 42]. To estimate the $\Phi_{\mathrm{tot}}$ value at which the transition occurs, we recall Eq. (4) and consider the series

$$\sum_{i=1}^{\infty}\phi_i = \sum_{i=1}^{\infty} i\left(\frac{\phi_1}{\tilde{\phi}}\right)^i \exp\left(\frac{\Delta\omega}{k_{\mathrm{B}}T}\, i^{\frac{d-1}{d}} - 1\right). \qquad \text{(D1)}$$

We note that when $N \to \infty$, this series converges only if $\phi_1/\tilde{\phi} \leq 1$. Thus, we get an upper bound for the series, namely

$$\sum_{i=1}^{\infty}\phi_i \leq \sum_{i=1}^{\infty} i \exp\left(\frac{\Delta\omega}{k_{\mathrm{B}}T}\, i^{\frac{d-1}{d}} - 1\right) \equiv \phi^{\mathrm{sg}}. \qquad \text{(D2)}$$

Approximating the series with the integral, we get an estimation for $\Phi^{sg}$:

$$\phi^{sg} = \begin{cases} 2\frac{(6-6\Delta\omega+3\Delta\omega^2-\Delta\omega^3)}{\Delta\omega^4}\tilde{\phi}, & d=2, \\ -\frac{3}{2}\frac{\Delta\omega^3}{(2-2\Delta\omega+\Delta\omega^2)}\tilde{\phi}, & d=3. \end{cases} \qquad \text{(D3)}$$

By the Maxwell construction, Eq. (7) with the free energy Eq. (12), we can study the interplay between the gelation transition and phase separation. Here, the parameter $\omega_\infty$ plays a crucial role. As discussed in Appendix B, $\omega_\infty$ contains an energetic and an entropic part, and is proportional to $e_{int} - Ts_{int}$, the coefficient depending on assembly dimension, and number and geometry of the binding sites. Here, for simplicity, we set

$$\omega_\infty = \frac{1}{2}\left(e_{int} - Ts_{int}\right). \qquad \text{(D4)}$$

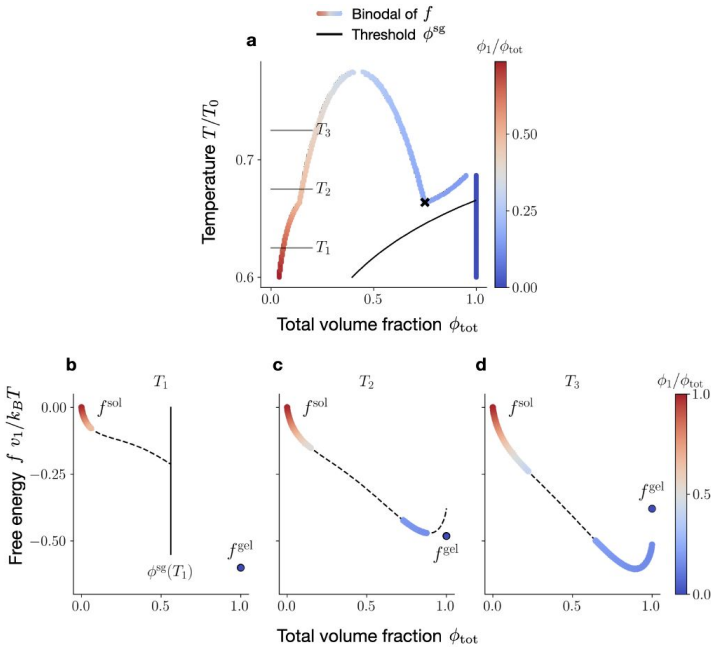
In **Fig. 7a**, we display the result of the construction (coloured curve), where the colour code depicts the monomer fraction $\Phi_1/\Phi_{tot}$ in the coexisting phases. Note that the branch of the binodal between the triple point (indicated with a cross) and $\Phi_{tot} = 1$, has the same trend as the curve stems from the estimate $\Phi^{sg}$ introduced in **Eq. (D3)** (black curve). This is expected since both curves correspond to the boundary between homogeneous mixtures and the gel state.

We also display the free energy for three temperature values corresponding to sol-gel coexistence (**Fig. 7b**), sol-sol and sol-gel coexistence (**Fig. 7c**), and sol-sol coexistence only (**Fig. 7d**). The dashed lines represent values where $f$ is not convex. Notice that, for consistency, we use values of $f_{sol}$ only up to $\Phi^{sg}$ (denoted by a vertical black line). This is because, as depicted in **Fig 6**, after this value a peak at $M$, the finite cut off used for the numerics, will appear.

## Appendix E: Mutual feedback between phase separation and assembly equilibria

We first discuss how assemblies can shape the phase diagram. For linear assemblies ($d = 1$) belonging to Class 1, assemblies facilitate phase separation. Indeed, as illustrated in **Fig. 8a-b**, increasing the relative assembly strength, i.e., decreasing $e_{int}/\chi$, leads to an upshift in critical temperature and a downshift in critical volume fraction. This trend can be explained by considering that assembly formation, even if energetically disfavoured, reduces the mixing entropy (see the first term in **Eq. (1)**). In **Fig. 8a**, we show the binodal lines corresponding to three representative values of the assembly strength: $e_{int}/\chi = 0, -1, -2$. We compare them to the black curve, which corresponds to a binary mixture made of monomers and solvent only (black curve). This reference case can be thought of as the limiting case in which assemblies have an infinite energy penalty, i.e, $e_{int}/\chi = \to \infty$. In **Fig. 8b**, we quantify the changes in critical temperature and critical volume fraction as a function of the relative assembly strength $e_{int}/\chi$. In **Fig. 8c-d**, we illustrate the behaviour of linear assemblies ($d = 1$) belonging to Class 2. In contrast to Class 2, assemblies can suppress phase separation. Indeed, making assemblies more favourable by decreasing $e_{int}/\chi$, the critical temperature decreases, and even if the critical density decreases and the binodal shrinks, see **Fig. 8c**. In **Fig. 8d**, we display critical temperatures and critical volume fraction variations as a function of the relative assembly strength $e_{int}/\chi$.

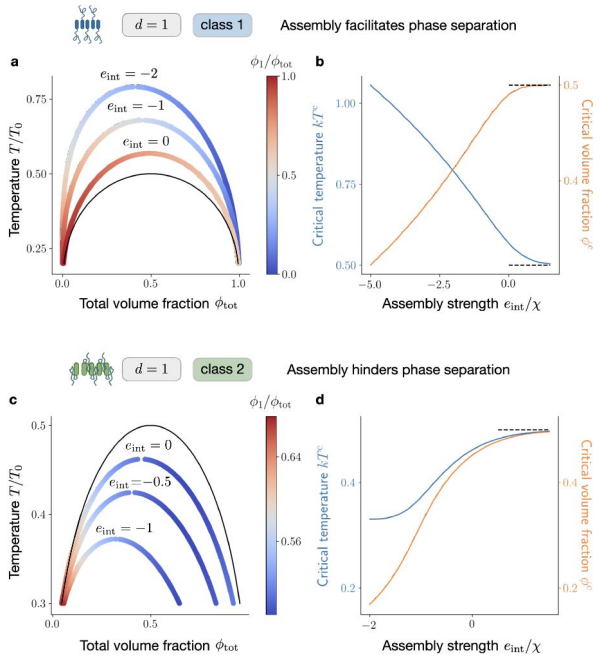
**Fig. 8** clearly shows that the presence of assemblies affects the phase equilibrium of a mixture. We now prove that, in turn, the total number of assemblies can differ between phase-separating and homogeneous systems with the same total protein volume fraction. To show this, we fix the interaction propensity $\chi$, the temperature $T/T_0$, and the total macromolecule volume fraction $\Phi_{tot}$ to values corresponding to two-phase coexistence at thermodynamic equilibrium. We then




## Figure 7

### Gel-sol free energies.

**a** The coloured curved indicates the binodal obtained with the Maxwell construction for $f = f_{sol} + f_{gel}$, together with the estimate $\Phi^{sg}(T)$ (black line, defined in **Eq. (D2)**) for the transition between homogeneous and gel states. The black cross 'x' indicates the triple point. **c-e** Maxwell construction for three different temperature values, the coloured and black, dashed curves represent convex and concave branches of $f$, respectively.




## Figure 8

### The influence of assemblies on the system phase behaviour.

**a** Focusing on systems with $d = 1$ belonging to class 1, we compare three binodals corresponding to assembly strength $e_{int}/\chi = 0.5, -1, -2$ (coloured curves) and the reference binary mixture composed of monomers and solvent only (black curve). The latter can be associated with the limit $e_{int}/\chi \to \infty$. The region enclosed by the binodal, corresponding to phase separation, expands even for assemblies with no assembly energy $e_{int}/\chi = 0$. This can be explained by the entropic advantage caused by size polydispersity. **b** Dependence of the critical volume fraction and critical temperature on the assembly strength $e_{int}/\chi$. The presence of assemblies causes $T^c$ and $\Phi^c$ to deviate from the reference values (black dashed lines) corresponding to a binary mixture with monomers and solvent only ($e_{int}/\chi \to \infty$). In particular, for Class 1, making assemblies more energetically favourable, i.e. decreasing $e_{int}/\chi$, induces an increase in $T^c$ and a decrease in $\Phi^c$, in turn making phase separation more accessible. Here $s_{int}/k_B = -2$, $M = \infty$. **c** Comparison between three binodal lines corresponding to systems belonging to class 2 and $d = 1$, with assembly energies $e_{int}/\chi = 0, -0.5, -1$ (coloured curves) and the reference binary mixture composed of monomers and solvent only (black curve). **d** For Class 2, decreasing $e_{int}/\chi$, causes $T^c$ and $\Phi^c$ to decrease, overall hindering phase separation. This is caused by the interaction propensity screening in monomers at the bulk of assemblies belonging to class 2, see **Eq. (10)**. Here $s_{int}/k_B = -2$, $M = \infty$. $\chi' = 0.2\chi$.

compare the assembly size distribution (after averaging over both phases), with the distribution in the corresponding homogeneous state, with the same values of $T$ and $\Phi_{\rm tot}$. Recalling that due to our choice of interaction propensity scaling in **Eq. (9)** that the size distribution in the homogeneous system, **Eq. (4)**, does not depend on $\chi$. For this reason, the homogeneous state can be thought of as an unstable state corresponding to the same $\chi$ as the phase separating one, which has not reached phase equilibrium yet, but also as the equilibrium state of a system with the same parameters as the phase separating one, but formed by assemblies that do not interact with the solvent ($\chi = 0$).

In **Fig. 9a**, we display results for linear assemblies ($d = 1$) with $T/T_0 = 0.2$, and $\Phi_{\rm tot} = 0.016$. We compare the size distribution in the homogeneous system $\phi_i^{\rm h}$, with the weighted average over compartments, defined as

$$\bar{\phi}_i = \frac{V^{\rm I}}{V}\phi_i^{\rm I} + \frac{V^{\rm II}}{V}\phi_i^{\rm II}\,, \tag{E1}$$

in the corresponding phase-separated system. Clearly, the two distributions differ, showing that the presence of compartments can lead to larger assemblies. The difference in size distributions can be quantified utilizing the so-called total variation distance, defined as

$$\delta(h,g) = \sup_i \left| \frac{h_i}{\sum_i h_i} - \frac{g_i}{\sum_i g_i} \right|. \tag{E2}$$

This quantity characterizes the distance between two normalised functions as the largest possible distance among values that they assign to the same argument. The distance between the homogeneous size distribution and the distribution defined in **Eq. (E1)** depends on the temperature $T$ and the total volume fraction $\Phi_{\rm tot}$, which in turn determines the droplet size. In **Fig. 9b**, we display distribution distances $\delta(\phi^{\rm h}, \bar{\phi})$ corresponding to different temperatures and droplet volumes. In the limits $V^{\rm I}/V \to 0$ and $V^{\rm I}/V \to 1$, the system becomes homogeneous. As a result, the distribution distance $\delta(\phi^{\rm h}, \bar{\phi})$ vanishes. Note that the volume corresponding to the maximum distribution distance shifts towards lower values.

## Appendix F: Assembly kinetics in homogeneous mixtures

In this section, we give the details on the kinetic theory for assembly in non-dilute homogeneous systems that can relax toward chemical equilibrium. Each component $i$ follows

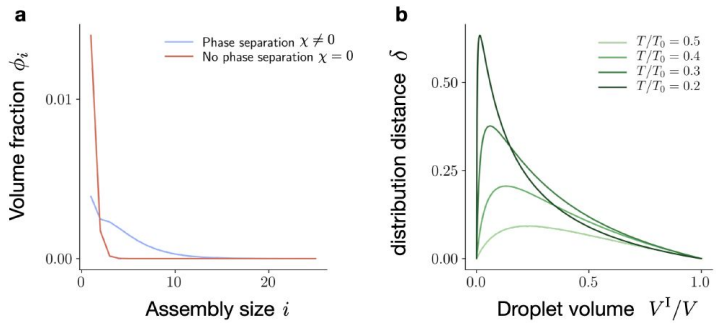
$$\frac{{\rm d}\phi_i}{{\rm d}t} = r_i\,. \tag{F1}$$

The assembly rates $r_i$ read

$$r_i = i\frac{1}{2}\sum_{j+k=i}\Delta r_{jk} - i\sum_j \Delta r_{ij}\,. \tag{F2}$$

These rates conserve the total volume fraction $\Phi_{\rm tot}$, i.e., $\partial_t \Phi_{\rm tot} = \Sigma_i\, r_i = 0$. The assembly flux between two assemblies of size $i$ and $j$, and the combined $(i+j)$-mer reads

$$\Delta r_{ij} = k_{ij}\left[\exp\left(\frac{i\mu_i + j\mu_j}{k_{\rm B}T}\right) - \exp\left(\frac{(i+j)\mu_{i+j}}{k_{\rm B}T}\right)\right], \tag{F3}$$




## Figure 9

### The influence of phase separation on assembly size.

**a** Comparison between the size distribution in a homogeneous system, and in the corresponding phase-separated system (averaged in both compartments). Here, we consider linear assemblies ($d$ = 1), $M \to \infty$, $\Phi_{\mathrm{tot}}$ = 0.016 and $T/T_0$ = 0.2. We note that the presence of compartments can favour assembly formation, even when the corresponding homogeneous mixture is populated mainly by monomers. The difference in distributions can be quantified utilizing the functional distance, defined in **Eq. (E2)**. **b** The magnitude of this distance depends on the droplet size and the temperature chosen. The volume corresponding to the maximum distribution distance shifts towards lower values with decreasing temperature $T/T_0$. The distributions separated by the maximum distance, for $T/T_0$ = 0.2, are the ones displayed in **a**. $e_{\mathrm{int}}/\chi$ = −0.5, $s_{\mathrm{int}}/k_{\mathrm{B}}$ = −2, $T/T_0$ = 0.25.

and is determined by differences in chemical potential per monomer. We now isolate the logarithmic part in the chemical potential and introduce chemical activities $\gamma_i$ via

$$\mu_i = i^{-1} \ln\left(\frac{\phi_i}{i}\gamma_i\right) . \qquad \text{(F4)}$$

Choosing $k_{ij} = 2\tilde{k}\exp\left[-[i\ln\gamma_i + j\ln\gamma_j]\right]$, we can recast the assembly flux in **Eq. (F2)** as

$$\Delta r_{ij} = 2\tilde{k}\left(\frac{\phi_i}{i}\frac{\phi_j}{j} - F_{ij}\frac{\phi_{i+j}}{i+j}\right) , \qquad \text{(F5)}$$

which leads to a finite flux $\Delta r_{ij}$ in the limit $\Phi_i \ll 1$. In the literature, $F_{ij}$ is known as fragmentation kernel, and in our case, it reads $F_{ij} = \exp\left[(i+j)\ln\gamma_{i+j} - i\ln\gamma_i - j\ln\gamma_j\right]$. For linear assemblies ($d = 1$) belonging to Class 1, the fragmentation kernel is constant $F_{ij} = \tilde{\phi} = \exp\left(\frac{\Delta\omega}{k_B T} - 1\right)$ in agreement with standard polymerization models [15]. For linear assemblies ($d = 1$) belonging to Class 2, $F_{ij} = F(\phi_{\rm tot}) = \exp\left(\frac{\Delta\omega}{k_B T} - 1 - \frac{\chi-\chi'}{k_B T}(1-\phi_{\rm tot})\right)$, i.e., the fragmentation kernel is still size independent but now depends on the total monomer volume fraction $\Phi_{\rm tot}$. Note that by making the kinetic rate coefficients corresponding to associations of small assemblies, i.e. $k_{ij}$ with $i, j \ll M$, explicitly dependent on the presence of large assemblies, our framework can be easily generalised to include primary and secondary nucleation [72]. Following again Ref. [15], we can express the time evolution of $\Phi_i(t)$ as follows:

$$\phi_i(t) = \phi_{\rm tot}\left(1-a(t)\right)^2 a(t)^{i-1} , \qquad \text{(F6)}$$

$$a(t) = \frac{1-\exp\left[-(\eta^{-1}-\eta)\phi_{\rm tot}t\right]}{\eta^{-1}-\eta\exp\left[-(\eta^{-1}-\eta)\phi_{\rm tot}t\right]} , \qquad \text{(F7)}$$

where $\eta$ is the following function of the fragmentation kernel $F$ ($\Phi_{\rm tot}$):

$$\eta = \frac{1+2\frac{\phi_{\rm tot}}{F} - \sqrt{1+4\frac{\phi_{\rm tot}}{F}}}{2\frac{\phi_{\rm tot}}{F}} . \qquad \text{(F8)}$$

## Appendix G: Assembly kinetics in phase-separated systems

Here, we generalise the assembly kinetics described in the previous section to the case of phase coexistence. To this end, we focus on passive systems that can relax to-ward thermodynamic equilibrium. Moreover, we restrict ourselves to systems that are at phase equilibrium at any time during the relaxation kinetics toward thermo-dynamic equilibrium and following the theory originally developed in Ref. [61]. Chemical kinetics constrained to phase equilibrium is valid if the chemical reaction rates are small compared to diffusion rates. By choosing initial average volume fractions corresponding to two-phase coexistence, we can consider the system volume to be divided into two homogeneous compartments as a result of phase separation. We then study the time evolution of compartment sizes and volume fractions due to chemical reactions, enforcing instantaneous phase equilibrium at all times. To this aim, we start with the variation of particle numbers in compartments I and II:

$$\frac{\mathrm{d}N_i}{\mathrm{d}t}^{\rm I/II} = -J_i^{\rm I/II} + R_i^{\rm I/II} , \qquad \text{(G1)}$$

where $R_i^{\mathrm{I/II}}$ are the variations due to chemical reactions and $J_i^{\mathrm{I/II}}$ describes the exchange of assemblies between the two phases. Particle conservation during crossing implies $J_i^{\mathrm{I}} = -J_i^{\mathrm{II}}$. Due to volume conservation in the two-phase, we have

$$V^{\mathrm{I/II}} = \sum_{i=0}^{M} N_i^{\mathrm{I/II}} \nu_i \,. \tag{G2}$$

Furthermore, $V = V^{\mathrm{I}} + V^{\mathrm{II}}$. We now introduce volume fractions $\phi_i^{\mathrm{I/II}} = N_i^{\mathrm{I/II}}/V^{\mathrm{I/II}}$ and the rescaled rates $j_i^{\mathrm{I/II}} = \nu_i J_i^{\mathrm{i/II}}/V^{\mathrm{I/II}}$ and $r_i^{\mathrm{I/II}} = \nu_i R_i^{\mathrm{I/II}}/V^{\mathrm{I/II}}$, leading to

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}t}^{\mathrm{I/II}} = -j_i^{\mathrm{I/II}} + r_i^{\mathrm{I/II}} - \phi_i^{\mathrm{I/II}} \frac{\mathrm{d}\ln V}{\mathrm{d}t}^{\mathrm{I/II}} , \tag{G3}$$

which correspond to Eq. (F1) generalised to two-phase coexistence. The rates in both phases $r_i^{\mathrm{I/II}}$ are given in Eq. (18). Eq. (G1) and Eq. (G2) can be combined to get $(d/dt)V^{\mathrm{I/II}} = V^{\mathrm{I/II}} \sum_{i=0}^{M} (r_i^{\mathrm{I/II}} - j_i^{\mathrm{I/II}})$. Using the volume conserving properties of the rates, $\sum_{i=1}^{M} r_i^{\mathrm{I/II}} = 0$ we finally obtain

$$\frac{\mathrm{d}\ln V}{\mathrm{d}t}^{\mathrm{I/II}} = -\sum_{i=0}^{M} j_i^{\mathrm{I/II}} \,. \tag{G4}$$

Assembly mass conservation at the interface implies

$$j_i^{\mathrm{I}} = -j_i^{\mathrm{II}} \frac{V^{\mathrm{II}}}{V^{\mathrm{I}}} \,, \tag{G5}$$

with the volume dynamics obeying $(d/dt)(V^{\mathrm{I}} + V^{\mathrm{II}}) = 0$. The currents $j_i^{\mathrm{I/II}}$ enforce that phase equilibrium is satisfied at all times, which can be expressed by taking a time derivative of Eq. (7):

$$\sum_{j=1}^{M} \frac{\partial \mu_i^{\mathrm{I}}}{\partial \phi_j^{\mathrm{I}}} \frac{\mathrm{d}\phi_j^{\mathrm{I}}}{\mathrm{d}t} = \sum_{j=1}^{M} \frac{\partial \mu_i^{\mathrm{II}}}{\partial \phi_j^{\mathrm{II}}} \frac{\mathrm{d}\phi_j^{\mathrm{II}}}{\mathrm{d}t} \,, \tag{G6a}$$

$$\sum_{j=1}^{M} \frac{\partial \Pi^{\mathrm{I}}}{\partial \phi_j^{\mathrm{I}}} \frac{\mathrm{d}\phi_j^{\mathrm{I}}}{\mathrm{d}t} = \sum_{j=1}^{M} \frac{\partial \Pi^{\mathrm{II}}}{\partial \phi_j^{\mathrm{II}}} \frac{\mathrm{d}\phi_j^{\mathrm{II}}}{\mathrm{d}t} \,, \tag{G6b}$$

provided that the initial phase volume and volume fractions $V^{\mathrm{I}}(t = 0)$, and $\phi_i^{\mathrm{I/II}}(t = 0)$ are a solution of Eq. (7). Once an expression for $\partial\mu_i/\partial\Phi_j$ and $\partial\Pi/\partial\Phi_j$ is calculated, we can derive an a set of $(M + 1)$ equations for $j_i^{\mathrm{I}}$ inserting Eq. (G3), Eq. (G4), and Eq. (G5) in Eq. (G6). These equations are linear and enable us to find an expression for $j_i^{\mathrm{I/II}}$ as a function of $\phi_i^{\mathrm{I/II}}$ and $V^{\mathrm{I}}/V$. We have finally all the ingredients to characterize the dynamics of the phase volume and volume fractions $\phi_i^{\mathrm{I/II}}(t)$ and $V^{\mathrm{I}}(t)$, integrating Eq. (G3) and Eq. (G4) and provided we can solve the initial phase equilibrium problem to find $V^{\mathrm{I}}(t = 0)/V$, and $\phi_i^{\mathrm{I/II}}(t = 0)$. This scheme can be used to study the kinetics of a system initially composed of two phases filled by monomers only that relax to its thermodynamic equilibrium. An illustration of such relaxation kinetics is depicted in **Fig. 10**. Note that the currents $j^{\mathrm{I/II}}$ restrict the trajectories to lie in the binodal manifold at all times.

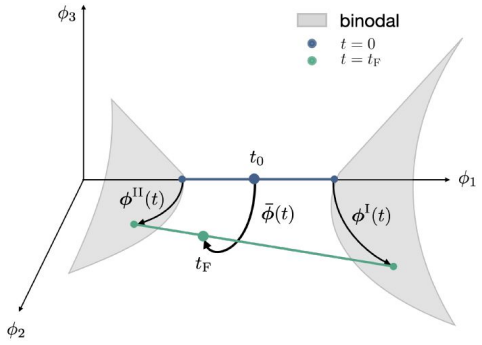



**Figure 10**

**Kinetic trajectory in the multicomponent phase diagram**

Illustration of the assembly kinetics at phase equilibrium, for systems corresponding to $M = 3$ and initially composed of monomers only.

## Supporting information

**Movie 1**

**Movie 2**

# References


[1] Doty P., Myers G. E. (1953) **Discussions of the Faraday Society 13**:51

[2] Hughes M. P., Sawaya M. R., Boyer D. R., Gold-schmidt L., Rodriguez J. A., Cascio D., Chong L., Gonen T., Eisenberg D. S. (2018) **Science 359**:698

[3] Fritsch A. W., Diaz-Delgadillo A. F., Adame-Arana O., Hoege C., Mittasch M., Kreysing M., Leaver M., Hyman A. A., Jülicher F., Weber C. A. (2021) **Proceedings of the National Academy of Sciences 118**:e2102772118

[4] Banani S. F., Lee H. O., Hyman A. A., Rosen M. K. (2017) **Nature Reviews Molecular Cell Biology 18**:285

[5] Nüske E., Marini G., Richter D., Leng W., Bogdanova A., Franzmann T. M., Pigino G., Alberti S. (2020) **Biology Open 9**:bio046391

[6] Zbinden A., Pèrez-Berlanga M., De Rossi P., Polymenidou M. (2020) **Developmental Cell 55**:45

[7] Wang B., Zhang L., Dai T., Qin Z., Lu H., Zhang L., Zhou F. (2021) **Signal Transduction and Targeted Therapy 6**:290

[8] Alberti S., Gladfelter A., Mittag T. (2019) **Cell 176**:419

[9] Zhao Y. G., Zhang H. (2020) **Developmental Cell 55**:30

[10] Agudo-Canalejo J., Schultz S. W., Chino H., Migliano S. M., Saito C., Koyama-Honda I., Stenmark H., Brech A., May A. I., Mizushima N., Knorr R. L. (2021) **Nature 591**:142

[11] Babl L., Merino-Salomón A., Kanwa N., Schwille P. (2022) **Scientific Reports 12**:17949

[12] Lu T., Liese S., Schoenmakers L., Weber C. A., Suzuki H., Huck W. T. S., Spruijt E. (2022) **Journal of the American Chemical Society 144**:13451

[13] Weber C. A., Zwicker D., Jülicher F., Lee C. F. (2019) **Reports on Progress in Physics 82**:064601

[14] Stockmayer W. H. (1943) **The Journal of Chemical Physics 11**:45

[15] Krapivsky P. L., Redner S., Ben-Naim E. (2010) **A Kinetic View of Statistical Physics** Cambridge: Cambridge University Press

[16] Israelachvili J. N. (2015) **Intermolecular and Surface Forces** Academic Press

[17] Safran S. A. (2019) **Statistical Thermodynamics of Surfaces, Interfaces, and Membranes** Boca Raton: CRC Press

[18] Hagan M. F., Grason G. M. (2021) **Reviews of Modern Physics 93**:025008

[19] Dumetz A. C., Chockla A. M., Kaler E. W., Lenhoff A. M. (2008) **Biophysical Journal 94**:570

[20] Shin Y., Brangwynne C. P. (2017) **Science 357**:eaaf4382

[21] McCall P. M., Srivastava S., Perry S. L., Kovar D. R., Gardel M. L., Tirrell M. V. (2018) **Biophysical Journal 114**:1636

[22] Wiegand T., Hyman A. A. (2020) **Emerging Topics in Life Sciences 4**:247

[23] Liu Z., Yang Y., Gu A., Xu J., Mao Y., Lu H., Hu W., Lei Q.-Y., Li Z., Zhang M., Cai Y., Wen W. (2020) **Nature Communications 11**:2266

[24] Küffner A. M., Linsenmeier M., Grigolato F., Prodan M., Zuccarini R., Palmiero U. C., Faltova L., Arosio P. (2021) **Chemical Science 12**:4373

[25] Beutel O., Maraspini R., Pombo-García K., Martin-Lemaitre C., Honigmann A. (2019) **Cell 179**:923

[26] Kar M., Dar F., Welsh T. J., Vogel L. T., Kühnemuth R., Majumdar A., Krainer G., Franzmann T. M., Alberti S., Seidel C. A. M., Knowles T. P. J., Hyman A. A., Pappu R. V. (2022) **Proceedings of the National Academy of Sciences 119**:e2202222119

[27] Li P., Banjade S., Cheng H.-C., Kim S., Chen B., Guo L., Llaguno M., Hollingsworth J. V., King D. S., Banani S. F., Russo P. S., Jiang Q.-X., Nixon B. T., Rosen M. K. (2012) **Nature 483**:336

[28] Franzmann T. M., Jahnel M., Pozniakovsky A., Mahamid J., Holehouse A. S., Nüske E., Richter D., Baumeister W., Grill S. W., Pappu R. V., Hyman A. A., Alberti S. (2018) **Science (New York, N.Y 359**:eaao5654

[29] Guillèn-Boixet J., Kopach A., Holehouse A. S., Wittmann S., Jahnel M., Schlüßler R., Kim K., Trussina I. R., Wang J., Mateju D., Poser I., Maharana S., Ruer-Gruß M., Richter D., Zhang X., Chang Y.-T., Guck J., Honigmann A., Mahamid J., Hyman A. A., Pappu R. V., Alberti S., Franzmann T. M. (2020) **Cell 181**:346

[30] Harmon T. S., Holehouse A. S., Rosen M. K., Pappu R. V. (2017) **elife 6**:e30294

[31] Martin E. W., Holehouse A. S., Peran I., Farag M., Incicco J. J., Bremer A., Grace C. R., Soranno A., Pappu R. V., Mittag T. (2020) **Science 367**:694

[32] Rana U., Brangwynne C. P., Panagiotopoulos A. Z. (2021) **The Journal of Chemical Physics 155**:125101

[33] Pappu R. V., Cohen S. R., Dar F., Farag M., Kar M. (2023) **Chemical Reviews** :acs.chemrev.2c00814

[34] Bray A. J., Rutenberg A. D. (1994) **Physical Review E 49**:R27

[35] Smoluchowski M. (1917) **Z. Phys. Chem 92**:129

[36] Oosawa F., Asakura S. (1975) **Thermodynamics of the Polymerization of Protein** London ; New York: Academic Press

[37] Blankschtein D., Thurston G. M., Benedek G. B. (1985) **Physical Review Letters 54**:955

[38] Li T., Rogers W. B., Jacobs W. M. (2023) **Physical Review E 108**:064501

[39] Semenov A. N., Rubinstein M. (1998) **Macromolecules 31**:1373

[40] Zilman A., Kieffer J., Molino F., Porte G., Safran S. A. (2003) **Physical Review Letters 91**:015901

[41] Tanaka F. (2011) **Polymer Physics: Applications to Molecular Association and Thermoreversible Gelation** Cambridge University Press

[42] Deviri D., Safran S. A. (2020) **Soft Matter 16**:5458

[43] Weber C. A., Michaels T., Mahadevan L. (2019) **eLife 8**:e42315

[44] Michaels T. C. T., Weber C. A., Mahadevan L. (2019) **Proceedings of the National Academy of Sciences 116**:14593

[45] Pönisch W., Michaels T. C. T., Weber C. A. (2023) **Biophysical Journal 122**:197

[46] Hagan M. F., Mohajerani F. (2023) **PLOS Computational Biology 19**:e1010652

[47] Kanaan N. M., Hamel C., Grabinski T., Combs B. (2020) **Nature Communications 11**:2809

[48] Ray S., Singh N., Kumar R., Patel K., Pandey S., Datta D., Mahato J., Panigrahi R., Navalkar A., Mehra S., Gadhe L., Chatterjee D., Sawner A. S., Maiti S., Bhatia S., Gerez J. A., Chowdhury A., Kumar A., Padinhateeri R., Riek R., Krishnamoorthy G., Maji S. K. (2020) **Nature Chemistry 12**:705

[49] Seim I., Posey A. E., Snead W. T., Stormo B. M., Klotsa D., Pappu R. V., Gladfelter A. S. (2022) **Proceedings of the National Academy of Sciences 119**:e2120799119

[50] Flory P. J. (1942) **The Journal of Chemical Physics 10**:51

[51] Huggins M. L. (1942) **Annals of the New York Academy of Sciences 43**:1

[52] Adame-Arana O., Weber C. A., Zaburdaev V., Prost J., Jülicher F. (2020) **Biophysical Journal 119**:1590

[53] Bartolucci G., Adame-Arana O., Zhao X., Weber C. A. (2021) **Biophysical Journal 120**:4682

[54] Bracha D., Walls M. T., Wei M.-T., Zhu L., Kurian M., Avalos J. L., Toettcher J. E., Brangwynne C. P. (2018) **Cell 175**:1467

[55] Mitrea D. M., Grace C. R., Buljan M., Yun M.-K., Pytel N. J., Satumba J., Nourse A., Park C.-G., Madan Babu M., White S. W., Kriwacki R. W. (2014) **Proceedings of the National Academy of Sciences 111**:4466

[56] Feric M., Vaidya N., Harmon T. S., Mitrea D. M., Zhu L., Richardson T. M., Kriwacki R. W., Pappu R. V., Brangwynne C. P. (2016) **Cell 165**:1686

[57] Luo F., Gui X., Zhou H., Gu J., Li Y., Liu X., Zhao M., Li D., Li X., Liu C. (2018) **Nature Structural & Molecular Biology 25**:341

[58] Hegde O., Li T., Sharma A., Borja M., Jacobs W. M., Rogers W. B. (2023) **Competition between self-assembly and phase separation governs high-temperature condensation of a DNA liquid,** *arXiv*

[59] Connor J. P., Quinn S. D., Schaefer C. (2022) **Frontiers in Molecular Neuroscience 15**:962526

[60] Jawerth L., Fischer-Friedrich E., Saha S., Wang J., Franzmann T., Zhang X., Sachweh J., Ruer M., Ijavi M., Saha S., Mahamid J., Hyman A. A., Jülicher F. (2020) **Science 370**:1317

[61] Bauermann J., Laha S., McCall P. M., Jülicher F., Weber C. A. (2022) **Journal of the American Chemical Society 144**:19294

[62] Miangolarra A. M., Duperray-Susini A., Coppey M., Castellana M. (2021) **Biophysical Journal 120**:2394

[63] Milo R., Phillips R. (2015) **Cell Biology by the Numbers** New York: Garland Science

[64] Bar-Even A., Noor E., Flamholz A., Buescher J. M., Milo R. (2011) **PLOS Computational Biology 7**:e1002166

[65] Mysels K. J. (1956) **Journal of Chemical Education 33**:178

[66] Patel A., Lee H. O., Jawerth L., Maharana S., Jahnel M., Hein M. Y., Stoynov S., Mahamid J., Saha S., Franzmann T. M., Pozniakovski A., Poser I., Maghelli N., Royer L. A., Weigert M., Myers E. W., Grill S., Drechsel D., Hyman A. A., Alberti S. (2015) **Cell 162**:1066

[67] Kato M., Han T. W., Xie S., Shi K., Du X., Wu L. C., Mirzaei H., Goldsmith E. J., Longgood J., Pei J., Grishin N. V., Frantz D. E., Schneider J. W., Chen S., Li L., Sawaya M. R., Eisenberg D., Tycko R., McK-night S. L. (2012) **Cell 149**:753

[68] Hartl F. U., Bracher A., Hayer-Hartl M. (2011) **Nature 475**:324

[69] Patel A., Malinovska L., Saha S., Wang J., Alberti S., Krishnan Y., Hyman A. A. (2017) **Science 356**:753

[70] Hautke A., Ebbinghaus S. (2023) **Biological Chemistry 404**:897

[71] Nandi S. K., Österle D., Heidenreich M., Levy E. D., Safran S. A. (2022) **Physical Review Letters 129**:128102

[72] Michaels T. C., Šarić A., Habchi J., Chia S., Meisl G., Vendruscolo M., Dobson C. M., Knowles T. P. (2018) **Annual Review of Physical Chemistry 69**:273

## Author information

### Giacomo Bartolucci

Max Planck Institute for the Physics of Complex Systems, Dresden, Germany, Faculty of Mathematics, Natural Sciences, and Materials Engineering: Institute of Physics, University of Augsburg, Universitätsstraße 1, Augsburg, Germany
ORCID iD: 0000-0002-3128-1154

**For correspondence:** bartolucci@pks.mpg.de

**Ivar S Haugerud**
Faculty of Mathematics, Natural Sciences, and Materials Engineering: Institute of Physics, University of Augsburg, Universitätsstraße 1, Augsburg, Germany
ORCID iD: 0000-0001-9931-3741

**Thomas CT Michaels**
ETH Zurich, Institute of Biochemistry, Zurich, Switzerland
ORCID iD: 0000-0001-6931-5041

**Christoph A Weber**
Faculty of Mathematics, Natural Sciences, and Materials Engineering: Institute of Physics, University of Augsburg, Universitätsstraße 1, Augsburg, Germany
ORCID iD: 0000-0001-6279-0405

**For correspondence:** christoph.weber@physik.uni-augsburg.de

## Editors

Reviewing Editor
**Pierre Sens**
Institut Curie, CNRS UMR168, Paris, France

Senior Editor
**Aleksandra Walczak**
CNRS, Paris, France

**Reviewer #3 (Public review):**

Summary:

The authors combine classical theories of phase separation and self-assembly to establish a framework for explaining the coupling between the two phenomena in the context of protein assemblies and condensates. By starting from a mean-field free energy for monomers and assemblies immersed in solvent and imposing conditions of equilibrium, the authors derive phase diagrams indicating how assemblies partition into different condensed phases as temperature and the total volume fraction of proteins are varied. They find that phase separation can promote assembly within the protein-rich phase, providing a potential mechanism for spatial control of assembly. They extend their theory to account for the possibility of gelation. They also create a theory for the kinetics of self-assembly within phase separated systems, predicting how assembly size distributions change with time within the different phases as well as how the volumes of the different phases change with time.

Review For Revision:

The revised manuscript provides better motivation and physical explanations for the equations, and the authors have addressed references, typos, and other minor technical issues identified in the review. These changes have significantly improved the manuscript.

https://doi.org/10.7554/eLife.93003.2.sa1

**Author response:**

The following is the authors' response to the original reviews.

***Public Reviews:***

***Reviewer 1 (Public Review):***

*Summary:*

*The authors present a mean-field model that describes the interplay between (protein) aggregation and phase separation. Different classes of interaction complexity and aggregate dimensionality are considered, both in calculations concerning (equilibrium) phase behavior and kinetics of assembly formation.*

*Strengths:*

*The present work is, although purely theoretical, of high interest to understanding biological processes that occur as a result of a coupling between protein aggregation and phase separation. Of course, such processes are abundant, in the living cell as well as in in-vitro experiments. I appreciate the consideration of aggregates with various dimensionality, as well as the categorization into different "interaction classes", together with the mentioning of experimental observations from biology. The model is convincing and underlines the complexity associated with the distribution of proteins across phases and aggregates in the living cell.*

*Weaknesses:*

*There are a few minor weaknesses.*

***Reviewer 2 (Public Review):***

*This work deals with a very difficult physical problem: relating the assembly of building blocks on a molecular scale to the appearance of large, macroscopic assemblies. This problem is particularly difficult to treat, because of the large number of units involved, and of the complex way in which these units-monomers-interact with each other and with the solvent. In order to make the problem treatable, the authors recur to a number of approximations: Among these, there is the assumption that the system is spatially homogeneous, i.e., its features are the same in all regions of space. In particular, the homogeneity assumption may not hold in biologically relevant systems such as cells, where the behavior close to the cell membrane may strongly differ from the one in the bulk. As a result, this hypothesis calls for a cautious consideration and interpretation of the results of this work. Another notable simplification introduced by the authors is the assumption that the system can only follow two possible behaviors: In the first, each monomer interacts equally with the solvent; no matter the size of the cluster of which it is part. In the second case, monomers in the bulk of a cluster and monomers at the assembly boundary interact with the solvent in a different way. These two cases are considered not only because they simplify the problem, but also because they are inspired by biologically relevant proteins.*

*With these simplifications, the authors trace the phase diagram of the system, characterizing its phases for different fractions of the volume occupied by the monomers and solvent, and for different values of the temperature. The results qualitatively reproduce some features observed in recent experiments, such as an anomalous distribution of cluster sizes below the system saturation threshold, and the gelation of condensed phases above such threshold.*

***Reviewer 3 (Public Review):***

*Summary:*

> *The authors combine classical theories of phase separation and self-assembly to establish a framework for explaining the coupling between the two phenomena in the context of protein assemblies and condensates. By starting from a mean-field free energy for monomers and assemblies immersed in solvent and imposing conditions of equilibrium, the authors derive phase diagrams indicating how assemblies partition into different condensed phases as temperature and the total volume fraction of proteins are varied. They find that phase separation can promote assembly within the protein-rich phase, providing a potential mechanism for spatial control of assembly. They extend their theory to account for the possibility of gelation. They also create a theory for the kinetics of self-assembly within phase separated systems, predicting how assembly size distributions change with time within the different phases as well as how the volumes of the different phases change with time.*
>
> *Strengths:*
>
> *The theoretical framework that the authors present is an interesting marriage of classic theories of phase separation and self-assembly. Its simplicity should make it a powerful general tool for understanding the thermodynamics of assembly coupled to phase separation, and it should provide a useful framework for analyzing experiments on assembly within biomolecular condensates.*
>
> *The key advance over previous work is that the authors now account for how self-assembly can change the boundaries of the phase diagram.*
>
> *A second interesting point is the explicit theoretical consideration for the possibility that gelation (i.e. self-assembly into a macroscopic aggregate) could account for widely observed solidification of condensates. While this concept has been broadly discussed, to date I have yet to see a rigorous theoretical analysis of the possibility.*
>
> *The kinetic theory in sections 5 and 6 is also interesting as it extends on previous work by considering the kinetics of phase separation as well as those of self-assembly.*
>
> *Weaknesses:*
>
> *A key point the authors make about their theory is that it allows, as opposed to previous research, to study non-dilute limits. It is true that they consider gelation when the 3D assemblies become macroscopic. However, dilute solution theory assumptions seem to be embedded in many aspects of their theory, and it is not always clear where else the non-dilute limits are considered. Is it in the inter-species interaction* $\chi ij$*? Why then do they never explore cases for which* $\chi ij$ *is nonzero in their analysis?*

We explicitly consider that monomers and aggregates are non-dilute with respect to solvent. This is evident in accounting for the mixing entropy of all components, including the solvent. Moreover, we account for interactions among the monomers and the different aggregates with the solvent. We consider the case where each monomeric unit, independent in aggregate it is part of, interacts the same way with the solvent. Please note that this case corresponds to a non-dilute scenario where interactions indeed drive phase separation.

> *The connection between this theory and biological systems is described in the introduction but lost along the main text. It would be very helpful to point out, for instance, that the presence of phase separation might induce aggregation of proteins. This point is described formally at the end of Section 3, but a more qualitative connection to biological systems would be very useful here.*

We thank the referee for the useful comment, we now mention this in the introduction (line 80) and point out the biological relevance of assembly formation and localization via the

presence of phase separation (lines 268 and 283).

> *Building on the previous point, it would be helpful to give an intuitive sense of where the equations derived in the Appendices and presented in the main text come from and to spell out clear physical interpretations of the results. For example, it would be helpful to point out that Eq. 4 is a form of the law of mass action, familiar from introductory chemistry. It would be useful to better explain how the current work extends on existing previous work from these authors as well as others. Along these lines, closely related work by W. Jacobs and B. Rogers [O. Hedge et al. 2023, https://arxiv.org/abs/2301.06134; T. Li et al. 2023, https://arxiv.org/abs/2306.13198] should be cited in the introduction. The results discussed in the first paragraph of Section 3 on assembly size distributions in a homogeneous system are well-known from classic theories of self-assembly. This should be acknowledged and appropriate references should be added; see for instance, Rev. Mod. Phys. 93, 025008 and Statistical Thermodynamics Of Surfaces, Interfaces, And Membranes by Sam Safran. Equation 14 for the kinetic of volume fractions is given with reference to Bauermann et al. 2022, but it should be accompanied by a better intuitive interpretation of its terms in the main text. In particular, how should one understand the third term in this equation? Why does the change in volume impact the change of volume fraction in this way?*

We thank the referee for the suggestions. We have included the missing references, with a particular emphasis on DNA nanostars that inhibit phase separation in DNA liquids in the definition of class II. We added intuitive explanations of the main equations, such as Eqs. (4), (8),(14), (17), and (18). Notice that, according to Mysels, Karol J., J. Chem. Educ., 33, 178 (1956) (https://pubs-acs-org.sire.ub.edu/doi/epdf/10.1021/ed033p178) we refer to (18) as the law of mass action.

> *The discussion in the last paragraph of Section 6 should be clarified. How can the total amount of protein in both phases decrease? This would necessarily violate either mass or volume conservation. Also, the discussion of why the volume is non-monotonic in time is not clear.*

A decrease in the total amount of protein in both phases does not violate mass conservation, if the volume of the phases varies accordingly. In particular, the volume of the denser phase should grow. This given, in the case presented the total protein amount in the dense phase decreases, while in the dilute phase increases. For this reason, we revised the paragraph and now explain the results in more detail (see lines starting from 407). The nonmonotonic volume change is indeed a puzzling finding that, as we now state in the manuscript, requires further investigation. Given the lack of analytical approaches available to tackle the complex kinetics in the presence of coexisting phases, we believe that this analysis goes beyond the scope of the present paper.

> ***Recommendations for the authors***
>
> ***Reviewer 1 (Recommendations For The Authors):***
>
> *Line 96: I feel a mentioning/definition/explanation and perhaps some discussion on the parameter M (limiting aggregate size) would have been in place in the introduction of Equation (1). Furthermore, in the usual interpretation, Flory interaction parameters (symbolized χ) are dimensionless, as, classically, they represent an exchange energy (normalized by kT), defined on a monomeric basis. Here they seem to carry the dimension of energy.*

We thank the reviewer for the observation. We have included a brief comment on M and mentioned that we use $\chi$ parameters that carry the dimension of energy such that, varying

*kBT*, we scale at the same time the term containing interaction propensities ($\chi$) and the one containing internal energies (_e_int). See the comment on line 127

> *Line 150: The choice of* ρi = i *physically implies that a single protein is assumed to have the same as a solvent molecule. This may be a bit of a stretch. This assumption leads to an overestimation of the translational entropy of the aggregates (first term in Equation (1)). Acknowledging that _ρ_1 >>* ρs *would give a pronounced desymmetrization of the phase diagram (I suspect).*

Indeed, in the case of monomers only, the assumption leads to a symmetric phase diagram which may be unrealistic. Once assemblies form, however, the phase diagram becomes asymmetric and for this reason we decided to assume $\rho i = i$, simplifying the theoretical analysis. We have added a clarifying sentence in the manuscript, see line 163

> *Furthermore, the pictures in Figure 1a-c suggest the presence of a disordered residue, the degree of swelling of which might affect binding strength (see for instance: https://doi.org/10.3389/fnmol.2022.962526).*

We added a comment on the possible coupling between internal free energies and interaction propensities, such as the swelling mechanism that affects binding sites, and included the reference above (line 215).

> *Line 154-156: It's unclear what is meant with "an internal bond that keeps each assembly together". How should this be interpreted on an intuitive physical level?*

We apologise for being unclear. We meant the internal bonds that lead to the formation of assemblies. We have now rephrased this sentence in the main text (lines starting from 169).

> *Line 254: The fact that* Φsg *is defined below does not mean it does not fall out of the air here. The same holds for the consideration of the limit* M $\rightarrow \infty$*. Ideally, the main text should stand on its own, in particular with respect to physical intuitiveness, as well as the necessity and interest of discussion topics. Technical details, derivations and additional information can be in an appendix.*

We agree with the referee and added some physical insights about the limit. We now also state clearly in the main text (line 298) that _Φ_sg is affected by temperature and the free energy of internal bonds.

> *Line 257: "Since we do not explicitly include the solvent in assembly formation we will consider the gel as a phase without solvent and thus* Φtot *= 1". I'm not sure if I can agree with this. I would say, a gel, certainly in biological context, almost per definition contains a large fraction of solvent, i.e. here water. The situation "*Φtot *= 1" would rather be a solid precipitate. Is gelation properly captured by this model?*

We thank the referee for this very relevant observation. We now state in the main text that the model predicts a macroscopic assembly which we call 'the gel phase', in agreement with previous literature. Then, to clarify, we added the sentence "Please note that, since we do not explicitly include the solvent in assembly formation (see reaction scheme in Fig.1a), in our model the gel corresponds to a phase without solvent, _Φ_tot = 1. To account for biological gels that can be rich in water, our theory can be straightforwardly extended by incorporating the solvent into the reaction scheme.", see main text line 300.

> *Line 268: Shouldn't "solvent" be "solution"? If* fsol *is given by Equation (1), surely not only the solvent is considered.*

Indeed, this is a typo, and we now use the term 'solution' instead of 'solvent.'

> *Line 273: At this stage, the only information provided in the main text is that ω∞ is "a constant that does not affect chemical nor phase equilibrium, except in the limit* M → ∞*" (see lines 153-154). This is a little bit too abstract for me. Again, the main text should stand on its own, meaning the reader should not have to rely on an appendix to at least have an intuitive physical understanding of any modeling or input parameter discussed in the main text.*

We thank the reviewer for pointing this out. We now comment on the physical interpretation of $\omega_\infty$ in the main text, see lines from 320 on.

> *Figure 4.* $\tilde{k}$ *appears in Equation (39) but it is not defined.*

We thank the reviewer for pointing this out. We have reshaped appendix 6A, making use of chemical activities and clarified the origin of the rate $\tilde{k}$.

> *Line 317. I don't fully understand the intention of the remark on the model being adaptable for "primary and secondary nucleation". How/in what way is this different from association and dissociation? For instance, classical nucleation theory is based on association and dissociation of monomeric units to and from clusters.*

We agree that the kinetic rate coefficients *kij* (appearing in the association and dissociation rates Δ*rij*, Eq. 17) in our manuscript already depend on assembly length, see Appendix 6 B, where we now clarified their definition. Please note that, however, that secondary nucleation is a special kind of association, for which the kinetic rate coefficients corresponding to associations of small assemblies, i.e. *kij* with_i,j_ ≪ *M*, explicitly depend on the presence of large assemblies with sizes $l \gg 1$. In our manuscript, we have not accounted for such a dependence. We now make this aspect clear in the manuscript, see Appendix 6 B.

> *Line 321. Why is Δ*rij *called the "monomer exchange rate"? In line 318 the same parameter is defined as the "reaction rate for the formation of a (i+j)-mer". Why should these be the same?*

We thank the reviewer for spotting this typo.

> *Line 323. Why do these calculations use M = 15?*

The exploration of a 15-dimensional phase space is already numerically challenging. We are currently working on a generalization of the numerical scheme to work with larger values of M but, to discuss the fundamental physical principles, we kept $M = 15$.

> ***Reviewer 2 (Recommendations For The Authors):***
>
> *The manuscript presents several issues, on both the scientific and presentational level, which need to be carefully addressed. Please find below a list of the points that need to be addressed by the authors, divided into major and minor points. Major issues:*
>
> *• A general, major concern about the results in the paper is the homogeneity assumption. I do understand that repeating the whole analysis presented in the manuscript by allowing for spatial inhomogeneities partially goes beyond the scope of this paper.*

> *However, the authors should at least discuss how such inhomogeneities may alter the results in a qualitative way, and treat explicitly the presence of inhomogeneity in one prototypical case treated in the manuscript. Namely, what happens if the volume fractions and relative molecular volumes in the free energy (1) depend on space, e.g.,* $\Phi i \rightarrow \Phi i(x)$*?*

We would like to stress that, in the present paper, we do account for spatial inhomogeneities. Indeed, in the case of phase separation, we consider systems which are divided into two phases, characterized by different values of the assemblies' volume fractions $\Phi i$. We do, however, consider the system to be homogeneous inside the phases, implying a jump in the value of the volume fraction at the interface between the two phases. In this sense, the analysis we carry out is valid in the thermodynamic limit, where gradients of the volume fractions $\Phi i(x)$ within the phases, can be neglected. On the other hand, considering the full spatial problem, i.e. solving the equations for $M = 15$ spatially varying fields, would be numerically extremely challenging.

> *• The authors' results relate molecular assembly- a phenomenon at the molecular scale- to phase separation-a mesoscopic or macroscopic phenomenon. The authors should stress the conceptual importance of this connection between scales, and present their results from the perspective of a multi-scale model.*

We thank the reviewer for pointing this out. We now emphasize the multi-scale feature of our model in the introduction (line 80).

> *• Starting from Section 1, the reader is not well guided through the sections that follow. The authors should provide an outline of the line of though that they are going to follow in the following sections, and logically connect each section to the next one with a short paragraph at the end of each section. This paragraph should resume what has been addressed in the current section, and the connection with the topic that will be addressed in the next one.*

We agree with the reviewer and have added a transitioning sentence at the end of each paragraph.

> *• 'We focus on linear assemblies (d = 1)': Given the striking differences of the results between d = 1 and* d > *1 shown above, the authors should discuss what happens for* d > *1 as well.*
>
> *• 'In figure Fig. 5a, we show the initial and final equilibrium binodals (black and coloured curve, respectively), for the case of linear assemblies (d = 1) belonging to class 1': Again, show what happens for* d > *1.*

We agree with the reviewer, the kinetics in $d > 1$ would be definitely interesting. However, in this case, one assembly can become macroscopic (i.e. $M$ must be set to $\infty$). This requires some substantial modification in the kinetic scheme, like introducing an absorbing boundary condition for monomers 'sucked in' the gel. We prefer to leave this for future work, and now state it explicitly in the manuscript (line 383).

> *• 'This difference arises because, within class 2, monomers in the bulk of an assembly have reduced interaction propensity with respect to the boundary ones. As a consequence, the formation of large clusters shifts the onset of phase separation to higher* $\Phi$tot *values.': To prove this argument, the authors should show Fig. 2g and h for* d > *1. In fact, by varying d, the effect of the boundary vs. bulk also varies.*

We prefer to discuss the thermodynamics of $d > 1$ in section 4 on gelation. There we present only a single phase diagram so as not to blow up the discussion on equilibrium too much.

> • *'referring for simplicity to systems belonging to Class 1': The authors should do the same analysis for Class 2.*

We agree with the reviewer. However, again not to blow up the discussion on equilibrium, we leave it for future work.

> • *'other, implying that the corresponding Flory-Huggins parameter* χij *vanishes': Why?*

The explanation based on a lattice model is reported in Appendix 2, and is now more clearly referenced (line 185).

> *Minor issues:*
>
> • *Eq. (10): Here the authors should explain in the main text, possibly in a simple and intuitive way, why the number of monomers i and the space dimension d enter the righthand side of this equation in this particular way.*

We thank the reviewer for pointing this out. We added the physical origin of the scaling with dimension in Eq. (10) and in Eq. (8), as pointed out by reviewer 3.

> • *'The second and fifth terms of* fsol *characterize the internal free energies': What do you mean by 'characterize the internal free energies'? Please clarify.*

As we now state more clearly (lines 114-120), these two contributions include the internal free energies _ω_s and $\omega i$, stemming from the free energy of internal bonds that lead to assembly formation.

> • *'depend on the scaling form of the': Scaling with respect to what ? Please clarify.*

We have now clarified that the scaling is with respect to the assembly size $i$.

> • *Figure 2 is way too dense: it should be split into two figures, and the legend of each of the two figures should be expanded to properly guide the reader to understand the figures.*

We understand the reviewer's point of view. To avoid altering the present flow, we decided not to split the figure, but we have included shaded boxes to better guide the reader.

> • *'this is a consequence of the gelation transition': Please clarify*
>
> • *'and this limitation can be dealt with by introducing explicitly the infinite-sized gel in the free energy': Why? Please clarify.*

We have now rephrased these sentences, hopefully in a clearer way. We now state: 'We know that this divergence is physical, and is caused by the gelation transition. This limitation can be dealt with by introducing explicitly a term in the free energy that accounts for an infinite-sized assembly (the gel)', see lines 320-322.

> • *Figure 4: Add plots of panels d, e, h and i with log scale on the y axis to make explicit an eventual exponential behavior, and revise the text accordingly*

Not to further complicate Figure 4, we preferred to display the logarithmic plots of the equilibrium distribution in the appendix, see Figure A3-1.

> *• '... an equilibrium distribution which monotonously decreases with assembly size': It is not the distributions that decreases but the cluster volume fraction, please rephrase.*

We thank the reviewer for pointing this out and have now rephrased this sentence (line 394).

> ***Reviewer 3 (Recommendations For The Authors):***
>
> *I could not obtain the exact form of Eq 29 in App 3, can the authors elaborate on this calculation. App 3: What does it mean binodal agrees well with Φsg? And doesn't Φsg depend on temperature through phi tilde? What temperature is this result for?*

We apologise for the unclear explanation. We now state in detail that Eq. (29) is obtained by plugging the expression of $\Phi i$ given in Eq. (24) into Eq. (1), in the main text. The dependence of $\Phi_1$ on $\Phi_{tot}$ is expressed in Eq. (26), and we have omitted linear terms in $\Phi_{tot}$, since they do not affect phase equilibrium (see lines 802-809). Moreover, $\Phi$sg depends indeed on $k_BT$. We refer to the comparison between the full curve $\Phi$sg in the $k_BT$–$\Phi_{tot}$ plane, and the branch of the binodal between the triple point (indicated now with a cross) and $\Phi_{tot} = 1$. The two curves are close, as expected since both correspond to the boundary between homogeneous mixtures and the gel state, obtained with different methods.

> *The references to Figures in the appendices are confusing. Please make it clear whether Figures in the main text or the appendices are being referenced. On a related note, the Appendix figures seem to be placed in appendices whose text describes something else - Appendix 2, Figure 1 should be moved to Appendix 3; Appendix 3, Figure 1 should be moved to Appendix 4; etc.*

We revised the appendix, corrected the figure positions and clarified their references.

https://doi.org/10.7554/eLife.93003.2.sa0