Biomolecular phase boundaries are described by a solubility product that accounts for variable stoichiometry and soluble oligomers

Sk Ashif Akram,[†] Aniruddha Chattaraj,[‡] Terry Salava,[†] Jonathon Ditlev,[¶] Leslie M. Loew,[‡] and Jeremy D. Schmit[*,†]

†*Department of Physics, Kansas State University, Manhattan, KS 66506, USA*

‡*R. D. Berlin Center for Cell Analysis and Modeling, University of Connecticut School of Medicine, Farmington, CT 06030*

¶*Program in Molecular Medicine, and Program in Cell and Systems Biology, Hospital for Sick Children, Toronto, ON M5G 1E8, Canada*

§*Department of Biochemistry, University of Toronto, Toronto, ON M5S 1A1, Canada*

E-mail: schmit@ksu.edu

Abstract

The solubility product is a rigorous description of the phase boundary for salt precipitation and has also been used to qualitatively describe the condensation of biomolecules. Here we present a derivation of the solubility product showing that the solubility product is also a robust description of biomolecules phase boundaries if care is taken to account for soluble oligomers and variable composition within the dense phase. Our calculation describes equilibrium between unbound monomers, the dense phase, and an ensemble of oligomer complexes with significant finite-size contributions to their free energy. The biomolecule phase boundary very nearly resembles the power law predicted by the solubility product when plotted as a function of the monomer concentrations. However, this simple form is confounded by the presence of oligomers in the dilute phase. Accounting for the oligomer ensemble introduces complexities to the power law phase boundary including re-entrant behavior and large shifts for stoichiometrically matched molecules. We show that allowing variable stoichiometry in the dense phase expands the two phase region, which appears as curvature of the phase boundary on a double-logarithmic plot. Furthermore, this curvature can be used to predict variations in the dense phase composition at different points along the phase boundary.

I. Introduction

Biomolecular phase separation has emerged as an important mechanism of cellular organization.[1,2] In some cases biomolecular condensate formation is driven by the phase separation of a single dominant scaffold molecule that occurs at a threshold concentration C_{sat}.[3–6] However, in many other cases the formation of biomolecular condensates requires multiple scaffold molecules and the condensation threshold depends on the concentration of all of the scaffolds.[7–16] A common case is when two scaffold species are required. This situation is analogous to salt precipitation, which requires both anions and cation components. In contrast to

the C_{sat} describing single-component phase separation, salts remain soluble when the product of the ion concentrations remains below a threshold value K_{SP}, known as the solubility product constant. Application of the solubility product concept to biological condensates has been promising,[16–19] however, biomolecule systems have complexities not present in simple salts that have confounded interpretation of experiments. For example, phase separating biomolecules often form small oligomers that remain in the dilute phase.[19,20] Additionally, it can be difficult to determine the effective valence of disordered proteins. Even in cases where the valence can be obtained from well-defined interaction modules, asymmetries due to non-interacting "spacers"[21] or electrostatic interactions can result in unexpected mixing ratios. Finally, biomolecular condensates tend to form without organized, repeating structural elements, which contributes to their liquid properties. This allows for heterogeneity in stabilizing interactions and variation in the stoichiometric composition of the condensed phase. This is biologically significant because stoichiometric variation has been linked to the functional state of condensates.[22–25]

Here we present a theory of the solubility product that accounts for many of the complexities inherent to biomolecular phase separation, specifically soluble oligomers and variable stoichiometry in the dense phase. We show that the phase boundary is most naturally defined in terms of the concentrations of free monomers. The presence of oligomers in the dilute phase complicates measurement of the monomer concentrations and leads to phenomena like the "magic number" effect[26–29] and re-entrant transitions.[30–32] Fluctuations in the dense phase composition expand the two-phase region of the phase diagram, which appears as curvature in the phase boundary when viewed on a double-logarithmic scale. This relationship can be inverted such that curvature in the phase boundary can be used to predict how composition of the dense phase changes with the solution composition. Finally, we show that the optimal stoichiometric composition of the condensed phase requires a balance of chemical potentials that does not coincide with balanced valences in the dilute phase (except in the special case of 1:1 stoichiometry). Together these results show how measurement of

the dilute phase can be informative about the properties of the condensed phase.

Results

Model describes the co-assembly of two molecular species

We present theory and simulations for a solution consisting of two macromolecular species, that we refer to as "A" and "B", immersed in a solvent that is implicitly accounted for by the interaction energies. The molecules can assemble into complexes that are each characterized by a free energy $F(n, m)$ where the argument n indicates the number of A molecules and m is the number of B molecules in the cluster. We use the term "cluster" generically to indicate a reversible, in the thermodynamic sense, supramolecular complex without implications of the solid/liquid or oligomer/phase properties of that complex.

The grand partition function of the solution is

$$Q = \sum_{\{n,m\}} e^{-F(n,m)+n\mu_A+m\mu_B} = \sum_{\{n,m\}} \left(\frac{C_a}{C_0}\right)^n \left(\frac{C_b}{C_0}\right)^m e^{-F(n,m)} \tag{1}$$

Where the second step takes the monomers to be the reference state (i.e., $F(1, 0) = F(0, 1) = 0$) and assumes the solution to be sufficiently dilute that the chemical potentials are $\mu_A = \ln(C_a/C_0)$ and $\mu_B = \ln(C_b/C_0)$, where C_0 is a reference concentration that we take to be 1 M in our comparison to simulations. The chemical potentials, free energies of association $F(n, m)$, and all subsequent energies, are expressed in units of k_BT, where k_B is Boltzmann's constant and T is the absolute temperature. Hereafter, we work in terms of dimensionless concentrations denoted with a lowercase c, as in $c_a = C_a/C_0$, $c_b = C_b/C_0$, etc. In our notation concentrations with a lowercase subscript indicate the monomer concentration while capital subscripts are reserved for the total concentration (with a superscript, if necessary, to indicate the dense or dilute phase).

Divergence in the partition function indicates the phase boundary

We divide the partition function into terms representing the monomers, oligomers, and condensed phase.

$$
\begin{aligned}
Q &= Q_{\text{mono}} + Q_{\text{oligo}} + Q_{\text{den}} && (2)\\
Q_{\text{mono}} &= c_a + c_b && (3)\\
Q_{\text{oligo}} &= \sum^{\text{oligomers}} c_a^n c_b^m e^{-F(n,m)} && (4)\\
Q_{\text{den}} &= \sum_{\{n \geq n_{\min}, m \geq m_{\min}\}} c_a^n c_b^m e^{-F(n,m)} && (5)
\end{aligned}
$$

We consider a cluster to be part of the dense phase when it exceeds threshold sizes $n_{\min}$ and $m_{\min}$ that are large enough that surface corrections to the free energy can be neglected. In other words, for clusters greater than these cutoffs the free energy scales linearly with size (i.e., $F(\lambda n, \lambda m) = \lambda F(n,m)$ for $\lambda \geq 1$, $n > n_{\min}$, and $m > m_{\min}$). All clusters that do not satisfy these conditions are included in Q_{oligo}.

We first consider the case where the stoichiometry of the dense phase is tightly constrained, which is valid in the tight binding limit where all bonds are satisfied in the dense phase. In this case, the number of B molecules in the dense phase scales in direct proportion to the number of A molecules, $m = sn$, where s is the stoichiometric ratio. The Boltzmann factors above the size cutoff can be written as $K_{n,m}^{-1} = \delta_{m,sn} e^{-F(n,m)}$, where δ is the Kronecker delta and $K_{n,m} = c_a^n c_b^m / c_{n,m}$ is the equilibrium dissociation constant for the (n,m) cluster. Applying this constraint to Q_{den}, the summation over m can be evaluated

$$
Q_{\text{den}} = \sum_{n \geq n_{\min}} c_a^n c_b^{sn} e^{-F(n,sn)} \qquad (6)
$$

Next, we use the fact that the free energy scales linearly with size to write $F(n, sn) \simeq n f_A$ where $f_A = \lim_{n\to\infty} F(n, ns)/n$ describes the free energy per A molecule in large clusters.

The summation is evaluated as

$$Q_{\text{den}} = \sum_{n=n_{\min}}^{\infty} \left(c_a c_b^s e^{-f_A}\right)^n = \frac{\left(c_a c_b^s e^{-f_A}\right)^{n_{\min}}}{1 - c_a c_b^s e^{-f_A}} \tag{7}$$

Note that Q_{den} diverges when $f_A - \ln c_a - s \ln c_b = 0$ which can be rearranged in the form of a solubility product

$$c_a c_b^s = e^{f_A} \tag{8}$$

The requirement that the partition function is finite and positive means that the concentration product $c_a c_b^s$ must be less than $K_{\text{SP}} = e^{f_A}$. This limit on the monomer concentrations is the phase boundary. For a predetermined value of c_a, the critical concentration of B molecules is $c_b = (K_{\text{SP}}/c_a)^{1/s}$ which is a straight line with slope $-s^{-1}$ on a double logarithmic plot. Note that omitting the small n, m terms from the Q_{den} sum does not affect condition for the phase boundary because $n_{\min}$ does not appear in Eq. 8.

Oligomers perturb the phase boundary by suppressing monomer concentrations

In many experiments the known quantities are the total concentrations and the partitioning between the dense and dilute phases. In such experiments the phase boundary is plotted as a function of the total concentrations c_A and c_B. Such phase boundaries do not show the expected power law behavior because the soluble fraction includes both monomers and oligomers.[33] Soluble oligomers can cause qualitative differences between the apparent phase boundary and the solubility product in both the subsaturated regime and super-stoichiometric mixtures.

To address these cases we note that, for suitably large $n_{\min}$, Q_{den} behaves like a step function because values of $c_a c_b^s / K_{\text{SP}} < 1$ are strongly suppressed by the exponent $n_{\min}$ in the numerator of Eq. 7 (see Fig. 1A). Therefore, below the solubility threshold it is reasonable

to approximate $Q \simeq Q_{\text{mono}} + Q_{\text{oligo}}$. Furthermore, in many cases Q_{oligo} will be dominated by a small number of favorable species.[18,34] Therefore, Q_{oligo} can be approximated by truncating the cluster summation to include only the most significant oligomers for a given system. This approximate form for the partition function is a polynomial that can be solved for c_a and c_b. This solution can be used to solve for the locus of points where $c_a c_b^s = K_{\text{SP}}$, which corresponds to the phase boundary.

To illustrate this approach, we consider a case $s = 1.4$ that, in the strong binding limit, corresponds to molecular valences that are related by $v_A = s v_B$. When the A molecule is present in excess, we expect the dominant oligomer will be the 1:1 heterodimer, which fully saturates the B binding sites. Under conditions of excess B the dominant oligomer will be the 1:2 heterotrimer, which fully saturates the A binding sites. Thus the partition function in the sub-saturated regime (i.e., no dense phase) can be approximated as

$$Q \simeq Q_{\text{mono}} + Q_{\text{oligo}} \tag{9}$$

$$= c_a + c_b + c_a c_b e^{-F(1,1)} + c_a c_b^2 e^{-F(1,2)} \tag{10}$$

The total concentration of B molecules is given by

$$c_B^{\text{dil}} = c_b \frac{dQ}{dc_b} = c_b + c_a c_b e^{-F(1,1)} + 2 c_a c_b^2 e^{-F(1,2)} \tag{11}$$

Similarly, the concentration of A monomers c_a is related to the total A concentration c_A^{dil} by

$$c_a = \frac{c_A^{\text{dil}}}{1 + c_b e^{-F(1,1)} + c_b^2 e^{-F(1,2)}} \tag{12}$$

where the denominator is the binding polynomial for A molecules.[35] Plugging Eq. 12 into Eq. 11 gives a cubic equation for c_b. The real, positive root of this equation gives c_b as a function of c_A^{dil} and c_B^{dil}, which can be used with Eq. 12 to give c_a.

Fig. 1B shows the solubility product $c_a c_B^s$ obtained from Eqs. 11 and 12 as a function

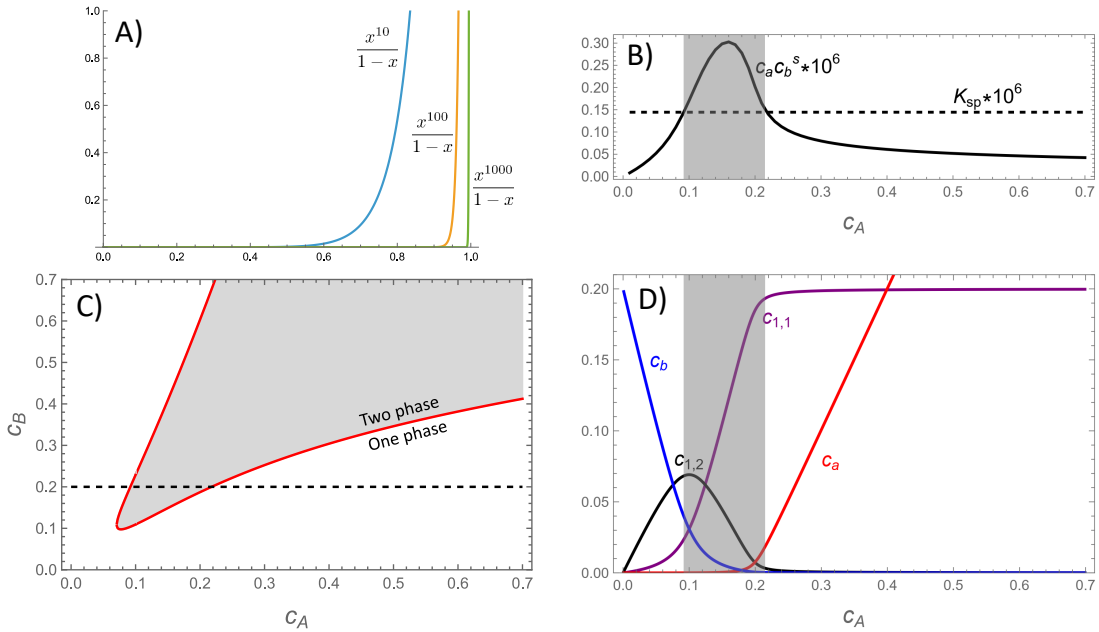

Figure 1: (A) The dense phase partition function Q_{den} has the characteristics of a step function, which allows the approximation $Q \simeq Q_{\text{mono}} + Q_{\text{oligo}}$ in the sub-saturated regime. Larger values of n_{min} increase the steepness of the step, while smaller values lead to more gradual phase transitions. (B) Plot of the monomer product $c_a c_b^s$ for fixed c_B (dotted line in panel C) computed from Eqs. 12 and 11. The product exceeds the threshold for condensation in the shaded region. (C) The phase boundary can be computed from the values where the solubility product reaches the K_{SP} threshold. (D) Plots of the oligomer concentrations reveal that the re-entrant behavior seen in panel C is due to an excess of 1:1 or 1:2 oligomers. Note that the oligomer concentrations in the shaded regions are not reliable because this calculation does not account for molecules in the dense phase. $s = 1.4$, $K_{1,1} = e^{-15/s}$, $K_{2,1} = e^{-15}$, and $f_A = -15.75$.

of c_A for a fixed value of c_B (indicated by the dotted line in Fig. 1C). At low and high concentrations of A the solubility product lies below the solubility threshold K_{SP}, indicating a single phase system. However, at intermediate concentrations, indicated by the shaded region, the solubility product rises above K_{SP}. The locus of points where $c_a c_B^s = K_{\mathrm{SP}}$ indicates the phase boundary, which is plotted as a red line in Fig. 1C. Fig. 1D shows how the population of oligomers shifts along the path shown in panels B and C (with the caveat that accounting for the dense phase will reduce the concentration in the shaded, two-phase region). In agreement with the assumptions of the model, the low and high concentration regimes are dominated by oligomers where the valence of the minority component is saturated. This saturation is responsible for the re-entrant behavior observed in Fig. 1C.

Stoichiometry matched oligomers increase sharply in concentration

In general, we expect that the molecular affinity in the dense phase, f_A, will be more attractive than the per molecule oligomer affinities $F(n,m)/n$ due to the additional conformational entropy and binding combinatorics when each molecule is surrounded by interaction partners. However, the dense phase is inhibited by the large translational entropy penalty coming from the $(c_a c_b^s)^{n_{\min}}$ factor in the numerator of Q_{den}. Therefore, condensation begins when the affinity advantage (f_A vs. $F(n,m)$) overwhelms the mixing entropy penalty, at which point Q_{den} dominates Q_{oligo}. However, the transition between oligomer accumulation and dense phase formation is highly sensitive to the molecular details. In particular, previous work has shown that large shifts in the phase boundary can occur when the valences of the two species differ by an integer ratio, which has been called the "magic number effect".[26–29]

To understand the large shift in the phase boundary at integer valence ratios, we take the simplest case where the A and B molecules have nearly matched valences (the general case is shown in the Appendix). For closely matched valences we can write $s = 1 + \epsilon$. Using ϵ as a small parameter, the phase boundary, Eq. 8, can be written as $c_a c_b = K_{SP} c_b^{-\epsilon}$. The dominant oligomer species will be the heterodimer, the concentration of which is given by

$c_{1,1} = c_a c_b / K_{1,1}$. Therefore, the concentration of dimer at the phase boundary is

$$c_{1,1} = \frac{K_{SP}}{K_{1,1}} c_b^{-\epsilon} \tag{13}$$

Phase separation will occur most readily when the dimer concentration is low, because this leaves more monomers to reach the K_{SP} threshold. In agreement with this expectation, we see that highly attractive dense phases (small K_{SP}) lower the dimer concentration, while high affinity dimers (small $K_{1,1}$) increase it.

Consider the case where the A molecules are unchanged and $\epsilon = s - 1$ is varied by changing the B valence. We further assume that the solution contains an equimolar mixture of molecules such that $c_a = c_b = c$. This allows us to rewrite the threshold dimer concentration as

$$c_{1,1} = \frac{K_{SP}}{K_{1,1}} \exp[-\frac{\epsilon}{2} \ln c^2] \tag{14}$$

$$= \frac{K_{SP}}{K_{1,1}} \exp[-\frac{\epsilon}{2} \ln(K_{1,1} c_{1,1})] \tag{15}$$

Next, we recursively plug Eq. 15 back into itself to obtain

$$c_{1,1} = \frac{K_{SP}}{K_{1,1}} \exp[-\frac{\epsilon}{2} \ln K_{SP}] \tag{16}$$

where we have dropped terms of order ϵ^2. While it is possible to sum the perturbation series to obtain an exact solution (see Appendix), the first approximation is sufficient to see the magic number phenomenon. We write the equilibria constants in terms of the free energies

$$c_{1,1} = \exp[f_A - F(1,1) - \frac{\epsilon}{2} f_A] \tag{17}$$

In the tight binding limit, the dimer free energy is determined by the maximum number of bonds that can be formed. Therefore, $F(1,1) = f_1 \min(v_A, v_B) = f_1 v_A \min(1, (1+\epsilon)^{-1})$,

where f_1 is the energy of a monovalent bond. Therefore, for $\epsilon < 0$ the threshold dimer concentration varies as

$$c_{1,1} = \exp[f_A - f_1 v_A - \epsilon f_A/2] \tag{18}$$

which increases exponentially as ϵ approaches zero. For $\epsilon > 0$ we have $F(1,1) = f_1 v_A/s \simeq f_1 v_A(1-\epsilon)$, so the threshold dimer concentration varies as

$$c_{1,1} = \exp[f_A - f_1 v_A + \epsilon(f_1 v_A - f_A/2)] \tag{19}$$

which *decreases* exponentially provided that $f_1 v_A < f_A/2$. This means that $c_{1,1}$ has a cusp at $\epsilon = 0$ with exponential drops on either side (Fig. 2A). The large increase dimer concentration depletes the monomer pool, which causes the phase boundary to shift to much higher concentrations (Fig. 2B).

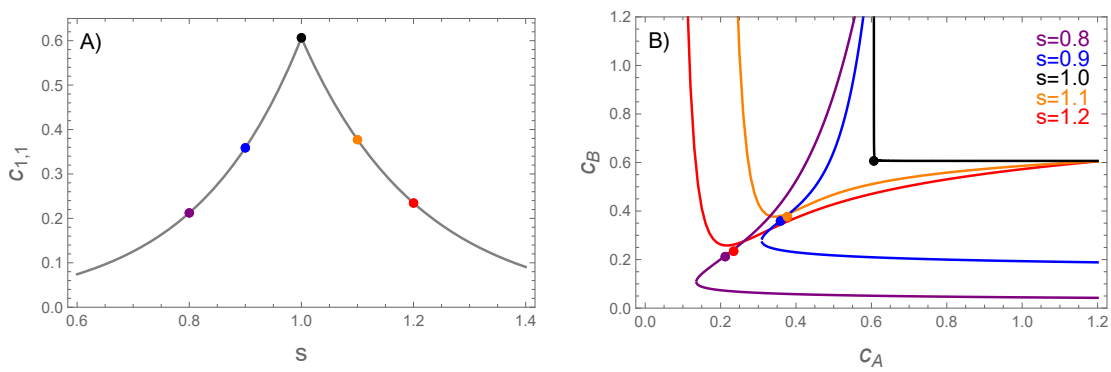

Figure 2: A) The heterodimer concentration at the phase boundary (computed from Eqs. 18 and 19) has a cusp peak at $s = 1$. B) As a result of this peak, the phase boundary shows a large shift at $s = 1$. The colored dots show the dimer concentrations at the points labeled in panel A. The dots lie very close to the phase boundary, indicating dimers account for most of the molecules in the dilute phase. $F(1,1) = f_1 v_A \min(1, 1/s)$, $f_A = 1.05 v_A f_1$, $v_A f_1 = 10$

Inspection of Eqs. 18 and 19 allows us to extend these arguments outside of the tight binding limit. The dense phase translational entropy, represented by the $f_A/2$ term, will always favor larger B valences. Similarly, the oligomer free energy $F(1,1)$ will also favor large v_B, but with substantially diminishing returns once the B valence exceeds v_A. Therefore, we can expect a magic number peak in the phase boundary whenever the change in the slope of

$F(1,1)$ is large compared to $f_A/2$. This will not be true in the weak binding regime where entropic contributions dominate the change in $F(1,1)$ or in cases where molecular geometry restricts bond saturation in the dimer.[27,36] However, in such cases there may be a larger oligomer that has a magic number peak at the phase boundary (see Appendix).

IV. Solubility with soft stoichiometry constraint

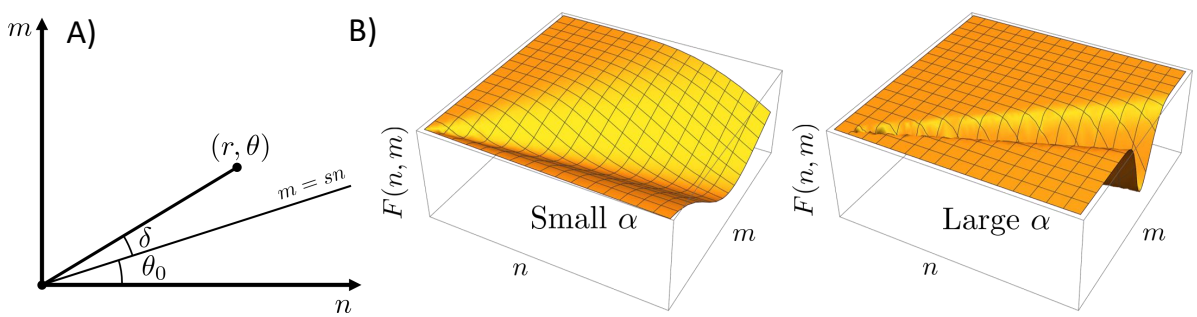

Figure 3: (A) Coordinate change from cartesian axes (n, m) to polar coordinates (r, θ). The radial integration is preformed over the angle $\delta = \theta - \theta_0$, which is measured relative to the line of optimal stoichiometry. (B) Schematic plots of $F(r, \theta)$ for small and large values of α. Larger values of α correspond to narrower low energy troughs in the free energy landscape and clusters that are more resistant to changes in composition.

Unlike salts, biomolecular condensates are not restricted to a specific stoichiometry. This flexibility can be attributed to the liquid-like disorder. To understand the effects of compositional variation in the dense phase, we begin with the dense phase partition function, Eq. 5. The summations in Eq. 5 encompass the upper right quadrant $n, m > 0$, which can also be covered using polar coordinates

$$Q_{\text{den}} = \int_0^\infty \int_0^{\pi/2} c_a^{r\cos\theta} c_b^{r\sin\theta} e^{-F(r,\theta)} r dr d\theta \tag{20}$$

The line of ideal stoichiometry $m = sn$ can be expressed in terms of an angle $\tan\theta_0 = s$ (see Fig. 3A). We expect that the free energy $F(r, \theta)$ will restrict the significant contributions to the integral to regions near the ideal stoichiometry. Therefore, we expand the condensed

phase free energy to lowest order in $\delta = \theta - \theta_0$.

$$F(r,\theta) = r f_r + r\frac{\alpha}{2}(\theta-\theta_0)^2 \tag{21}$$

$$= r f_r + r\frac{\alpha}{2}\delta^2 \tag{22}$$

Here f_r describes how the free energy scales with r when the condensate has the ideal stoichiometry

$$f_r = \lim_{n\to\infty}\frac{F(n,ns)}{r} = \frac{f_A}{\sqrt{1+s^2}} \tag{23}$$

and α is a parameter that describes how the free energy varies as the stoichiometry deviates from the ideal ratio

$$\alpha = \lim_{r\to\infty}\left.\frac{\partial^2 F(r,\theta)/r}{\partial\theta^2}\right|_{\theta_0} \tag{24}$$

The tight stoichiometry limit can be recovered in the limit $\alpha \to \infty$. The gaussian approximation in Eq. 22 will be valid when the stoichiometry fluctuations are small enough that anharmonic terms ($\mathcal{O}(\delta^3)$) in the free energy are negligible. For consistency with the expansion of $F(r,\theta)$, we also Taylor expand the chemical potential terms to second order in δ. The partition function with the expanded free energy takes the form (see Appendix)

$$Q_{\text{den}} \simeq \int_0^\infty\int_0^{\pi/2} e^{-r\delta^2 a + r\delta b - rc} r\,dr\,d\delta \tag{25}$$

where

$$a = \frac{1}{2\sqrt{1+s^2}}\left(\alpha_A + \ln\left(c_a c_b^s\right)\right) \tag{26}$$

$$b = \frac{1}{\sqrt{1+s^2}}\ln\frac{c_b}{c_a^s} \tag{27}$$

$$c = \frac{1}{\sqrt{1+s^2}}\left(f_A - \ln\left(c_a c_b^s\right)\right) \tag{28}$$

where $\alpha_A = \sqrt{1+s^2}\alpha$. Evaluation of the integrals in Eq. 25 (see Appendix) yields

$$Q_{\text{den}} = \frac{\pi(1+s^2)}{\sqrt{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)}\left(f_A - \ln\left(c_a c_b^s\right) - \frac{\left(\ln\frac{c_b}{c_a^s}\right)^2}{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)}\right)^{3/2}} \tag{29}$$

Composition fluctuations expand the two-phase regime

The phase boundary lies at the point were the partition function diverges. Inspection of Eq. 29 shows that this occurs when

$$c_a c_b^s = \exp\left[f_A - \frac{\left(\ln\frac{c_b}{c_a^s}\right)^2}{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)}\right] \tag{30}$$

Using the recursive approximation $\ln c_a c_b^s \simeq f_A$, we can define the effective resistance to composition fluctuations $\alpha' = \alpha_A + f_A$. This simplifies the solubility threshold to

$$c_a c_b^s = \exp\left[f_A - \frac{1}{2\alpha'}\left(\ln\frac{c_b}{c_a^s}\right)^2\right] \tag{31}$$

Eq. 31 has nearly the same form as the expression we found for the solubility product in the limit of a tight stoichiometry constraint (Eq. 8). The difference is a correction term that accounts for fluctuations in the composition of the dense phase. Note that the fluctuation term is always negative since α' must be positive for the gaussian integral to converge. Therefore, the fluctuation term lowers the free energy of the dense phase relative to the tight stoichiometry case. This is expected because loosening the stoichiometry constraint contributes additional terms to the partition sum, which must lower its free energy.

Eq. 31 can be solved for c_b (see Appendix), which yields an expression for the phase boundary

$$c_b = \exp\left[s\alpha'\left(-1 + \frac{\ln c_a}{\alpha'} + \sqrt{1 - \frac{2\ln c_a}{\alpha'} + \frac{2f_A}{\alpha' s^2} - \frac{2\ln c_a}{\alpha' s^2}}\right)\right] \tag{32}$$

Eq. 32 is plotted for various values of α in Fig. 4. We see that decreasing α has the effect of introducing curvature in the phase boundary when plotted on double logarithmic axes.

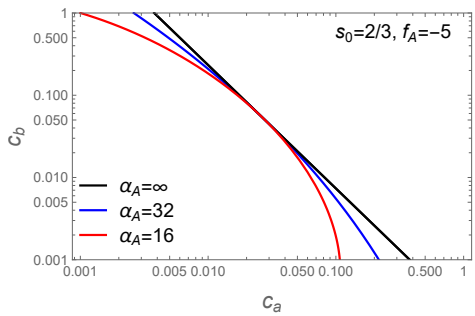

Figure 4: Plot of Eq. 32 for $s = 2/3$. The $\alpha = \infty$ limit is linear with slope $-s^{-1}$ on a log-log axes (black lines) due to the power law behavior of the tight stoichiometry case (Eq. 7). Lower values of α introduce curvature to the phase boundary and increase the phase separated region (upper right) at the expense of the single phase region (lower left). The curved phase boundary is tangent to the black line when $s \ln c_a - \ln c_b = 0$.

Phase boundary curvature describes variations in the dense phase stoichiometry

Our next task is to determine how the composition of the dense phase varies with changes in the solution stoichiometry. In cases where the concentrations of the dense phase are known, this calculation is equivalent to determining the tie lines connecting the dense and dilute phase. Measurement of the tie line slope has proven to be a useful tool to understand the balance of attractive and repulsive interactions driving phase separation.[37]

The total concentration of dense phase molecules can be computed using

$$c_A^{\text{den}} = c_a \frac{dQ_{\text{den}}}{dc_a} \tag{33}$$

$$c_B^{\text{den}} = c_b \frac{dQ_{\text{den}}}{dc_b} \tag{34}$$

Evaluation of the derivatives gives the stoichiometric ratio (see Appendix)

$$\frac{c_B^{\text{den}}}{c_A^{\text{den}}} = \frac{s + \frac{1}{2\alpha'} \ln\left(\frac{c_b}{c_a^s}\right)}{1 - \frac{s}{2\alpha'} \ln\left(\frac{c_b}{c_a^s}\right)} \tag{35}$$

Both the numerator and denominator have correction terms that are proportional to α'^{-1}, so they become negligible in the tight stoichiometry limit ($\alpha \to \infty$). In this limit we recover the optimal stoichiometry, s. Additionally, we obtain the optimal stoichiometry when $s \ln c_a - \ln c_b = 0$. This expression describes the balance of chemical potentials required for a growth increment of Δn type A molecules to incur the same mixing entropy penalty as recruiting $s\Delta n$ molecules of type B. Notably, this condition (i.e., $c_a^s = c_b$) is different than the more intuitive expectation that the dense phase has the optimal stoichiometry when the dilute phase has a matching stoichiometry $sc_a = c_b$ (except for the special case $s = 1$). When $s \ln c_a - \ln c_b \neq 0$ the dense phase stoichiometry is perturbed such that the molecule that is present in excess in the dilute phase is also over-represented in the dense phase.

Comparison of Eqs. 31 and 35 reveals that a single parameter, α', controls both the compositional variation in the dense phase and the deviation of the phase boundary from a $-s^{-1}$ power law. Therefore, the curvature of the phase boundary, when plotted on double logarithmic axes can be used to predict the composition of the dense phase.

Simulations of sticker-spacer polymers show a curved phase boundary

To test the predicted relationship between the phase boundary and the dense phase composition, we conducted simulations of a bead-spring polymer model. Within each chain, certain beads behave as stickers while others serve as spacers to impart flexibility[21] (Fig. 5A). Our model only includes heterotypic interactions in which a bond requires one sticker of type A and one of type B. The number of stickers per molecule defines the molecular valences, v_A

and v_B.

Each simulation starts with a predetermined number of each molecule, uniformly distributed in a cubic simulation volume. The systems evolve by Langevin dynamics using the SpringSaLaD software package.[38] As a result of binding interactions, these multivalent molecules form clusters, where a cluster includes all beads connected by intra-chain or intermolecular bonds (Fig. 5B).

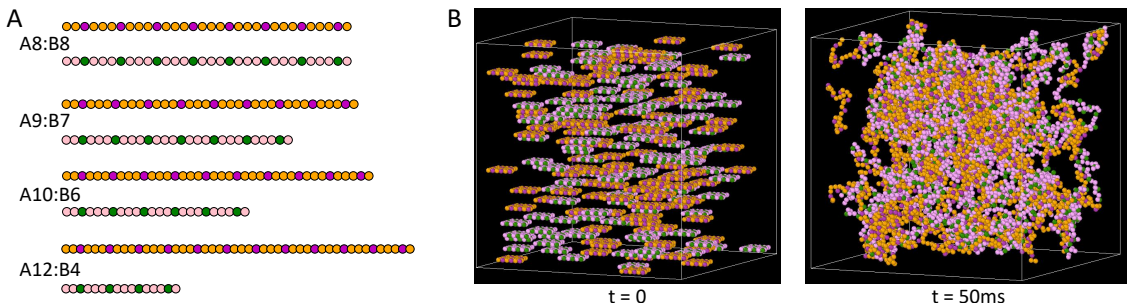

Figure 5: (A) Schematic representations of the bead-spring polymers used in the SpringSaLaD simulations. Green and magenta beads represent the A and B-type stickers, respectively, while the orange and pink beads represent spacer segments. (B) Snapshots of the A8:B8 system with $N_A = N_B = 120$ molecules at the start of the simulation and after 50ms.

We first establish the convergence of our simulation by tracking the time evolution of the cluster size distribution and bond saturation (Fig. 6A-C). Multiple stochastic runs are performed to extract an average behavior of the system. Fig. 6A shows how the system approaches steady state, where the system bifurcates between a small fraction of molecules in very small clusters and a large fraction in clusters containing > 20 molecules (total number of A+B molecules is 240). Figs. 6B and C display the average clustering and binding saturation dynamics. Once the system fluctuates around a steady state, we sample the cluster size distribution that consists of monomers (unbound molecules), oligomers (smaller clusters) and large clusters. Fig. 6D shows the extent of bond saturation as the clusters grow large in size showing that, while small clusters vary considerably in bond saturation, large clusters converge to a constant value of about 0.45.

To obtain the phase boundary, we varied the number of A and B molecules across eight

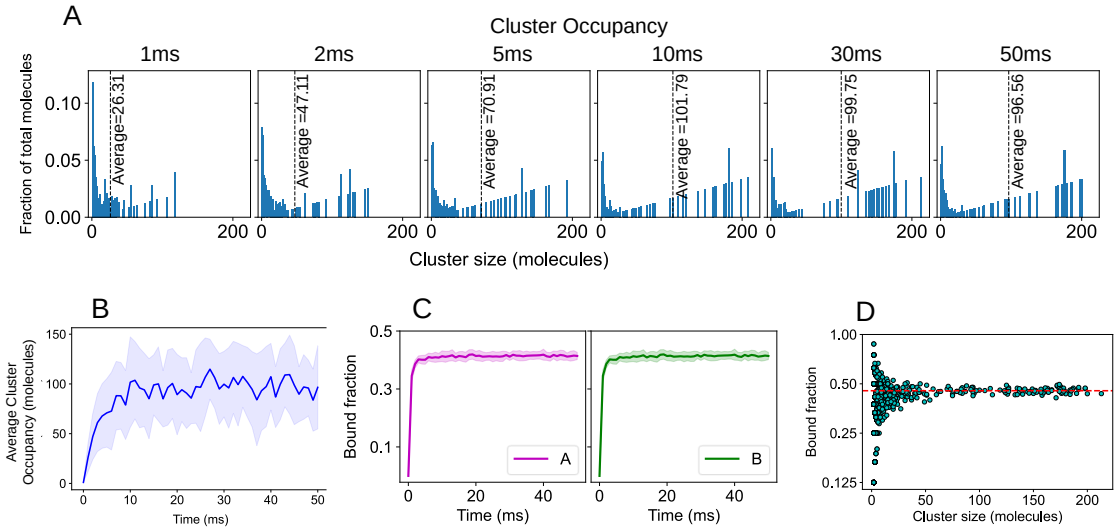

Figure 6: SpringSaLaD simulation of A8:B8 system. (A) Histograms of the normalized cluster occupancy, which we define as the fraction of molecules ($Nc_N/(c_A + c_B)$) in clusters of size N, where N is the total number of A and B molecules in the cluster $N = n + m$ and c_N is the concentration of all clusters of that size $c_N = \sum_{n+m=N} c_{n,m}$. The average cluster occupancy (dotted line) grows initially before converging to a constant value at about 10 ms. Each panel shows an average over 25 stochastic runs. (B) Time evolution of the average cluster occupancy averaged over 25 runs. The solid line shows the mean and the shaded area is standard deviation. (C) Bonded sticker fraction as a function of time, averaged over 25 runs. (D) Bonded sticker fraction as a function of cluster size obtained from 4 times (20, 30, 40, 50ms) across 25 runs (total 100 frames). Length of the cubic simulation box = 120 nm and molecular count, $N_A = N_B = 120$, molecular concentration ~ 115 μM.

values for a total of 64 mixtures. According to Eq. 8, the phase boundary is defined in terms of the monomer concentrations. Fig. 7A,C shows that the monomer concentrations are clustered along a curved line. The absence of points above the line indicates the presence of a phase boundary. The scatter of points below and/or left of the boundary is due to either initial conditions that are sub-saturated or highly supersaturated simulations where obtaining an accurate monomer concentration would require correcting for the volume occupied by the dense phase. To avoid these complications, we approximate the phase boundary by the points along the upper edge of the cluster.

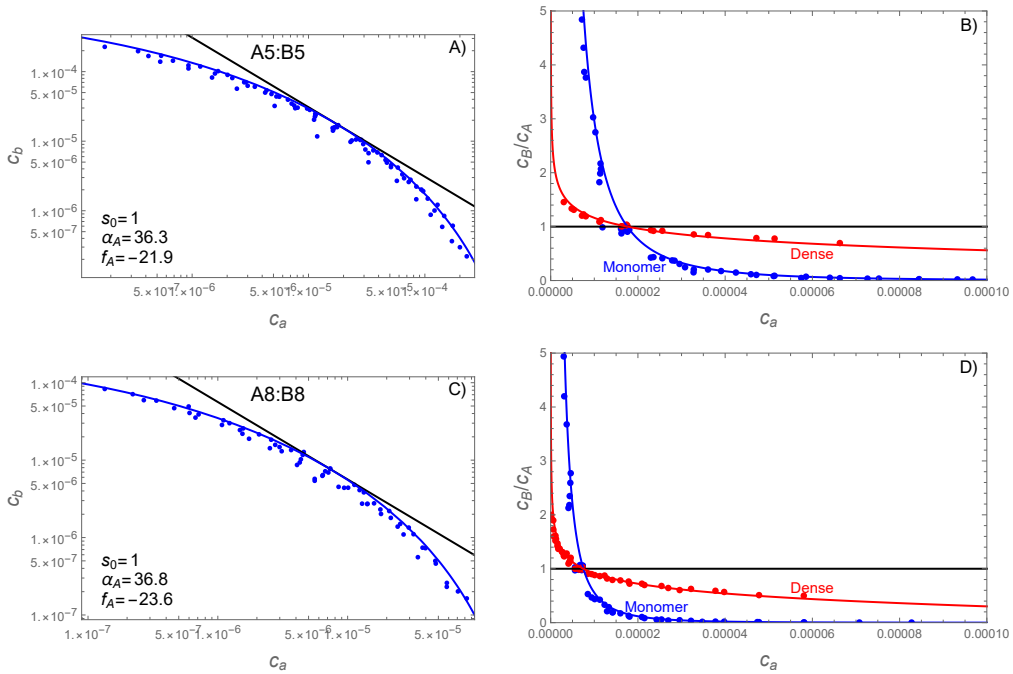

Figure 7: SpringSaLaD simulations reveal a curved phase boundary that predicts composition variations in the dense phase. Panels A,B show molecules with 5 stickers per chain ($v_A = v_B = 5$) while panels C,D show $v_A = v_B = 8$. Left panels (A,C) show the average concentration of free monomers for each of the simulated mixing ratios. Fitting Eq. 32 to the uppermost points yields values for f_A and α_A. Right panels (B,D) show the stoichiometric ratios B:A for both the monomers (blue) and dense phase (red). Note that the monomer lines and data (blue) are equivalent between panels with only a change in the vertical axis. The monomer ratio is calculated from Eq. 32 divided by c_a and the dense phase ratio is calculated from Eq. 35 with Eq. 32 substituted for c_b. The stoichiometric composition of the dense phase agrees well with the composition of clusters $n + m \geq 20$ molecules observed in the simulations. In all panels simulation data are shown as points and theory as a solid line.

The simulations where both the A and B molecules contain 5 stickers (hereafter "A5:B5") are shown in Fig. 7A while the A8:B8 simulations are shown in Fig. 7C. In both cases the phase boundary curves substantially away from the $c_b \propto c_a^{-1}$ power law (black lines) expected for a strict $s = 1$ stoichiometry. However, the phase boundaries can be adequately fit by Eq. 32 a two-step process in which f_A is first adjusted such that $c_b = e^{f_A}/c_a$ (black line) is tangent to the observed phase boundary. The second step is to adjust α_A to match the observed curvature (blue lines). Because the most reliable data for the phase boundary lie at the upper edge of the distribution, all fits were done visually (not by least squares regression).

Phase boundary curvature predicts composition of the dense phase.

Having obtained f_A and α_A from the phase boundary, we can predict the composition of the dense phase without additional parameters. Fig. 7B,D shows the B:A ratio for both the monomers and dense phase. The red lines show the predicted composition of the dense phase, obtained by plugging Eq. 32 into Eq. 35. This is in good agreement with the observed composition (red points) of clusters containing $n + m \geq 20$.

For asymmetric simulations, where the A and B molecules have different valences, the situation is slightly more complicated because the optimal composition s of the dense phase is not known *a priori*. This creates uncertainty in which power law should be used as the reference tangent to the phase boundary. However, inspection of Fig. 4 reveals a second condition that can be used to obtain s. Specifically, the $-s^{-1}$ power law must be tangent to the phase boundary at the point of balanced chemical potentials $s \ln c_a = \ln c_b$ (see Eq. 31). This second requirement again allows for a two-step fitting procedure where f_A and s are determined first, such that the power law and large α_A lines share a common tangent to the phase boundary, and then α_A is adjusted to match the observed curvature.

Fig. 8 compares the theory to the simulations of A9:B7, A10:B6, and A12:B4 molecular pairs. Again, the parameters obtained from fitting the phase boundary do an excellent job of

predicting the composition of the dense phase across the phase boundary. Interestingly, the s values are far below what would be expected from the valence ratio. While the A9:B7 system has a valence ratio of 1.29, the obtained value of s was 1.02; for A10:B6 we fit 1.06 instead of the expected 1.67; and for A12:B3 we fit 1.10 instead of the expected 3.0. Reassuringly, the stoichiometry values obtained from matching the tangent of the phase boundary are close to the large cluster compositions in the simulations (see Appendix Fig. 10). Of course, the valence ratio will only give the correct stoichiometry in the strong-binding limit, where nearly all bonds are satisfied in the dense phase. In contrast, our simulations used a relatively weak affinity of 350 μM where even the minority molecule only forms about half of its possible bonds (see Appendix Fig. 10). Under such conditions other contributions to the free energy, primarily the mixing entropy, will bias the dense phase composition toward more balanced ratios. The low values we found for s are most likely due to large activity coefficients in the dense phase that make it difficult to achieve the high concentrations needed for the low-valence species.

Inspection of Fig. 8F shows that the red, blue, and black lines do not intersect at a common point. As a result, when the monomer phase has the ideal stoichiometry for the dense phase $s = 1.1$, the dense phase has a composition ratio of ~ 1.4. This is because the dense phase reaches the optimal composition with the chemical potential balance condition $c_a^s = c_b$ (see discussion following Eq. 35) and not when the monomer phase has the matching composition.

Discussion

Our calculation shows that the solubility product is a robust description for salts due to two criteria that are often inappropriate for multi-component biomolecular condensates. The first of these is the neglect of soluble species larger than monomers, which is valid for salts due to a large surface tension suppressing clusters smaller than the critical nucleus. In contrast,

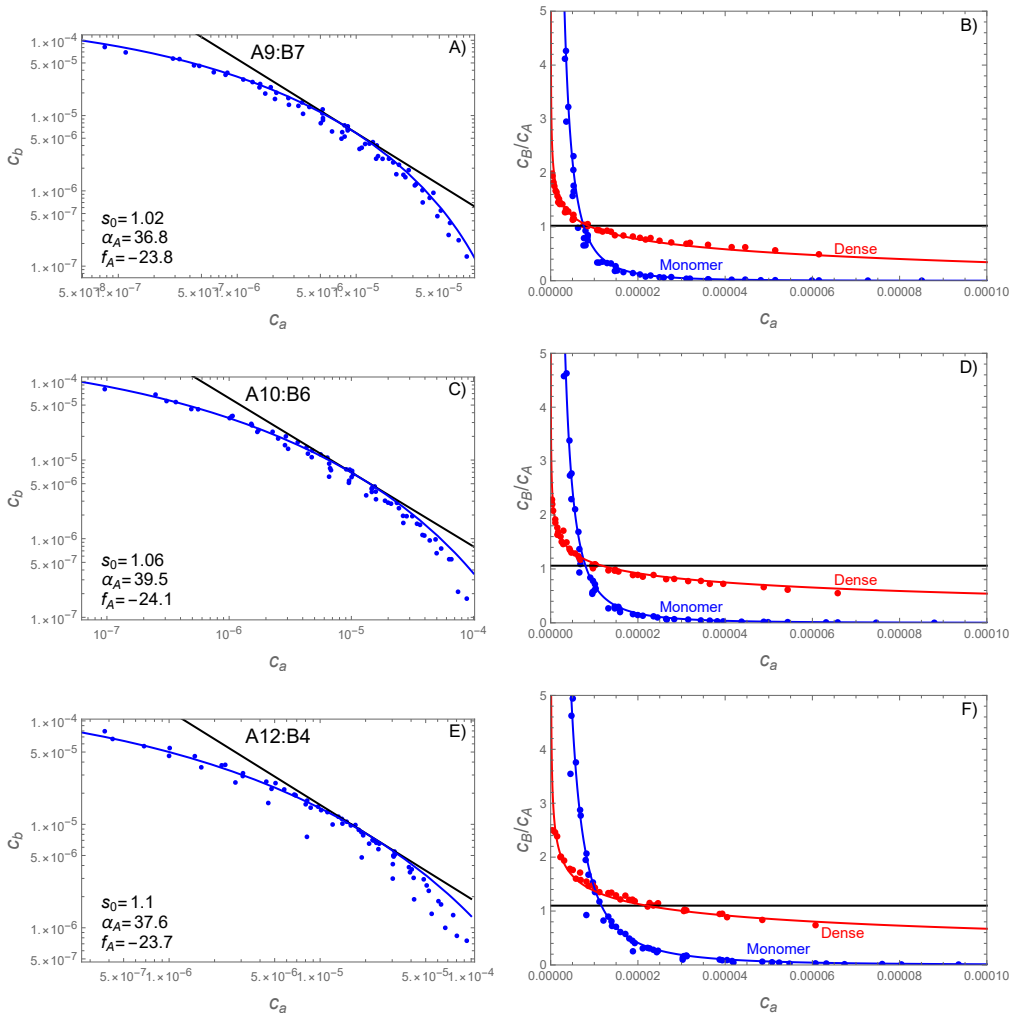

Figure 8: SpringSaLaD simulations reveal a curved phase boundary that predicts composition variations in the dense phase. Panels A,B show $v_A = 9$, $v_B = 7$; panels C,D show $v_A = 10$, $v_B = 6$; panels E,F show $v_A = 12$, $v_B = 4$;. Left panels (A,C,E) show the average concentration of free monomers for each of the 64 simulated mixing ratios. Fitting Eq. 32 to the uppermost points yields values for f_A and α_A. Right panels (B,D,F) show the stoichiometric ratios B:A for both the monomers (blue) and dense phase (red). Note that the monomer lines and data (blue) are equivalent between panels with only a change in the vertical axis. The monomer ratio is calculated from Eq. 32 divided by c_a and the dense phase ratio is calculated from Eq. 35 with Eq. 32 substituted for c_b. The stoichiometric composition of the dense phase agrees well with the composition of clusters $n + m \geq 20$ molecules observed in the simulations. In all panels simulation data are shown as points and theory as a solid line.

biomolecular condensates tend to have comparatively low surface tension.[39] Biomolecular condensates also have a high propensity to form clusters in the sub-saturated regime[20,40] due to multiple factors including stoichiometry matching,[26,34] conformational states,[41] and hierarchical assembly involving multiple interactions.[42] The complication due to oligomers can be alleviated by considering the solution to be a three-state equilibrium between monomers, the condensed state, and the oligomer ensemble (Fig. 9). The equilibrium between the monomer and condensed phases resembles the power law of the solubility product and the sharpness of the condensation transition allows for straightforward calculation of the oligomer equilibrium in the sub-saturated regime.

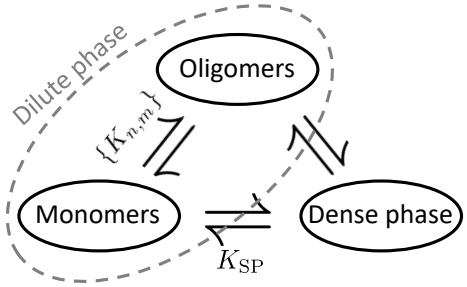

Figure 9: The equilibrium between monomers and oligomers can be calculated from dissociation constants while the equilibrium between monomers and the dense phase is described by the solubility product. The lack of a simple relationship between oligomers and the dense phase is responsible for the complexities associated with describing partitioning between the dense and dilute states.

The second complication is that, unlike salts, biomolecular condensates do not have tightly constrained stoichiometry in the dense phase. Compositional variation lowers the free energy of the dense phase, expands the two-phase region of the phase diagram, and causes the monomer-condensate phase boundary to deviate from a simple power law. However, the power law deviation can be informative because it is predictive of how the composition of the dense phase varies.

In recent years, the search for phase separated compartments in cells has revealed a growing list of molecules that form puncta within cells. Since many of these puncta are quite small (< 100 nm), it raises the question of whether these structures should be con-

sidered biomolecular condensates, or simply clusters of molecules.[43] Our decomposition of the solution into monomer, oligomer, and dense phases provides a framework to resolve this issue. Specifically, molecular assemblies should be considered condensates when they meet the linear scaling requirements necessary for inclusion in Q_{den}. Assemblies that do not meet this criteria belong in Q_{oligo} and should be considered molecular clusters. There are at least two mechanisms by which an assembly might fail the criteria for Q_{den}. First, generically we expect a surface contribution to the free energy that scales as $n^{2/3}$, making it significant for small clusters but negligible for macroscopic phases. Second, biomolecular condensates often assemble from a variety of interaction motifs ranging from specific interactions between globular domains to weak interactions between IDRs. This will naturally lead to hierarchical assembly[42] in which the stronger interactions form multimer clusters and the weaker interactions drive coalescence of multimer subunits. The relatively high affinity interactions responsible for the initial clusters are not representative of the free energy of the larger assembly and, therefore, these clusters should be included in Q_{oligo}.

While our calculation provides intuition into factors that shape biomolecule phase boundaries, it does not account for many of the complexities of membraneless organelles in cells. Most severely, our calculation assumes the system is in equilibrium and, therefore, cannot account for the myriad of non-equilibrium and active processes that influence condensate formation.[44] We have also limited our calculation to two components. In contrast, biomolecular condensates can have hundreds of components that are enriched above the background level, although usually only a small number of scaffold species are necessary to reconstitute condensation *in vitro*.[45] While it would be straightforward to generalize our method to more components, this would come at the cost of more parameters. Finally, many condensates have multiple structural states due to aging[46] or biological responsiveness.[24,42,47] Such transitions are due to multiple minima in the condensed phase free energy. Our calculation assumes a single free energy minimum and, therefore, cannot capture changes in the phase boundary due to a transition between conformational states. This deficiency could be exploited, how-

ever, because deviations from our theory could indicate the presence of biologically relevant states.

Conclusion

Many of the complexities in understanding the phase boundaries of biomolecular condensates can be attributed to the appearance of partitioning between two states, dense and dilute. It is conceptually easier, instead, to consider a three-state partitioning between monomers, the dense phase, and the oligomer ensemble (Fig. 9).[19,28,33,48] In the three-state picture, the monomer-oligomer equilibrium can be addressed with association constants, while the monomer-dense phase equilibrium is captured by the solubility product. This separation allows for greater intuition into how factors like bond saturation and compositional variations shape experimentally determined phase boundaries.

Appendix

Magic number effect

Consider a (n, m) oligomer where m/n is close to the stoichiometry of the dense phase. The oligomer and dense phase equilibrium conditions are

$$c_{n,m} = c_a^n c_b^m e^{-F(n,m)} \tag{36}$$

$$c_a^n c_b^{ns} = e^{n f_A} \tag{37}$$

where the second line is Eq. 8 raised to the nth power. Using $\epsilon = s - m/n$ as a small parameter, the phase boundary can be manipulated to find the concentration of the (n, m)

oligomer at the onset of phase separation

$$c_a^n c_b^m = e^{nf_A} c_b^{-n\epsilon} \tag{38}$$

$$c_{n,m} e^{F(n,m)} = e^{nf_A} c_b^{-n\epsilon} \tag{39}$$

Solving for the oligomer concentration we have

$$c_{n,m} = \exp[nf_A - F(n,m) - n\epsilon \ln c_b] \tag{40}$$

We assume that the A valence is held fixed while the B valence is varied and that the solution mixing ratio is held fixed at a value $S = c_b/c_a$ as the valence is changed. In this case $c_a^n c_b^m = c_b^{n+m} S^{-n}$. Therefore the threshold oligomer concentration can be written as

$$c_{n,m} = \exp\left[nf_A - F(n,m) - \frac{n\epsilon}{n+m} \ln c_b^{n+m}\right] \tag{41}$$

$$= \exp\left[nf_A - F(n,m) - \frac{n\epsilon}{n+m} \ln c_a^n c_b^m S^n\right] \tag{42}$$

$$= \exp\left[nf_A - F(n,m) - \frac{n\epsilon}{n+m} \ln(c_{n,m} e^{F(n,m)} S^n)\right] \tag{43}$$

This expression can be used recursively to obtain a power series in ϵ. Applying the recursion iteratively gives

$$c_{n,m} = \exp\left[nf_A - F(n,m) - \frac{n\epsilon}{n+m}\ln(S^n e^{nf_A - \frac{n\epsilon}{n+m}\ln(c_{n,m}e^{F(n,m)}S^n)})\right] \tag{44}$$

$$= \exp\left[nf_A - F(n,m) - \frac{n\epsilon}{n+m}\left(n\ln S + nf_A - \frac{n\epsilon}{n+m}\ln(c_{n,m}e^{F(n,m)}S^n)\right)\right] \tag{45}$$

$$= \exp\left[nf_A - F(n,m) - \frac{n\epsilon}{n+m}(n\ln S + nf_A) + \left(\frac{n\epsilon}{n+m}\right)^2 \ln(c_{n,m}e^{F(n,m)}S^n)\right] \tag{46}$$

$$= \exp\left[nf_A - F(n,m) - \frac{n\epsilon}{n+m}(n\ln S + nf_A) + \left(\frac{n\epsilon}{n+m}\right)^2 (n\ln S + nf_A) - \left(\frac{n\epsilon}{n+m}\right)^3 \ln(c_{n,m}e^{F(n,m)}S^n)\right] \tag{47}$$

$$= \exp\left[nf_A - F(n,m) - (n\ln S + nf_A)\frac{n\epsilon}{n+m}\sum_{j=0}^{\infty}\left(\frac{-n\epsilon}{n+m}\right)^j\right] \tag{48}$$

$$= \exp\left[nf_A - F(n,m) - (n\ln S + nf_A)\frac{n\epsilon}{n+m+n\epsilon}\right] \tag{49}$$

Note that the final term is a smooth function of ϵ because the pole at $\epsilon = -1 - m/n$ corresponds to the non-physical value $s = -1$. Eq. 49 allows for the calculation of the concentration of any oligomer along the phase boundary, provided $F(n,m)$ is known. To understand the magic number effect, we need to make assumptions about $F(n,m)$, in which case it is also convenient to employ the first order approximation for $c_{n,m}$ by taking Eq. 46 and dropping terms of order ϵ^2.

$$c_{n,m} = \exp\left[nf_A - F(n,m) - \frac{n^2\epsilon}{n+m}(f_A + \ln S)\right] \tag{50}$$

In the tight binding limit the oligomer free energy is dominated by the maximum number of bonds

$$F(n,m) = f_1 \min(v_A n, v_B m) \tag{51}$$

$$= f_1 n v_A \min\left(1, \frac{m}{ns}\right) \tag{52}$$

$$= f_1 n v_A \min\left(1, \frac{1}{1+\frac{n\epsilon}{m}}\right) \tag{53}$$

where f_1 is the free energy of a single intermolecular bond. Therefore, for $\epsilon < 0$ the threshold oligomer concentration varies as

$$c_{n,m} = \exp\left[n f_A - f_1 n v_A - \frac{n^2\epsilon}{n+m}(f_A + \ln S)\right] \tag{54}$$

which increases exponentially as ϵ approaches zero. For $\epsilon > 0$ we have $F(n,m) \simeq f_1 n v_A(1 - n\epsilon/m)$, so in the small ϵ limit the oligomer concentration at the phase boundary varies as

$$c_{n,m} = \exp\left[n f_A - n f_1 v_A - \frac{n^2\epsilon}{n+m}(f_A + \ln S) + \frac{n^2\epsilon}{m} f_1 v_A\right] \tag{55}$$

which *decreases* exponentially provided that $f_1 v_A < (f_A + \ln S)m/(n+m)$. This inequality can be interpreted as follows. The left side is the binding energy of the oligomer normalized by the number of A molecules. Similarly, f_A is the free energy per A molecule in the dense phase. However, at equilibrium f_A is balanced against the entropic cost of recruiting molecules to the condensate. Therefore, the right side represents the entropic cost of recruiting the B molecules to the dense phase. This interpretation is supported by the $m/(n+m)$ factor that selects only the B molecules in a stoichiometric unit and the $\ln S$ factor that corrects the translational entropy cost for the solution stoichiometry. Therefore, oligomer concentration has a magic number peak when the benefit of a B molecule binding in the oligomer state outweighs the translational entropy of recruitment to the dense phase.

Evaluation of partition function integrals

Our goal is to evaluate the double integral

$$Q_{\text{den}} = \int_0^\infty \int_0^{\pi/2} c_a^{r\cos\theta} c_b^{r\sin\theta} e^{-F(r,\theta)} r dr d\theta \tag{56}$$

The chemical potential terms can be Taylor expanded around θ_0 to obtain an expression consistent with the second-order expansion of the free energy. For the A molecules we have

$$\begin{aligned} n &= r\cos\theta && (57)\\ &= r\cos(\delta+\theta_0) && (58)\\ &= r(\cos\delta\cos\theta_0 - \sin\delta\sin\theta_0) && (59)\\ &= r((1-\delta^2/2)\cos\theta_0 - \delta\sin\theta_0) && (60)\\ &= \frac{r}{\sqrt{1+s^2}}\left(1 - s\delta - \frac{\delta^2}{2}\right) && (61) \end{aligned}$$

and for the B molecules

$$\begin{aligned} m &= r\sin\theta && (62)\\ &= r\sin(\delta+\theta_0) && (63)\\ &= r(\sin\delta\cos\theta_0 + \cos\delta\sin\theta_0) && (64)\\ &= r(\delta\cos\theta_0 + (1-\delta^2/2)\sin\theta_0) && (65)\\ &= \frac{r}{\sqrt{1+s^2}}\left(s + \delta - s\frac{\delta^2}{2}\right) && (66) \end{aligned}$$

where the last steps used the relations

$$\begin{aligned} \cos\theta_0 &= \frac{1}{\sqrt{1+s^2}} && (67)\\ \sin\theta_0 &= \frac{s}{\sqrt{1+s^2}} && (68) \end{aligned}$$

The expanded free energy (Eq. 22) and chemical potentials Eqs. 61 and 66 can be inserted back into the expression for the partition function

$$Q_{\text{den}} = \int_0^\infty \int_0^{\pi/2} c_a^{r\cos\theta} c_b^{r\sin\theta} e^{-F(r,\theta)} r dr d\theta \tag{69}$$

$$\simeq \int_0^\infty \int_0^{\pi/2} e^{-r\delta^2 a + r\delta b - rc} r dr d\delta \tag{70}$$

where

$$a = \frac{1}{2\sqrt{1+s^2}} \left[\alpha\sqrt{1+s^2} + \ln c_a + s \ln c_b\right] \tag{71}$$

$$b = \frac{1}{\sqrt{1+s^2}} \left(\ln c_b - s \ln c_a\right) \tag{72}$$

$$c = \frac{1}{\sqrt{1+s^2}} \left(f_A - \ln c_a - s \ln c_b\right) \tag{73}$$

Next we complete the square in the exponent

$$Q_{\text{den}} = \int_0^\infty \int_0^{\pi/2} e^{-r\left(\delta^2 a - \delta b + c\right)} r dr d\delta \tag{74}$$

$$= \int_0^\infty \int_0^{\pi/2} e^{-r\left(\delta\sqrt{a} - \frac{b}{2\sqrt{a}}\right)^2 + r\frac{b^2}{4a} - rc} r dr d\delta \tag{75}$$

$$= \int_0^\infty e^{r\frac{b^2}{4a} - rc} \int_0^{\pi/2} e^{-ra\left(\delta - \frac{b}{2a}\right)^2} r dr d\delta \tag{76}$$

To perform the integral over δ we extend the limits of integration to $\pm\infty$ so that the integral yields $(\pi/ra)^{1/2}$. While the extended limits introduce error by admitting regions where n, m are negative, contributions from these regions will be strongly suppressed by the unfavorable free energies at extreme stoichiometries. Therefore, this error is likely smaller than the neglect of anharmonic terms.

Now we perform the integral over r

$$
\begin{aligned}
Q_{\text{den}} &= \int_0^\infty \sqrt{\frac{\pi}{ra}} e^{r\frac{b^2}{4a} - rc} r dr && (77)\\
&= \sqrt{\frac{\pi}{a}} \frac{1}{\left(c - \frac{b^2}{4a}\right)^{3/2}} \int_0^\infty \sqrt{x} e^{-x} dx && (78)\\
&= \frac{\pi}{2\sqrt{a}\left(c - \frac{b^2}{4a}\right)^{3/2}} && (79)
\end{aligned}
$$

where the integration variable is $x = r(c - \frac{b^2}{4a})$ and the final integral is $\Gamma(3/2) = \sqrt{\pi}/2$. Here we have introduced another approximation by extending the integration range to zero. This captures terms in the oligomer range, however, the divergence that determines the phase boundary comes from the large r portion of the integral.

Inserting the values for a, b, and c into the partition function we have

$$
Q_{\text{den}} = \frac{\pi(1+s^2)}{\sqrt{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)}\left(f_A - \ln\left(c_a c_b^s\right) - \frac{\left(\ln \frac{c_b}{c_a^s}\right)^2}{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)}\right)^{3/2}} \quad (80)
$$

In this expression we see that the integral approximation to the partition function has introduced a second source of error in addition to extending the limits of the δ integration. Specifically, the above expression will not quantitatively reproduce the tight stoichiometry result. This can be seen by taking the limit $\alpha_a \to \infty$, which causes the partition function to vanish. This happens because the Gaussian that results from Taylor expanding the free energy $F(r,\theta) \simeq r(f + \alpha\delta^2)$ has vanishing width, and hence vanishing area, in the limit $\alpha \to \infty$. The denominator in the above expression is somewhat different than the tight stoichiometry limit (Eq. 7), although this is partly due to the integral approximation of the summation. Evaluating Eq. 7 using an integral, rather than a sum, yields a denominator $f_A - \ln c_a c_b^s$, which is equivalent to the small argument expansion of the exponential. Another significant change is the increase in the exponent of the divergent term from -1 to $-3/2$, which comes from including the second dimension in the partition sum.

Phase boundary

We can use the solubility relation (Eq. 31) to derive a condition for the phase boundary.

$$c_a c_b^s = \exp\left[f_A - \frac{1}{2\alpha'}\left(\ln\frac{c_b}{c_a^s}\right)^2\right] \tag{81}$$

With the definitions $x = \ln c_a$ and $y = \ln c_b$ this becomes

$$e^x e^{sy} = \exp\left[f_A - \frac{1}{2\alpha'}\left(y - sx\right)^2\right] \tag{82}$$

Taking the logarithm of both sides and rearranging

$$2\alpha'(x + sy - f_A) \;=\; -\left(y - sx\right)^2 \tag{83}$$

which is a quadratic equation for y. The physically relevant root is

$$y = sx - \alpha' s + \sqrt{(sx - \alpha' s)^2 - 2\alpha'(x - f_A) - s^2x^2} \tag{84}$$

$$= s\alpha'\left(-1 + \frac{x}{\alpha'} + \sqrt{1 - \frac{2x}{\alpha'} + \frac{2f_A}{\alpha' s^2} - \frac{2x}{\alpha' s^2}}\right) \tag{85}$$

In the limit of large α' we recover the slope of $-s^{-1}$ expected for a power law

$$y \simeq \frac{2f_A}{s} - \frac{x}{s} \tag{86}$$

Replacing x and y in Eq. 85 yields the phase boundary in the main text (Eq. 32).

It is straightforward, albeit slightly more cumbersome, to derive the phase boundary without employing the recursive approximation $\alpha' \simeq \alpha_A + \ln c_a c_b^s$. However, this more complicated expression yielded similar results in our comparison with simulations.

Dense phase stoichiometry

To compute the stoichiometric composition of the dense phase, we begin with the dense phase partition function

$$Q_{\text{den}} = \frac{\pi(1+s^2)}{\sqrt{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)}\left(f_A - \ln\left(c_a c_b^s\right) - \frac{\left(\ln\frac{c_b}{c_a^s}\right)^2}{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)}\right)^{3/2}} \tag{87}$$

The total number of molecules in the dense phase can be computed using

$$c_A^{\text{den}} = c_a \frac{dQ_{\text{den}}}{dc_a} \tag{88}$$

$$c_B^{\text{den}} = c_b \frac{dQ_{\text{den}}}{dc_b} \tag{89}$$

We note that the partition function has the form

$$Q_{\text{den}} = \frac{\pi(1+s^2)}{g^{1/2}h^{3/2}} \tag{90}$$

so that taking a derivative gives

$$Q'_{\text{den}} = Q_{\text{den}}\left(-\frac{g'}{2g} - \frac{3h'}{2h}\right) \tag{91}$$

We are interested in the composition of the dense phase, which will dominate the solution when $h \simeq 0$. Under these conditions the derivative is dominated by the second term of Eq. 91. The stoichiometry of the dense phase is therefore

$$\frac{c_B^{\text{den}}}{c_A^{\text{den}}} = \frac{c_b \frac{dQ}{dc_b}}{c_a \frac{dQ}{dc_a}} \simeq \frac{c_b \frac{dh}{dc_b}}{c_a \frac{dh}{dc_a}} \tag{92}$$

Evaluating the derivatives gives

$$c_a \frac{dh}{dc_a} = c_a \left(-\frac{1}{c_a} - \frac{-s\frac{1}{c_a}\ln\frac{c_b}{c_a^s}}{\alpha_A + \ln(c_a c_b^s)} + \frac{\frac{1}{c_a}\left(\ln\frac{c_b}{c_a^s}\right)^2}{2\left(\alpha_A + \ln\left(c_a c_b^s\right)\right)^2} \right) \tag{93}$$

$$\simeq \left(-1 - \frac{-s\ln\frac{c_b}{c_a^s}}{\alpha'} + \frac{\left(\ln\frac{c_b}{c_a^s}\right)^2}{2\alpha'^2} \right) \tag{94}$$

$$c_b \frac{dh}{dc_b} = c_b \left(-\frac{s}{c_b} - \frac{\frac{1}{c_b}\ln\frac{c_b}{c_a^s}}{\alpha_A + \ln(c_a c_b^s)} + \frac{\frac{s}{c_b}\left(\ln\frac{c_b}{c_a^s}\right)^2}{2\left(\alpha_A + \ln(c_a c_b^s)\right)^2} \right) \tag{95}$$

$$\simeq \left(-s - \frac{\ln\frac{c_b}{c_a^s}}{\alpha'} + \frac{s\left(\ln\frac{c_b}{c_a^s}\right)^2}{2\alpha'^2} \right) \tag{96}$$

The quantity $\ln(c_b c_a^{-s})/\alpha'$ is the driving force for composition fluctuations divided by the fluctuation stiffness. This can be used as a small parameter in the weak perturbation limit, implying that the third terms in the above equations are negligible. Keeping only the first two terms, we find that the the stoichiometric ratio is

$$\frac{c_B^{\text{den}}}{c_A^{\text{den}}} = \frac{s + \frac{1}{\alpha'}\ln\frac{c_b}{c_a^s}}{1 - \frac{s}{\alpha'}\ln\frac{c_b}{c_a^s}} \tag{97}$$

$$\simeq s + \frac{(1+s)}{\alpha'}\ln\frac{c_b}{c_a^s} \tag{98}$$

SpringSaLaD Simulations

SpringSaLaD software[38] has been used in multiple publications[17–19,23,34] to study clustering of multivalent biomolecules and its implications in phase separation. In this framework, each molecule is represented by a collection of beads connected by harmonic bonds (springs). Motion of particles are governed by an overdamped Langevin equation. The springs transmit force to the connected beads; additionally, a random force (uncorrelated in time) is applied to each bead at each timestep whose magnitude is determined by the diffusion coefficient

assigned to that bead. Each bead also has a radius.

In SpringSaLaD, certain beads engage in binding interactions while others serve as structural sites to mimic the actual molecular geometry. Each binding interaction is assigned association and dissociation rates whose ratio determines the macroscopic equilibrium constant. These rates are then converted to microscopic probabilities[38] in a thermodynamically consistent manner which guarantees the equilibrium distribution of molecular species. When particles diffuse in the simulation volume, reversible binding interactions are modelled with a Smoluchowski-like framework. When complementary binding sites are within a cutoff radius, they form a bond based on the probability computed from the association rate, particle radius, and diffusion coefficient. Similarly, the bond is broken with the probability calculated from the dissociation rate.

In our simulation setup, we first consider a pair of octavalent molecules (A8, B8, Fig. 1A). Each molecular type has 8 stickers (binding sites) and 24 spacers (structural sites). The binding affinity between sticker pairs is 350 μM (association rate = 10 $\mu\text{M}^{-1}\text{s}^{-1}$ and dissociation rate = 3500/s). We used a bead radius = 1 nm, diffusion coefficient = 2 $\mu\text{m}/s$, and the length between two consecutive beads is 2.5 nm.

We titrate up molecular count in a fixed volume to change the concentration. For all simulations except A5:B5 we used volume = $120 \times 120 \times 120\text{nm}^3$ and the molecular counts, N_A and N_B are taken from {15, 30, 45, 60, 75, 90, 105, 120}. As a reference, in the given volume, 1 molecule ~ 0.96 μM. We create a grid of concentrations (64 points) by taking all combinations of N_A and N_B from the aforementioned list. For the A5:B5 simulations the simulation volume is $100 \times 100 \times 100\text{nm}^3$ and the molecular counts, N_A and N_B are taken from {20, 40, 60, 80, 100, 120, 140, 160, 180}.

To create asymmetric molecular pairs, we extended (or reduced) the molecular length keeping the sticker:spacer ratio and total number of beads fixed. The chain characteristics are summarized using the notation $L(i, j)$ where i and j are the sticker and spacer count per molecule ($L = i + j$) respectively (see Fig. 10).

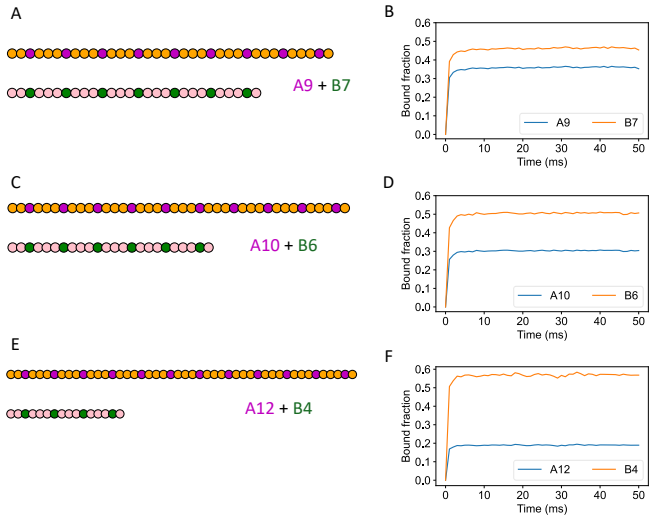

Figure 10: Panels A, C, and E show a schematic representation of the sticker and spacer patterns for the asymmetric simulations. Panels B, D, and F show bond saturation for the corresponding system. The bond saturation is $\sim 50\%$ for the minority component and even lower for the majority component. This low bond saturation allows for variations in the dense phase stoichiometry.

A8 + B8: L(8,24) + L(8,24)

A9 + B7: L(9,27) + L(7,21)

A10 + B6: L(10,30) + L(6,18)

A12 + B4: L(12,36) + L(4,12)

Due to the difference in valence (sticker stoichiometry), the asymmetric simulations show different degrees of bond saturation for A-type and B-type chains (Fig. 10B, D, and F). This difference grows as the sticker stoichiometry gets more imbalanced.

As the clusters grow in size, their compositions converge to a fixed value (Fig. 11). Smaller clusters show significant variation, but large clusters tend to have similar compositions.

Acknowledgements S.A.A., T.S., and J.D.S. acknowledge support from NIH grant R01GM141235. L.L. and A.C. acknowledge NIH grants R01GM132859 and R24GM137787. J.A.D. acknowledges support from Natural Sciences and Engineering Research Council Discovery Grant RGPIN-2022-03274.

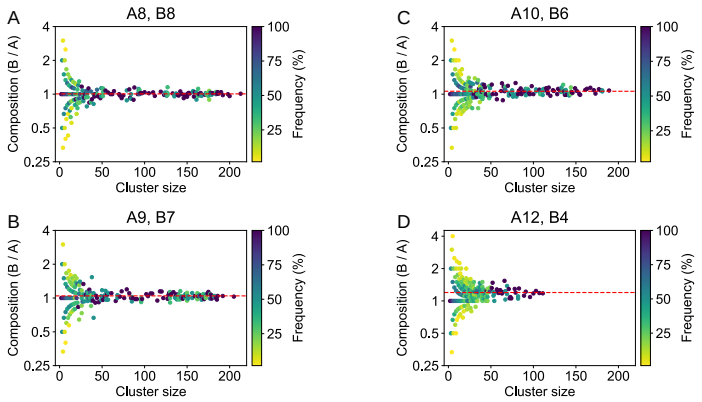

Figure 11: Summary of cluster composition for asymmetric (length and valence) molecule pairs in simulations with $N_A = N_B = 120$ molecules. The average composition for clusters larger than 50 molecules is given by the red dashed lines. The average values are 1.01 for A8:B8, 1.04 for A9:B7, 1.06 for A10:B6, and 1.20 for A12:B4.

Competing interests: Jonathon Ditlev is an advisor for Dewpoint Therapeutics and a scientific advisory board member for Neurophase Therapeutics.

References

(1) Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, eaaf4382.

(2) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* **2017**, *18*, 285–298.

(3) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M. et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162*, 1066–1077.

(4) Jain, A.; Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **2017**, *546*, 243–247.

(5) Woodruff, J. B.; Wueseke, O.; Viscardi, V.; Mahamid, J.; Ochoa, S. D.; Bunkenborg, J.; Widlund, P. O.; Pozniakovsky, A.; Zanin, E.; Bahmanyar, S. et al. Regulated assembly of a supramolecular centrosome scaffold in vitro. *Science* **2015**, *348*, 808–812.

(6) Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **2015**, *163*, 123–133.

(7) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F. et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483*, 336–340.

(8) Han, T. W.; Kato, M.; Xie, S.; Wu, L. C.; Mirzaei, H.; Pei, J.; Chen, M.; Xie, Y.; Allen, J.; Xiao, G. et al. Cell-free Formation of RNA Granules: Bound RNAs Identify Features and Components of Cellular Assemblies. *Cell* **2012**, *149*, 768–779.

(9) Su, X.; Ditlev, J. A.; Hui, E.; Xing, W.; Banjade, S.; Okrut, J.; King, D. S.; Taunton, J.; Rosen, M. K.; Vale, R. D. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **2016**, *352*, 595–599.

(10) Banani, S. F.; Rice, A. M.; Peeples, W. B.; Lin, Y.; Jain, S.; Parker, R.; Rosen, M. K. Compositional Control of Phase-Separated Cellular Bodies. *Cell* **2016**, *166*, 651–663.

(11) Mitrea, D. M.; Cika, J. A.; Stanley, C. B.; Nourse, A.; Onuchic, P. L.; Banerjee, P. R.; Phillips, A. H.; Park, C.-G. G.; Deniz, A. A.; Kriwacki, R. W. Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. *Nature Communications* **2018**, *9*, 842.

(12) Strom, A. R.; Emelyanov, A. V.; Mir, M.; Fyodorov, D. V.; Darzacq, X.; Karpen, G. H. Phase separation drives heterochromatin domain formation. *Nature* **2017**, *547*, 241–245.

(13) Larson, A. G.; Elnatan, D.; Keenen, M. M.; Trnka, M. J.; Johnston, J. B.; Burlingame, A. L.; Agard, D. A.; Redding, S.; Narlikar, G. J. Liquid droplet formation by HP1 suggests a role for phase separation in heterochromatin. *Nature* **2017**, *547*, 236–240.

(14) Kim, T. H.; Tsang, B.; Vernon, R. M.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* **2019**, *365*, 825–829.

(15) Riback, J. A.; Zhu, L.; Ferrolino, M. C.; Tolbert, M.; Mitrea, D. M.; Sanders, D. W.; Wei, M.-T.; Kriwacki, R. W.; Brangwynne, C. P. Composition-dependent thermodynamics of intracellular phase separation. *Nature* **2020**, 1–6.

(16) Qian, D.; Acker, J.; Scrutton, R. M.; Fischer, C. M.; Qamar, S.; Sneideris, T.; Borodavka, A.; Knowles, T. P. J. Scaling relations of multicomponent phase coexistence boundaries. 2025; `http://biorxiv.org/lookup/doi/10.1101/2025.04.02.645070`.

(17) Chattaraj, A.; Blinov, M. L.; Loew, L. M. The solubility product extends the buffering concept to heterotypic biomolecular condensates. *eLife* **2021**, *10*.

(18) Chattaraj, A.; Loew, L. M. The maximum solubility product marks the threshold for condensation of multivalent biomolecules. *Biophysical Journal* **2023**, *122*, 1678–1690.

(19) Chattaraj, A.; Baltaci, Z.; Chung, S.; Mayer, B. J.; Loew, L. M.; Ditlev, J. A. Measurement of solubility product reveals the interplay of oligomerization and self-association for defining condensate formation. *Molecular Biology of the Cell* **2024**, *35*, 122–123.

(20) Kar, M.; Dar, F.; Welsh, T. J.; Vogel, L. T.; Kühnemuth, R.; Majumdar, A.; Krainer, G.; Franzmann, T. M.; Alberti, S.; Seidel, C. A. et al. Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proceedings of the National Academy of Sciences of the United States of America* **2022**, *119*, e2202222119.

(21) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics* **2020**, *49*, 107–133.

(22) Ditlev, J. A.; Michalski, P. J.; Huber, G.; Rivera, G. M.; Mohler, W. A.; Loew, L. M.; Mayer, B. J. Stoichiometry of Nck-dependent actin polymerization in living cells. *Journal of Cell Biology* **2012**, *197*, 643–658.

(23) Case, L. B.; Zhang, X.; Ditlev, J. A.; Rosen, M. K. Stoichiometry controls activity of phase-separated clusters of actin signaling proteins. *Science* **2019**, *363*, 1093–1097.

(24) Schmit, J. D.; Bouchard, J. J.; Martin, E. W.; Mittag, T. Protein Network Structure Enables Switching between Liquid and Gel States. *Journal of the American Chemical Society* **2020**, *142*, 874–883.

(25) Bhandari, K.; Cotten, M. A.; Kim, J.; Rosen, M. K.; Schmit, J. D. Structure–Function Properties in Disordered Condensates. *The Journal of Physical Chemistry B* **2021**, *125*, 467–476.

(26) Rosenzweig, E. S. F.; Xu, B.; Cuellar, L. K.; Martinez-Sanchez, A.; Schaffer, M.; Strauss, M.; Cartwright, H. N.; Ronceray, P.; Plitzko, J. M.; Förster, F. et al. The Eukaryotic CO2-Concentrating Organelle Is Liquid-like and Exhibits Dynamic Reorganization. *Cell* **2017**, *171*, 148–162.e19.

(27) Xu, B.; He, G.; Weiner, B. G.; Ronceray, P.; Meir, Y.; Jonikas, M. C.; Wingreen, N. S. Rigidity enhances a magic-number effect in polymer phase separation. *Nature Communications* **2020**, *11*, 1561.

(28) Zhang, Y.; Xu, B.; Weiner, B. G.; Meir, Y.; Wingreen, N. S. Decoding the physical principles of two-component biomolecular phase separation. *eLife* **2021**, *10*.

(29) He, G.; GrandPre, T.; Wilson, H.; Zhang, Y.; Jonikas, M. C.; Wingreen, N. S.; Wang, Q. Phase-separating pyrenoid proteins form complexes in the dilute phase. *Communications Biology* **2023**, *6*, 19.

(30) Banerjee, P. R.; Milin, A. N.; Moosa, M. M.; Onuchic, P. L.; Deniz, A. A. Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. *Angewandte Chemie International Edition* **2017**, *56*, 11354–11359.

(31) Portz, B.; Shorter, J. Biochemical Timekeeping Via Reentrant Phase Transitions. *Journal of Molecular Biology* **2021**, *433*, 166794.

(32) Arter, W. E.; Qi, R.; Erkamp, N. A.; Krainer, G.; Didi, K.; Welsh, T. J.; Acker, J.; Nixon-Abell, J.; Qamar, S.; Guillén-Boixet, J. et al. Biomolecular condensate phase diagrams with a combinatorial microdroplet platform. *Nature Communications 2022 13:1* **2022**, *13*, 1–10.

(33) Nehm, S. J.; Rodríguez-Spong, B.; Rodríguez-Hornedo, N. Phase solubility diagrams of cocrystals are explained by solubility product and solution complexation. *Crystal Growth and Design* **2006**, *6*, 592–600.

(34) Chattaraj, A.; Youngstrom, M.; Loew, L. M. The Interplay of Structural and Cellular Biophysics Controls Clustering of Multivalent Molecules. *Biophysical Journal* **2019**,

(35) Dill, K. A.; Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*; Garland Science, 2002; p 686.

(36) Grandpre, T.; Zhang, Y.; Pyo, A. G. T.; Weiner, B.; Li, J.-L.; Jonikas, M. C.; Wingreen, N. S. Impact of Linker Length on Biomolecular Condensate Formation. *PRX LIFE* **2023**, *1*, 23013.

(37) Qian, D.; Welsh, T. J.; Erkamp, N. A.; Qamar, S.; Nixon-Abell, J.; Krainer, G.; George-Hyslop, P. S.; Michaels, T. C.; Knowles, T. P. Tie-Line Analysis Reveals Interactions Driving Heteromolecular Condensate Formation. *Physical Review X* **2022**, *12*, 041038.

(38) Michalski, P. J.; Loew, L. M. SpringSaLaD: A Spatial, Particle-Based Biochemical Simulation Platform with Excluded Volume. *Biophysical Journal* **2016**, *110*, 523–529.

(39) Jawerth, L. M.; Ijavi, M.; Ruer, M.; Saha, S.; Jahnel, M.; Hyman, A. A.; Jülicher, F.; Fischer-Friedrich, E. Salt-Dependent Rheology and Surface Tension of Protein Condensates Using Optical Traps. *Physical Review Letters* **2018**, *121*, 258101.

(40) Seal, M.; Jash, C.; Jacob, R. S.; Feintuch, A.; Harel, Y. S.; Albeck, S.; Unger, T.; Goldfarb, D. Evolution of CPEB4 Dynamics across its Liquid-Liquid Phase Separation Transition. *Journal of Physical Chemistry B* **2021**, *125*, 12947–12957.

(41) Golani, G.; Seal, M.; Kar, M.; Hyman, A. A.; Goldfarb, D.; Safran, S. Mesoscale properties of protein clusters determine the size and nature of liquid-liquid phase separation (LLPS). *Communications Physics* **2025**, *8*, 1–10.

(42) Schmit, J. D.; Feric, M.; Dundr, M. How Hierarchical Interactions Make Membraneless Organelles Tick Like Clockwork. *Trends in Biochemical Sciences* **2021**, *46*, 525–534.

(43) Forman-Kay, J. D.; Ditlev, J. A.; Nosella, M. L.; Lee, H. O. What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates? *RNA* **2022**, *28*, 36–47.

(44) Zwicker, D.; Paulin, O. W.; ter Burg, C. Physics of droplet regulation in biological cells. **2025**,

(45) Ditlev, J. A.; Case, L. B.; Rosen, M. K. Who's In and Who's Out—Compositional Control of Biomolecular Condensates. *Journal of Molecular Biology* **2018**, *430*, 4666–4684.

(46) Jawerth, L.; Fischer-Friedrich, E.; Saha, S.; Wang, J.; Franzmann, T.; Zhang, X.; Sachweh, J.; Ruer, M.; Ijavi, M.; Saha, S. et al. Protein condensates as aging Maxwell fluids. *Science* **2020**, *370*, 1317–1323.

(47) Acker, J.; Wang, X.; Desirò, D.; Agarwal, T.; Colyer, A.; Haller, C.; Scrutton, R.; Sherry, L.; Saar, K. L.; Murray, R. et al. Phosphorylation enables allosteric control of a viral condensate. *bioRxiv* **2025**, 2025.05.24.655949.

(48) Seim, I.; Posey, A. E.; Snead, W. T.; Stormo, B. M.; Klotsa, D.; Pappu, R. V.; Gladfelter, A. S. Dilute phase oligomerization can oppose phase separation and modulate material properties of a ribonucleoprotein condensate. *Proceedings of the National Academy of Sciences* **2022**, *119*, 2021.04.19.440511.