Analytical Solution to the Flory–Huggins Model

Daoyuan Qian, Thomas C. T. Michaels,* and Tuomas P. J. Knowles*

ABSTRACT: A self-consistent analytical solution for binodal concentrations of the two-component Flory–Huggins phase separation model is derived. We show that this form extends the validity of the Ginzburg–Landau expansion away from the critical point to cover the whole phase space. Furthermore, this analytical solution reveals an exponential scaling law of the dilute phase binodal concentration as a function of the interaction strength and chain length. We demonstrate explicitly the power of this approach by fitting experimental protein liquid–liquid phase separation boundaries to determine the effective chain length and solute–solvent interaction energies. Moreover, we demonstrate that this strategy allows us to resolve differences in interaction energy contributions of individual amino acids. This analytical framework can serve as a new way to decode the protein sequence grammar for liquid–liquid phase separation.

The formation of membrane-less organelles through liquid–liquid phase separation (LLPS) has emerged as an important mechanism used by cells to regulate their internal biochemical environments, and it is also closely related to the development of neurodegenerative diseases.1–8 The Flory–Huggins model9–11 is a foundational theoretical picture that describes the phenomenology of LLPS, driven by a competition of entropy and interaction energy. Despite the generality of the Flory–Huggins model, analytical solutions describing the binodal line have not been available, and progress has instead been made through numerical methods.11–13 Here, we propose an analytical self-consistent form for the binodal concentrations and demonstrate the high accuracy comparable to numerical schemes. This can then be used to efficiently fit experimental binodal data and determine key underlying physical parameters.

The two-component Flory–Huggins theory describes mixing of a polymer species of length \( N \) with a homogeneous solvent. Denoting the volume fraction of polymers as \( \phi \), the volume fraction of the solvent is simply \( 1 - \phi \) by volume conservation. The model uses an effective lattice site contact energy \( \chi \equiv \frac{z}{2} \epsilon_{s} - \epsilon_{ps} - \epsilon_{pp} \) between the polymer and solvent in which \( z \) is a coordination constant, and \( \epsilon_{ps} \), \( \epsilon_{s} \), and \( \epsilon_{pp} \) are bare polymer–solvent, solvent–solvent, and polymer–polymer contact energies. The free energy density of the Flory–Huggins model is given by9–12

\[
\frac{f(\phi)}{k_{B}T} = \frac{\phi}{N} \ln \phi + (1 - \phi) \ln(1 - \phi) + \chi \phi (1 - \phi)
\]

(1)

where \( k_{B} \) is the Boltzmann constant and \( T \) is the absolute temperature. In the following, we consider energies relative to the thermal energy and set \( k_{B}T = 1 \) to simplify notation. The first two terms on the right-hand side of eq 1 represent the entropic free energy of mixing, while the third term denotes the effective contact energy. Two important quantities can be calculated: the spinodal concentration and binodal concentration. The spinodal is the boundary between locally stable/unstable regions and can be solved exactly, while the binodal separates globally stable/unstable regions, and the system can still be locally stable on the binodal boundary itself. It is also straightforward to generalize eq 1 to include more components or surface tension,11,14,15 and over the decades more detailed models have been proposed to include electrostatic interactions16 and sticker-spacer behaviors17,18 or to calculate free energy density from first principles using a field-theoretic approach.19–21 It thus appears that the Flory–Huggins theory is an outdated model due to its oversimplifying, mean-field nature, while we note that even then an analytical solution for the binodal concentrations is lacking for this most basic picture of LLPS.

Mathematical formulations of spinodal and binodal concentrations are briefly summarized here. The free energy density becomes locally unstable at \( f'(\phi) \leq 0 \), and consequently the spinodal boundary \( \phi^{sp} \) is defined at the transition point \( f'(\phi^{sp}) = 0 \). Solving for this condition, we obtain the dense \( (\phi^{sp}) \) and dilute phase \( (\phi^{dp}) \) spinodal concentrations.

Received: June 27, 2022
Accepted: August 5, 2022
Published: August 17, 2022
and these are qualitatively different from the exponential scaling of the dilute phase binodal concentrations, as will be shown using the self-consistent equations.

Next, we outline the steps to obtain binodal concentrations. The binodal concentrations are found by assuming the existence of two distinct phases characterized by polymer volume fractions \( \phi_+ \) and \( \phi_- \). The equilibrium condition requires minimization of the total energy \( F_{\text{tot}} \equiv V_+ f_+(\phi_+) + V_- f_-(\phi_-) \) subject to total volume and mass conservation conditions \( V_+ + V_- = V_{\text{tot}} \) and \( V_+\phi_+ + V_-\phi_- = V_{\text{tot}}\phi_{\text{tot}} \). Using Lagrange minimization, we identify the chemical potential \( \mu(\phi) \equiv f'(\phi) + \Pi(\phi) \) and osmotic pressure \( \Pi(\phi) \equiv \phi f'(\phi) - f(\phi) \) as Lagrange multipliers that have to hold the same values in the two compartments:

\[
\left\{ \begin{array}{l}
\mu(\phi_+) = \mu(\phi_-) \\
\Pi(\phi_+) = \Pi(\phi_-)
\end{array} \right.
\]

Graphically, in a \( [\phi f(\phi)] \) plot, the \( \mu(\phi_+) = \mu(\phi_-) \) condition forces the two points describing the coexisting phases to have the same gradient, and \( \Pi(\phi_+) = \Pi(\phi_-) \) aligns the two tangent lines to have the same \( y \)-intercept and as such represent a common tangent construction (Figure 1A). Using the Flory–Huggins free energy eq 1, we have

\[
\phi_{\text{eq}}^{(0)} = \left( \frac{1}{2} - \frac{\gamma}{4\chi} \right) \pm \sqrt{\left( \frac{1}{2} - \frac{\gamma}{4\chi} \right)^2 - \frac{1}{2\chi N}} \tag{2}
\]

with \( \gamma \equiv 1 - \frac{1}{N} \), which goes to zero in the symmetric \( N = 1 \) case. The critical point of LLPS occurs when the dense and dilute phases coincide, corresponding to a critical interaction strength \( \chi_c \), and concentration \( \phi_c \):

\[
\chi_c = \frac{1}{2} \left( 1 + \frac{1}{\sqrt{N}} \right)^2, \quad \phi_c = \frac{1}{1 + \sqrt{N}} \tag{3}
\]

Near the critical point with \( \delta\chi \equiv \chi - \chi_c \approx 0 \), the spinodal concentrations have the approximate form

\[
\phi_{\text{eq}}^{(0)} = \phi_c \pm \sqrt{\frac{\delta\chi}{2\chi_c^2\sqrt{N}}} + O(\delta\chi) \tag{4}
\]

Note that in the opposite limit of large \( N \) or large \( \chi \) the dilute phase concentration has a power-law scaling

\[
\phi_{\text{eq}}^{(0)} \approx \frac{1}{(2\chi - 1)N} \quad \text{for } N \gg 1
\]

\[
\frac{1}{2\chi N} \quad \text{for } \chi \gg 1 \tag{5}
\]

Figure 1. Flory–Huggins model (A–C) and self-consistent solution for the symmetric \( N = 1 \) case (D–F). (A) Common tangent construction at \( N = 3, \chi = 1.5 \) gives the dense and dilute phase concentrations \( \phi_{\pm} \). The gradient of the common tangent is the chemical potential \( \mu(\phi) \), and the \( y \)-intercept is \( -1 \) times the osmotic pressure \( \Pi(\phi) \). (B, C) Complete phase diagram of the \( N = 3 \) system in linear and logarithmic \( \phi \) scales. The binodal is calculated numerically. Near the critical point, the Ginzburb–Landau binodal approximates the exact binodal well, but at large \( \chi \) the two quickly diverges and the Ginzburb–Landau solution enters the unphysical range \( \phi < 0 \) and \( \phi > 1 \) (gray zones). (D) Plot of \( |g'|(\phi)| = 1 \) in \( \phi, \chi \) space. The black solid line is binodal, and the hollow circle is the critical point. Dashed lines are contours of constant \( |g'|(\phi)| = 1 \). The blue region with \( |g'|(\phi)| < 1 \) has stable orbits, while the red regions are unstable. (E, F) Comparison between the numerical binodal (black solid line) with self-consistent schemes with 0, 1, and 2 iterations (colored dashed lines).
The Ginzburg–Landau solution describes the binodal near the critical point (Figure 1B,C). We now aim to extend this solution to cover χ far away from χc, through a self-consistent approach, summarized as the following. Suppose we need to solve the equation η = ℬ(η) with some operator ℬ. Instead of solving directly, we treat ℬ as a discrete map and start with a solution η(0) and apply ℬ iteratively to generate the orbit η(i) = ℬ[η(i−1)] with i = 1, 2, ... With a suitable form of ℬ, the orbit converges to the fixed point limi→∞ η(i) = η, which then solves the equation η = ℬ(η). This is the contractive mapping principle. The self-consistent approach has been previously employed to approximate the protein aggregation kinetics curves and here we show a similar procedure allows efficient and accurate computation of the binodal concentrations.

Starting with the simple case of unit polymer length N = 1, the free energy (eq 1) is invariant under a reflection around Φ = ±1/2, i.e., Φ → −1−Φ, and the binodal is given by the condition f′(Φ) = 0, leading to 1−Φ = exp(−2Φ + χ). Upon rearrangement, the binodal equation is

\[ \Phi = 1 \left/ \left[ 1 + \exp(-2\Phi + \chi) \right] \right. \equiv g(\Phi) \]

and we use g(Φ) to define the 1D map

\[ \Phi^{(i)} = g^{(i)}(\Phi^{(i−1)}) \]

with the initial guess Φ(0) to be determined later. The fixed points are the binodal concentrations Φ±. To study the convergent properties of the map, we expand g(Φ) near the fixed point, writing Φ± = Φ± + δΦ± + O(δΦ±2). This gives δΦ± = g′(Φ±)δΦ±, so for a convergent orbit we require |g′(Φ±)| < 1, and quick convergence can be expected for |g′(Φ±)| ≈ 0. We calculate |g′(Φ)| = 1 in the φ,χ space and observe that near the binodal convergence is fast in the high-χ regime with |g′(Φ)| ≈ 0, while it is much slower near criticality χ ≈ χc ≈ 2 and becomes 0 at exactly the critical point (Figure 1D). We thus need the initial guess to be close to the binodal just at χ ≈ χc and both the spinodal and approximate binodal may seem to be appropriate choices. It is worth writing down these
In the case of \( N = 1 \), we recover the \( e^{-x} \) scaling discussed above. The convergence behavior of the 2D map can also be studied similar to the 1D case. Defining the Jacobian matrix \( J \) as

\[
J = \begin{pmatrix}
\frac{\partial G_i}{\partial \phi_i} & \frac{\partial G_i}{\partial \phi_-} \\
\frac{\partial G_+}{\partial \phi_i} & \frac{\partial G_+}{\partial \phi_-}
\end{pmatrix}
\]

and writing \( \phi^{(i)} = \phi^{\text{bin}} + \delta \phi^{(i)} \), we get \( \delta \phi^{(i+1)} = J \delta \phi^{(i)} \). Stability requires moduli of eigenvalues of \( J \) to be less than 1. Since both the eigenvalues and eigenvectors of \( J \) are in general complex, to better visualize the convergence of the orbit \( \phi^{(i)} \) we instead promote the discrete map to a continuous flow equation parametrized by \( t \): \( \dot{\phi} = \mathbf{G}(\phi) - \phi \). The velocity field \( \dot{\phi} \) then contains the behavior of the orbit \( \phi^{(i)} \) in the limit of small time steps, and three fixed points can be identified: one stable fixed point corresponding to the binodal and two unstable fixed points on the \( \phi_i = \phi_- \) diagonal (Figure 2A). We observe the orbit \( \phi^{(i)} \) is indeed convergent (Figure 2B,C).

To obtain analytical forms for the general binodal, we first simplify notations by defining

\[
\alpha \equiv N^{1/4},
\]

\[
\Delta = \frac{\chi - \chi_c}{\chi_c},
\]

and express all other parameters in terms of \( \alpha, \Delta \). This allows us to write

\[
\gamma = 1 - \frac{1}{N} = \frac{4}{\alpha^2} \sinh \ln \alpha \cosh \ln \alpha
\]

\[
\phi_i = \frac{1}{1 + \sqrt{N}} = \frac{1}{2\alpha \cosh \ln \alpha}
\]

\[
\chi_c = \frac{1}{2} \left( 1 + \frac{1}{\sqrt{N}} \right)^2 = \frac{2}{\alpha^2} \cosh^2 \ln \alpha
\]

The Ginzburg–Landau solution (eq 21) then takes the simple form

\[
\phi^\text{bin} \approx e^{-N(x-y)}
\]
\[ \phi^{(0)}_\pm = \frac{1 \pm \alpha \sqrt{3\Delta}}{2\alpha \cosh \ln \alpha} \]  
\[ \text{and } x, y \text{ as defined in eq 17 are now} \]
\[ x = \frac{4 \cosh^2 \ln \alpha}{\alpha^2} (1 + \Delta)(\phi_+ - \phi_-) \]
\[ y = \frac{4 \sinh \ln \alpha \cosh \alpha}{\alpha^2} (\phi_+ - \phi_-) \]
\[ + \frac{2 \cosh^2 \ln \alpha}{\alpha^2} (1 + \Delta)(\phi_+ - \phi_-)(\phi_+ + \phi_-) \]
\[ \text{Direct substitution gives} \]
\[ x^{(1)} = \frac{4 \cosh \ln \alpha}{\alpha^2} (1 + \Delta) \sqrt{3\Delta} \]
\[ y^{(1)} = \frac{2 \sqrt{3\Delta}}{\alpha} (1 + \Delta^2) \]
\[ \text{so at one self-consistent step we have} \]
\[ \phi^{(1)}_\pm = \frac{1 - e^{\mp 2\Delta}}{1 - e^{2\Delta + 2B}} \]
\[ \text{where} \]
\[ A = \frac{1}{\alpha} \left( 1 + \frac{\Delta}{\alpha^2} \right) \sqrt{3\Delta} \]
\[ B = \alpha \left( 1 + \alpha^2 \Delta \right) \sqrt{3\Delta} \]
\[ \text{A and B are related through the transformation } \alpha \rightarrow \frac{1}{\alpha} \text{ as} \]
\[ A_{\alpha \rightarrow 1/\alpha} = B \text{ and vice versa. The large-} \chi \text{ behavior is incompletely captured in log scale (Figure 2C). We thus again calculate the second-order self-consistent solution. At second order, we substitute eq 30 into eq 28 and obtain} \]
\[ x^{(2)} = \frac{8D \cosh \ln \alpha}{\alpha^2} (1 + \Delta) \]
\[ y^{(2)} = \frac{8D \sinh \ln \alpha}{\alpha^2} + \frac{8D \coth B}{\alpha^2} (1 + \Delta) \]
\[ \text{where} \]
\[ D = \frac{\cosh \alpha}{\coth A + \coth B} \]
\[ \text{and D is invariant under the transformation } \alpha \rightarrow \frac{1}{\alpha}. \text{ The second-order expression is then} \]
\[ \phi^{(2)}_\pm = \left\{ 1 - \exp \left[ \mp 8D \left( \frac{\sinh \ln \alpha}{\alpha^2} + D(1 + \Delta) \right) \right] \right\} \]
\[ \left( \frac{\coth B}{\alpha^2} \right) \]
\[ \left\{ 1 - \exp \left[ \pm 8D \left[ \frac{1}{\alpha^2} - \alpha^2 \sinh \ln \alpha + D(1 + \Delta) \right] \right] \right\} \]
\[ \left( \frac{\coth B}{\alpha^2} + \alpha^2 \coth A \right) \}
\[ \text{Notice again that the denominator is invariant under the transformation } \alpha \rightarrow \frac{1}{\alpha}. \text{ The second-order analytical form approximates the exact binodal to a high degree (Figure 2B,C) even at the large-N regime.} \]

Although the self-consistent solutions are exact near critical points and convergent at large \( \chi \), the convergence is slow in the transition region. Here we show that we can improve the maps by performing a first-order expansion of the self-consistent operator. Starting from a general self-consistent equation \( \eta = \mathcal{A}(\eta) \) and an initial guess \( \eta^{(0)} \), we want to find a step \( \delta \eta \) such that the next guess \( \eta^{(1)} \equiv \eta^{(0)} + \delta \eta \) solves the self-consistent equation to first order. We thus write \( \delta \eta = \frac{\eta^{(0)} - \mathcal{A}(\eta^{(0)})}{\mathcal{A}(\eta^{(0)}) - 1} \). Expanding \( \mathcal{A}(\eta^{(0)}) + \delta \eta \) to first order, and solving for \( \delta \eta \) we obtain
\[ \delta \eta = \frac{\eta^{(0)} - \mathcal{A}(\eta^{(0)})}{\mathcal{A}(\eta^{(0)}) - 1} \]
\[ \text{so the next best guess is} \]
\[ \eta^{(1)} = \frac{\mathcal{A}(\eta^{(0)}) \eta^{(0)} - \mathcal{A}(\eta^{(0)})}{\mathcal{A}(\eta^{(0)}) - 1} \]

The above results can readily be applied to improve the maps \( g(\phi) \) and \( G(\phi) \). In the \( N = 1 \) case, we define the new map \( h(\phi) \) as
\[ h(\phi) \equiv \frac{g'(\phi) - g(\phi)}{g'(\phi) - 1} \]

and it reduces to the original map when \( g'(\phi) = 0 \). We thus define the new orbit \( \phi^{(1)}_h \equiv h'(\phi^{(0)}_h) \). The convergent property can be studied by expanding the above with \( \phi^{(i+1)}_h = \phi^{(i)}_h + \delta \phi^{(i)}_h \), and we arrive at
\[ \delta \phi^{(i+1)}_h = \frac{g'(\phi^{(i)}) [g'(\phi^{(i)}) - \phi^{(i)}]}{[g'(\phi^{(i)}) - 1]^2} \]
\[ \delta \phi^{(i)}_h \]

and near the fixed point the numerator approaches 0, so the convergence is rapid (Figure 3A). In the case of general \( N \), we similarly obtain
\[ \delta \phi = (J - 1)^{-1} [\phi - G(\phi)] \]

with \( 1 \) a 2 by 2 identity matrix. The improved operator is

![Figure 3](https://doi.org/10.1021/acs.jpcl.2c01986)

**Figure 3.** Improved self-consistent solutions from self-consistent expansion (orange dashed lines) converge to the numerical solution (black solid lines) more quickly than the original ones (blue dashed lines), in both the (A) symmetric \( N = 1 \) case and (B) the general \( N \) case. The numerical binodal virtually overlaps with the improved solutions.
Figure 4. Flory–Huggins fit of binodal data from ref 27, fitted using eq 34. A constant N is maintained across all variants. Solid triangle markers are dilute and dense phase concentration measurements, and light hollow triangle markers are estimates of the critical point from cloud point measurements. Dashed lines are the best-fit curves, and hollow circles are critical points. The WT binodal is the same in all three plots and is plotted in the solid line. Colors of the plot represent the $E \equiv -N\Delta \epsilon$ values of the variant. The $\pm nX$ in variant names indicate $n$ of $X$ residues are added (+) or removed (−) from WT.

Table 1. Fitting Results for the A1-LCD Data$^a$

| variant    | aromatic series $\Delta E$ (kJ/mol) | polar series $\Delta E$ (kJ/mol) | ionic series $\Delta E$ (kJ/mol) |
|------------|--------------------------------------|----------------------------------|----------------------------------|
| −12F +12Y  | −5.5                                 | +23G −23S                        | +7R +12D                         |
| −7F −7Y    | +7.4                                 | −10G +10S                        | +7K +12D                         |
| −4F −2Y    | +22.0                                | −20G +20S                        | +12D                             |

$^a$Differences in $\Delta E$ across variants allow contributions of individual residues to be inferred.

$H(\phi) \equiv (J - 1)^{-1}[J\phi - G(\phi)]$  \hspace{1cm} (40)

Good agreement with numerical results is achieved for the new orbit $\phi_{i}^{(N)} \equiv H(\phi_{i}^{(N)})$ within three iterations for a large $N = 100$ (Figure 3B).

The self-consistent solution allows efficient computation of binodal concentrations, and we use it to fit experimental LLPS data and extract the interaction parameters. In the following fitting, we use eq 34 to compute the binodal concentrations. Binodal concentrations for the prion-like low-complexity (WT) A1-LCD binodal was also measured. During fitting, the second series involves nonequivalent polar spacers glycine (G) and serine (S); and the third series involves ionic residues aspartic acid (D), arginine (R), and lysine (K). The wild type (WT) A1-LCD binodal was also measured. During fitting, the chain length $N$ is set as a global fitting parameter. The interaction parameter $\chi$ has the form $\chi = \frac{\Delta \epsilon}{kT}$, and we fit $\Delta \epsilon$ for each variant. To convert concentrations to volume fractions, we use a protein density of 1.35 g/cm$^3$ and average molecular weight of 13.1 kDa to obtain the conversion ratio from concentration $c$ (M) to volume fraction $\phi$ as $\phi = \frac{13.1 c}{1.35 M}$.

The fitting results give the effective chain length $N = 158.6$, larger than 137, the number of residues. These results can appear counterintuitive, as past studies have postulated an effective protein segment length larger than the size of an amino acid,\(^{31,29}\) so the effective $N$ should be smaller than the number of residues. The discrepancy arises from a subtle difference in the definition of $N$ in the Flory–Huggins picture as compared to the polymer picture: here the fitted $N$ represents the number of lattice sites occupied by each solute and does not depend on its polymeric nature. As a result, the same $N$ can be defined for nonpolymer solutes such as micelle clusters, and thus the $N$ estimated here should not in any way relate to the effective segment length of the polymer. In the present case, the lattice site volume is determined by the underlying medium, i.e., water, so the effective $N$ will be larger than the number of residues owing to the larger size of amino acids compared to water molecules. The ratio $\eta \equiv \frac{N}{137} = 1.16$ then represents the average number of lattice sites occupied by one residue. The fitted $\Delta \epsilon$ values represent the site-to-site contact energy, and a larger $\Delta \epsilon$ indicates a stronger attraction between proteins. To highlight the difference across variants, we first calculate the protein-to-protein contact energy $E \equiv -N\Delta \epsilon$ and define the deviation from WT as $\Delta E_{\text{variant}} \equiv E_{\text{variant}} - E_{\text{WT}}$. Fitted curves are plotted in Figure 4, and $\Delta E$ results are listed in Table 1. Each variant series then allows quantitative interpretation of impacts of different residues on LLPS propensity. In principle, the effective interaction energy $E_{\text{variant}}$ is a function of the whole amino acid sequence that depends on both the composition and arrangement of individual residues. For example, relating the detailed sequence information to the effective interaction energy has been achieved for polyelectrolytes through the sequence charge decoration (SCD) parameter\(^{30}\) with pairwise binding constants and second virial coefficient expressed in terms of SCD.\(^{31}\) Finding this function for a generic protein can be hard,
although machine-learning techniques could potentially be used with enough protein sequences and corresponding binodal data. In the present case, only limited data are available, so we assume a simple, linear functional form of $E_{\text{variant}}$ to illustrate the utility of the self-consistent solution. To this end, we simply assume $E_{\text{variant}} = E_0 + \sum n_i \Delta E_i$, with $E_0$ a constant and $n_i \Delta E_i$ the number and effective contribution of the amino acid residue $i$. This then allows us to construct linear simultaneous equations from the fitted values (Table 1) and quantify the energetic contribution of individual residues.

Results from the aromatic series indicate that tyrosine is a stronger sticker than phenylalanine, in line with previous observations.\textsuperscript{27} We further infer the individual contribution of each Tyr and Phe residue, $\Delta E_{\text{Tyr}}$ and $\Delta E_{\text{Phe}}$, using values from Table 1:

$$
\begin{cases}
-12 \Delta E_{\text{Phe}} + 12 \Delta E_{\text{Tyr}} = -5.5 \text{ kJ/mol} \\
+7 \Delta E_{\text{Phe}} - 7 \Delta E_{\text{Tyr}} = +7.4 \text{ kJ/mol} \\
-4 \Delta E_{\text{Phe}} - 2 \Delta E_{\text{Tyr}} = +22.0 \text{ kJ/mol}
\end{cases}
$$

(41)

The first two equations give $\Delta E_{\text{Phe}} - \Delta E_{\text{Tyr}} = 0.8 \pm 0.3 \text{ kJ/mol}$, with the error arising from the difference in the measured per-residue energy change. This can then be combined with the third equation to give $\Delta E_{\text{Tyr}} = -4.2 \pm 0.2 \text{ kJ/mol}$ and $\Delta E_{\text{Phe}} = -3.4 \pm 0.1 \text{ kJ/mol}$. Both Tyr and Phe are thus stickers with Tyr stronger than Phe.

For the polar series, we perform similar calculations and extract the difference $\Delta E_{\text{Gly}} - \Delta E_{\text{Ser}} = -0.43 \pm 0.11 \text{ kJ/mol}$, indicating the destabilizing effect of serine residues. This can be understood as the OH group in Ser forming favorable interactions with water, thus destabilizing the condensate.

The ionic series data are harder to interpret since the overall protein charge can have a non-monotonic effect on LLPS propensity.\textsuperscript{27} We can however still compare the $+7$R $+12$D and $+7$K $+12$D variants since they have the same overall charge. The energy difference between arginine and lysine is $\Delta E_{\text{Arg}} - \Delta E_{\text{Lys}} = -3.4 \text{ kJ/mol}$, indicating stronger sticker behavior for arginine. This can arise due to the electron delocalization in the guanidinium group and higher charge–charge contact efficiency with other charged residues, or stronger cation–π interaction from coplanar packing.\textsuperscript{32,33} Furthermore, it should be noted that despite the simple linear functional form of $E_{\text{variant}}$ assumed here, the resulting energy difference between Arg and Lys is in line with atomistic simulation results: using the Kim-Hummer model,\textsuperscript{34,35} the difference in average residue–residue pair interaction involving either Arg or Lys is estimated to be $(1.48 - 2.22)k_B T = -0.74k_B T \approx -1.84 \text{ kJ/mol}$.\textsuperscript{32} This is roughly half of $-3.4 \text{ kJ/mol}$ as estimated from LLPS data, and a probable reason is that one Arg or Lys residue might be involved in more than one residue–residue contact, giving a higher overall contribution to the contact energy.

To conclude, we have developed a self-consistent solution for the binodal concentration of the two-component Flory–Huggins phase-separating system. The proposed self-consistent operators shed light on the scaling behavior of the dilute phase binodal, which is qualitatively different from the scaling of the spinodal and explains why LLPS of proteins occurs over a concentration range spanning several orders of magnitude. Using the well-known Ginzburg–Landau binodal approximate solution as the initial guess, the self-consistent solution achieves numerical accuracy within two to three iterations and allows highly efficient fitting of experimental binodal data. Explicit analytical forms of the binodal concentrations are also proposed to approximate the binodal. Using the developed solution, we fitted experimental data measured for variants of the A1-LCD protein and extracted effective interaction energies, which can be used to further decode the impact of individual amino acid residues on LLPS. Our analytical solution to the Flory–Huggins model thus allows systematic investigation of sequence grammar of LLPS-prone proteins and, with sufficient experimental data, can lead to development of a wholistic framework for predicting LLPS propensity from sequence information.

### ASSOCIATED CONTENT

#### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.2c01986.

Transparent Peer Review report available (PDF)

### AUTHOR INFORMATION

**Corresponding Authors**

Thomas C. T. Michaels — Department of Physics and Astronomy, Institute for the Physics of Living Systems and Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, U.K.; Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.; Email: t.michaels@ucl.ac.uk

Tuomas P. J. Knowles — Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.; Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, U.K.; orcid.org/0000-0002-7879-0140; Email: tpjk2@cam.ac.uk

**Author**

Daoyuan Qian — Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.; orcid.org/0000-0002-8539-3346

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acs.jpclett.2c01986

**Notes**

The authors declare no competing financial interest.

Code availability: A python library containing the self-consistent maps and auxiliary functions can be found at: https://github.com/KnowlesLab-Cambridge/FloryHuggins.

**ACKNOWLEDGMENTS**

This study is supported by the Laboratory for Molecular Cell Biology, University College London (T.C.T.M.), the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007-2013) through the ERC grant PhysProt (Agreement No. 337969) (T.P.J.K.), the BBSRC (T.P.J.K.), the Newman Foundation (T.P.J.K.), and the Wellcome Trust Collaborative Award 203249/Z/16/Z (T.P.J.K.). The authors thank Alexander Röntgen for helpful discussions.
REFERENCES

(1) Alberti, S.; Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. Nat. Rev. Mol. Cell Biol. 2021, 22, 196–213.

(2) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. Annu. Rev. Phys. Chem. 2020, 71, 53–75.

(3) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. 2017, 18, 285–298.

(4) Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. Science 2017, 357. DOI: 10.1126/science.aaf4382

(5) Babincak, W. M.; Sureauwicz, W. K. Liquid-Liquid Phase Separation and Its Mechanistic Role in Pathological Protein Aggregation. J. Mol. Biol. 2020, 432, 1910–1925.

(6) Söding, J.; Zwicker, D.; Sohrabi-Jahromi, S.; Boehning, M.; Kirschbaum, J. Mechanisms for Active Regulation of Biomolecular Condensates. Trends in Cell Biology 2020, 30, 4–14.

(7) 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.

(8) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer physics of intracellular phase transitions. Nat. Phys. 2015, 11, 899–904.

(9) Florý, P. J. Thermodynamics of High Polymer Solutions. J. Chem. Phys. 1942, 10, 51–61.

(10) De Gennes, P.-G.; Gennes, P.-G. Scaling Concepts in Polymer Physics; Cornell University Press, 1979.

(11) Mao, S.; Kulidinow, D.; Haataja, M. P.; Košmrlj, A. Phase behavior and morphology of multicompontent liquid mixtures. Soft Matter 2019, 15, 1297–1311.

(12) Lin, Y.-H.; Wessén, J.; Pal, T.; Das, S.; Chan, H. S. Numerical Techniques for Applications of Analytical Theories to Sequence-Dependent Phase Separations of Intrinsically Disordered Proteins. arXiv 2022, arXiv:2201.01920.

(13) Ariono, D.; Aryanti, P. T.; Hakim, A. N.; Subagio, S.; Wenthen, I. G. Determination of thermodynamic properties of polysulfone/PEG membrane solutions based on Flory-Huggins model. AIP Conf. Proc. 2017, 1840, 090008.

(14) Deviri, D.; Safran, S. A. Physical theory of biological noise buffering by multicomponent phase separation. Proc. Natl. Acad. Sci. U.S.A. 2021, 118, 1–34.

(15) Berry, J.; Brangwynne, C. P.; Haataja, M. Physical principles of intracellular organization via active and passive phase transitions. Rep. Prog. Phys. 2018, 81, 046601.

(16) Overbeek, J. T. G.; Voorn, M. J. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. Journal of Cellular and Comparative Physiology 1957, 49, 7–26.

(17) semenov, A. N.; Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. Macromolecules 1998, 31, 1373–1385.

(18) Wessén, J.; Pal, T.; Das, S.; Lin, Y. H.; Chan, H. S. A simple explicit-solvent model of polyanlypholyte phase behaviors and its ramifications for dielectric effects in biomolecular condensates. J. Phys. Chem. B 2021, 125, 4337–4358.

(19) McCarty, J.; Delaney, K. T.; Danielsen, S. P.; Fredrickson, G. H.; Shea, J. E. Complete Phase Diagram for Liquid-Liquid Phase Separation of Intrinsically Disordered Proteins. J. Phys. Chem. Lett. 2019, 10, 1644–1652.

(20) Lin, Y. H.; Song, J.; Forman-Kay, J. D.; Chan, H. S. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. J. Mol. Liq. 2017, 228, 176–193.

(21) Chen, G. P.; Voora, V. K.; Agee, M. M.; Balasubramani, S. G.; Furche, F. Random-Phase Approximation Methods. Annu. Rev. Phys. Chem. 2017, 68, 421–445.

(22) Granas, A.; Dugundji, J. Fixed Point Theory; Springer Monographs in Mathematics; Springer New York: New York, NY, 2003.

(23) Zeidler, E. Nonlinear Functional Analysis and its Applications: Part 1: Fixed-Point Theorems; Springer: New York, 1985.

(24) Cohen, S. I.; Vendruscolo, M.; Welland, M. E.; Dobson, C. M.; Terentjev, E. M.; Knowles, T. P. Nucleated polymerization with secondary pathways. I. Time evolution of the principal moments. J. Chem. Phys. 2011, 135, 065105.

(25) Cohen, S. I.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. Nucleated polymerization with secondary pathways. II. Determination of self-consistent solutions to growth processes described by nonlinear master equations. J. Chem. Phys. 2011, 135, 065106.

(26) Strogatz, S. H. Nonlinear Dynamics and Chaos; CRC Press, 2018.

(27) Bremer, A.; Farag, M.; Borchers, W. M.; Peran, I.; Martin, E. W.; Pappu, R. V.; Mittag, T. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. Nat. Chem. 2022, 14, 196–207.

(28) Dill, K. A. Theory for the folding and stability of globular proteins. Biochemistry 1985, 24, 1501–1509.

(29) Song, J.; Li, J.; Chan, H. S. Small-Angle X-ray Scattering Signatures of Conformational Heterogeneity and Homogeneity of Disordered Protein Ensembles. J. Phys. Chem. B 2021, 125, 6451–6478.

(30) Sawle, L.; Ghosh, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. J. Chem. Phys. 2015, 143, 085101.

(31) Amin, A. N.; Lin, Y. H.; Das, S.; Chan, H. S. Analytical Theory for Sequence-Selective Binary Fuzzy Complexes of Charged Intrinsically Disordered Proteins. J. Phys. Chem. B 2020, 124, 6709–6720.

(32) Das, S.; Lin, Y. H.; Vernon, R. M.; Forman-Kay, J. D.; Chan, H. S. Comparative roles of charge, π, and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U.S.A. 2020, 117, 28795–28805.

(33) Crowley, P. B.; Golovin, A. Cation-π interactions in protein-protein interfaces. Proteins: Struct., Funct., Genet. 2005, 59, 231–239.

(34) Kim, Y. C.; Hummer, G. Coarse-grained Models for Simulations of Multiprotein Complexes: Application to Ubiquitin Binding. J. Mol. Biol. 2008, 375, 1416–1433.

(35) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. PLoS Comput. Biol. 2018, 14, e1005941.