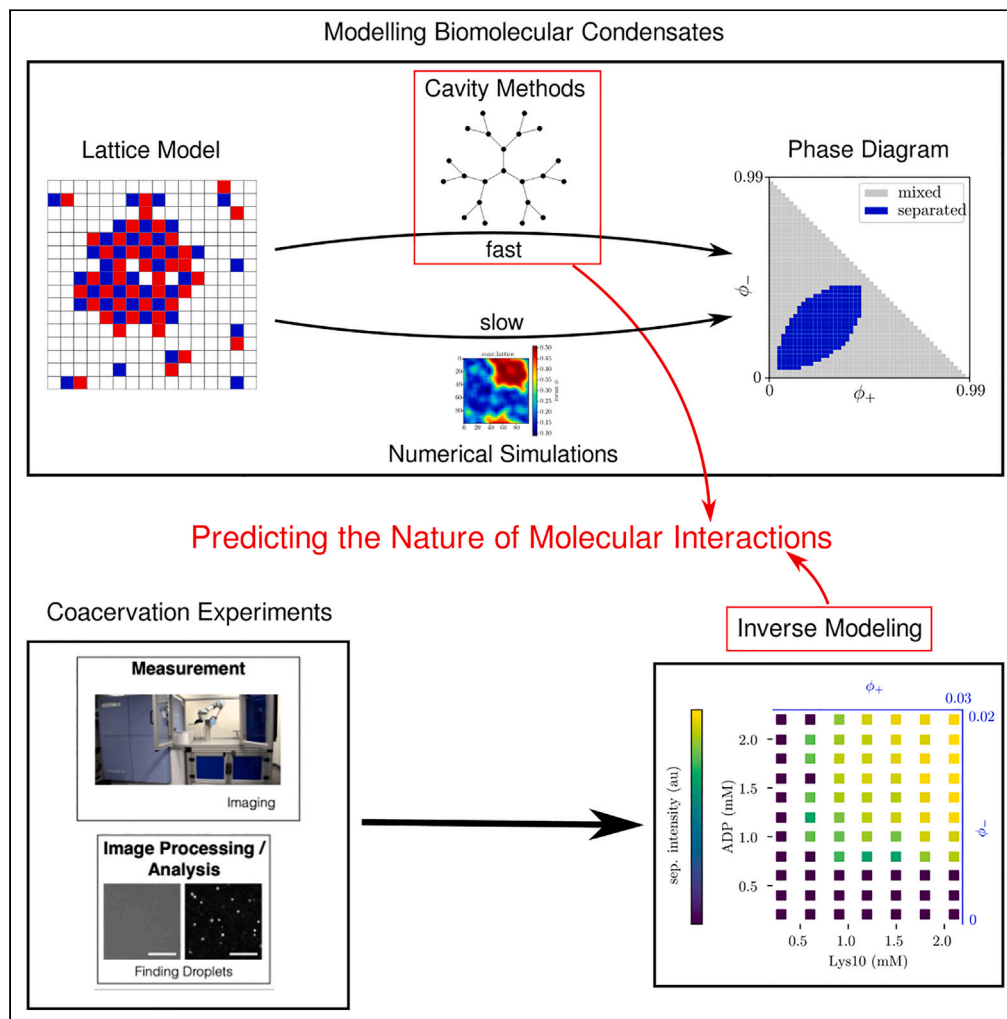


Article

Statistical mechanics of biomolecular condensates via cavity methods



Nino Lauber,
Ondrej Tichacek,
Rudrarup Bose, ...,
T-Y Dora Tang,
Kepa Ruiz-Mirazo,
Daniele De
Martino

daniele.demartino@ehu.es

Highlights

We apply cavity methods to calculate the free energy of biomolecular condensates

We reproduce simulations of ternary systems where one-factor approximations fail

Cavity methods are fast and apt to inverse model coacervation experiments

We reconstruct the phase diagram of polylysine-nucleotides associative demixing



Article

Statistical mechanics of biomolecular condensates via cavity methods

Nino Lauber,^{1,2,3,10} Ondrej Tichacek,^{4,10} Rudrarup Bose,⁵ Christoph Flamm,⁶ Luca Leuzzi,^{7,8} T-Y Dora Tang,⁵ Kepa Ruiz-Mirazo,^{1,3} and Daniele De Martino^{1,9,11,*}

SUMMARY

Physical mechanisms of phase separation in living systems play key physiological roles and have recently been the focus of intensive studies. The strongly heterogeneous nature of such phenomena poses difficult modeling challenges that require going beyond mean-field approaches based on postulating a free energy landscape. The pathway we take here is to calculate the partition function starting from microscopic interactions by means of cavity methods, based on a tree approximation for the interaction graph. We illustrate them on the binary case and then apply them successfully to ternary systems, in which simpler one-factor approximations are proved inadequate. We demonstrate the agreement with lattice simulations and contrast our theory with coacervation experiments of associative de-mixing of nucleotides and poly-lysine. Different types of evidence are provided to support cavity methods as ideal tools for modeling biomolecular condensation, giving an optimal balance between the consideration of spatial aspects and fast computational results.

INTRODUCTION

The spatial organization of the components of biological cells is a very important aspect of their physiology¹ and its nature is eminently physical. For instance, with regard to metabolism, different processes require in principle different environmental conditions and segregation mechanisms to ensure an efficient orchestration of cellular functionalities through compartmentalization. Classical, well-understood examples include oxidative phosphorylation and photosynthesis (performed in specialized organelles, mitochondria, and chloroplasts, respectively²). It has been recently proposed that, apart from compartmentalization through lipid membranes, living systems could deal with the problem of creating and controlling microenvironments by means of the physical mechanism of phase separation, where liquid mixtures spatially segregate.³ Examples range from ATP concentration in stress granules⁴ to control of gene expression by chromatin condensation^{5,6} and formation of protein complexes,⁷ while a better-established mechanism is the storage of carbohydrates into starch and/or glycogen,⁸ avoiding potential osmotic imbalance. On the flip side, it is well known that the inadequate formation of biomolecular condensates is the physical correlate of many prion-based pathologies, such as mad cow or Alzheimer's disease, for instance. Further evidence is accumulating to support that phase separation could play a key role in gene regulation in diseases, including cancer.^{9–12}

Additionally, in the field of origins of life the interest in coacervation has been "rediscovered" in recent years^{13–16} as a simple and highly plausible compartmentalization mechanism under prebiotic conditions (as it was actually suggested in the early days of the field¹⁷). It is becoming increasingly apparent, particularly among researchers in the protocell camp, that both physics and chemistry must come together to open the way toward biological complexity, in a scenario in which it is necessary to combine not only a significant diversity of molecular components but also of interactions and transformation processes.¹⁸

One main difference with respect to classical physical and chemical studies on phase separation is the extremely heterogeneous and complex nature of biological components,¹⁹ with thousands of different species of microscopic units (that can be complex themselves, such as polymers) even in a relatively simple bacterium such as *E. coli*.²⁰ Besides, the specific focus of investigations in life sciences is centered on problems of

¹Biofisika Institute (CSIC, UPV/EHU), Barrio Sarriena s/n. 48940 Leioa, Bizkaia, Spain

²Donostia International Physics Center (DIPC), Paseo Manuel de Lardizabal, 4, 20018 Donostia-San Sebastian, Gipuzkoa, Spain

³Department of Philosophy (UPV/EHU), Avenida de Tolosa 70, 20018 Donostia-San Sebastian, Gipuzkoa, Spain

⁴Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo náměstí 542/2, 160 00 Praha 6, Czech Republic

⁵Max Planck Institute of Molecular Cell Biology & Genetics, Pflotenhauerstraße 108, 01307 Dresden, Germany

⁶Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria

⁷Department of Physics, Università di Roma la Sapienza, Piazzale Aldo Moro 5, 00185 Rome, Italy

⁸Institute of Nanotechnology, Soft and Living Matter Laboratory, Consiglio Nazionale delle Ricerche (CNR-NANOTEC), Piazzale Aldo Moro 5, 00185 Rome, Italy

⁹Ikerbasque Foundation, Alameda Urquijo, 36, 48011 Bilbao, Bizkaia, Spain

¹⁰These authors contributed equally

¹¹Lead contact

*Correspondence: daniele.demartino@ehu.es
<https://doi.org/10.1016/j.isci.2023.106300>



control, design, and inverse modeling. These aspects spurred the wide use of mean-field approximations for theoretical and computational studies, in particular regarding the extension of the regular solution model.^{21–25}

In the case of polymer solutions, the classical Flory-Huggins (FH) model can be used to describe the segregative de-mixing with the formation of multiple phases, each enriched in one respective polymer.^{26,27} This model was later extended by Voorn and Overbeek (VO) for solutions of oppositely charged poly-ions (i.e., charged polymers) which usually display associative de-mixing, with the formation of one phase enriched in multiple poly-ions.²⁸ Both models are mean-field approximations that build on the interplay between the entropy that drives the mixing of the system and the enthalpy, which results from the interaction energies (derived from van der Waals or ionic forces) between the molecules. As such, they do not explicitly deal with the partition function of the system.

In recent years these mean-field models have been extended successfully to the case of complex coacervation, in particular taking into account the effect of the sequence and focusing on the problem of phase separation of intrinsically disordered proteins, by means of random phase approximation theory.^{29,30} In this article we will focus on the multi-canonical model defined in^{21–25} and its proposed mean-field approximation, namely the regular solution model. We chose to analyze this model for the sake of simplicity, since it is a generalization of the Potts model, which is a minimal model representing an instance of universality class in phase transitions.³¹

One common shortcoming of the aforementioned mean-field approaches is that they tend to neglect spatial correlations by recurring to one-factor approximations, akin to the well-known Curie-Weiss (CW) approximation in magnetic systems.³¹ This is known to lead to difficulties in presence of idiosyncratic, repulsive interactions and frustration, yielding multi-equilibrium. To overcome these difficulties more refined mean-field approximations were developed within the framework of magnetic systems, among which we can find the Bethe-Peierls (BP) approximation,^{32,33} recently reformulated as cavity methods,³⁴ or message passing and belief propagation algorithms.³⁵ The latter are considered important standard methods for the statistical physics of spin glasses and disordered systems, with applications that include inference, information theory and resolution of combinatorial optimization problems.³⁶

In this work we will apply the BP mean-field technique to describe the self-assembly of biomolecular condensates, focusing more specifically on the problem of reproducing numerical simulations of a general grand-canonical heterogeneous lattice model. The article is organized as follows. First, we will introduce the BP approach and compare it with the regular solution model on a simple standard binary system, providing an analytical formula for the spinodal line that improves quantitatively the match with numerical simulations (in comparison with the classical formula coming from the regular solution model). Then a simple ternary system will be considered, to explore a case where the regular solution model is clearly inadequate (i.e., unable to show even a qualitative agreement with numerical simulations) for the case of mixed repulsive and attractive interactions, known to lead to associative de-mixing. We will demonstrate that the BP approach reproduces much better numerical simulations and provides an immediate method to draw phase diagrams with a semi-quantitative controlled match. This supports the use of BP to classify much more accurately de-mixing phenomena in ternary systems in the interaction coupling space, where we successfully recapitulate the main three modes of phase separation, i.e. i) associative, ii) segregative, and iii) counter-ionic de-mixing. Thus the way is paved for inverse modeling and inference of couplings from experimental results, so our last section will be devoted to the reconstruction of de-mixing phase diagrams from high-throughput data. A relatively simple but well-controlled *in vitro* system with poly-lysine and nucleotides in buffer solution was chosen because it allowed us to explore parameter space more systematically, it is relevant in the context of prebiotic chemistry and it fulfills our quantitative modeling purposes. Our findings and the potential implications of our work will be finally summarized in a conclusion section.

RESULTS

As a starting point, we will consider the microscopic coarse-grained multi-component solution model defined e.g. in²², which can be seen as a particular instance of the generalized Potts model.³⁷ The space is discretized into a regular lattice with N sites, where each site- i is in a state $\sigma_i = 0, 1 \dots q$, standing for the presence of a particle of a given type (e.g., solvent or various solutes), q being the number of components. The interaction between two lattice sites σ_i, σ_j is described by a given function $J(\sigma_i, \sigma_j)$ (for the usual

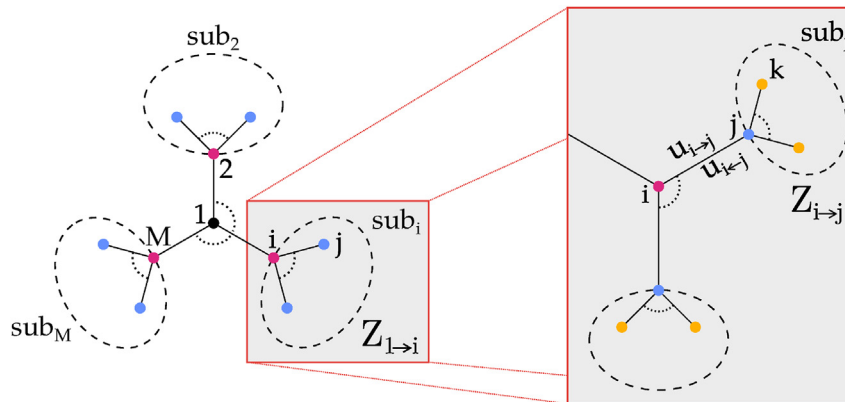


Figure 1. Illustration of the tree-like approximation for the calculation of the partition function

Starting from one node, the latter can be decomposed in terms of the conditional partition functions of sub-systems sprouting from neighboring sites and the procedure can be iterated recursively.

Potts model $J(\sigma_i, \sigma_j) = \delta_{\sigma_i, \sigma_j}$ and the number of different kinds of particles is controlled by their chemical potentials $\mu(\sigma_i)$. The Hamiltonian of the system is, therefore:

$$H(\vec{\sigma}) = - \sum_{\langle i,j \rangle} J(\sigma_i, \sigma_j) + \sum_i \mu(\sigma_i) \quad (\text{Equation 1})$$

where the first sum runs over all neighboring lattice sites $\langle i,j \rangle$. In contrast with the regular solution model, we will not postulate a form for the free energy, but rather derive it from microscopic interactions by computing the partition function. This is the fundamental quantity that bridges between the molecular microscopic interactions and the collective macroscopic behavior of the system, alongside with its thermodynamic properties.³ Its computation makes it possible to map between the energy as a function of the diverse microscopic configurations and the free energy as a function of macroscopic variables (e.g. concentrations and/or chemical potentials, temperature). Here its expression is ($\beta = 1/T$ is the inverse temperature)

$$Z = \sum_{\vec{\sigma}} e^{-\beta H(\vec{\sigma})} = \sum_{\sigma_1, \dots, \sigma_N} e^{\beta \left[\sum_{\langle i,j \rangle} J(\sigma_i, \sigma_j) - \sum_i \mu(\sigma_i) \right]} \quad (\text{Equation 2})$$

We remark that the computation of the partition function is in general a very difficult computational task since an evaluation of the sum on the RHS could require to count objects with involved combinatorics.³⁶ For instance, its calculation for the simple Ising model without an external magnetic field corresponds to the enumeration of all closed loop in the underlying graph.³⁸ We will compute it by approximating the lattice in terms of a tree-graph branching out from any given site. In this way, the lattice is decomposed into sub-systems that are connected only by the sprouting site. Once the state value of the latter is fixed, the partition function can be factorized in terms of the partition functions of the sub-systems and the procedure can be iterated recursively (see Figure 1).

We end up with the equations

$$Z_{i \rightarrow j}(\sigma_i) = \sum_{\sigma_j} e^{\beta [J(\sigma_i, \sigma_j) - \mu(\sigma_j)]} \prod_{k \in \partial_{ji}} Z_{j \rightarrow k}(\sigma_j) \quad (\text{Equation 3})$$

where \prod stands for the product, ∂_{ji} are all the sites connected to j except from i , and $Z_{i \rightarrow j}(\sigma_i)$ is the partition function of the sub-system starting from site- j , given that site- i is fixed to the value σ_i . In general, we have: $Z_{i \rightarrow j}(\sigma_i) \neq Z_{j \rightarrow i}(\sigma_j)$

Binary system

For a simple binary phase separation we have $q = 2$ and $\sigma_i = 0, 1$. Fixing $J(\sigma_i, 0) = \mu(0) = 0$ the cavity equations will be

$$Z_{i \rightarrow j}(\sigma_i) = \prod_{k \in N_{ji}} Z_{j \rightarrow k}(0) + e^{\beta [J(\sigma_i, 1) - \mu]} \prod_{k \in N_{ji}} Z_{j \rightarrow k}(1) \quad (\text{Equation 4})$$

Parametrizing $Z_{i \rightarrow j}(\sigma_i) = A_{i \rightarrow j} e^{\beta u_{i \rightarrow j} \sigma_i}$, it is possible to see that Equation 4 leads to a set of self-consistent equations for the messages $u_{i \rightarrow j}$:

$$u_{i \rightarrow j} = \frac{1}{\beta} \log \left(\frac{1 + e^{\beta J - \beta \mu + \beta \sum_{k \in N_{ji}} u_{j \rightarrow k}}}{1 + e^{-\beta \mu + \beta \sum_{k \in N_{ji}} u_{j \rightarrow k}}} \right) \quad (\text{Equation 5})$$

Considering that the tree-graph is a Caley-Graph with a branching of $C = K + 1$, and assuming homogeneity $u_{i \rightarrow j} = u$ ($\forall i, j$) we get for Equation 5:

$$u = \frac{1}{\beta} \log \left[\frac{1 + e^{\beta(J - \mu + Ku)}}{1 + e^{\beta(-\mu + Ku)}} \right] \quad (\text{Equation 6})$$

In addition one can assume that the average site occupation or density (equivalent to the occupation probability of the lattice site by the solute) $\langle \sigma \rangle = \varphi$ will verify the following equation (see STAR Methods S.1 for further details):

$$\begin{aligned} \varphi &= \frac{e^{-\beta \mu} \prod_{i \in N_1} Z_{1 \rightarrow i}(1)}{\prod_{i \in N_1} Z_{1 \rightarrow i}(0) + e^{-\beta \mu} \prod_{i \in N_1} Z_{1 \rightarrow i}(1)} \\ &= \frac{e^{-\beta \mu + \beta(K+1)u}}{1 + e^{-\beta \mu + \beta(K+1)u}} \end{aligned} \quad (\text{Equation 7})$$

These equations express implicitly the state equation $\varphi(\mu)$, from which the phase separation curve $(\beta J)(\varphi)$ can be obtained, in implicit form, by standard thermodynamic stability analysis upon introducing parameter $w = e^{\beta u}$:

$$\begin{aligned} \frac{\varphi}{1 - \varphi} &= \frac{w(w - 1)}{e^{\beta J} - w}, \quad w_{1,2} = \frac{-b \pm \sqrt{b^2 - 4K^2 e^{\beta J}}}{-2K}, \\ b &= Ke^{\beta J} + K - e^{\beta J} + 1 \end{aligned} \quad (\text{Equation 8})$$

where the values $w_{1,2}$ correspond to the two branches of the spinodal line. The above parametric formula can be compared now with the one obtained from the Curie-Weiss model (see STAR Methods S.3 for a derivation)

$$\beta J = \frac{1}{(K+1)(1 - \varphi)\varphi} \quad (\text{Equation 9})$$

and contrasted with microscopic numerical simulations of the model on a regular square lattice (for which $K = 3$) as illustrated in Figure 2, where we show the binodal lines as well (calculated via Maxwell construction, see STAR Methods S.4). As it can be easily observed, even though the model on a nearest-neighbor lattice is extremely far from being tree-like, numerical simulations are in better quantitative agreement with the predictions of the cavity method, compared with the regular solution equations. An example of the biological application of the phase separation in a binary model to receptor clustering can be found in³⁹.

Ternary system

We next investigated the phase separation between two types of solutes and a solvent. We define configurations as $\sigma_i \in \{-1, 0, +1\}$. For the sake of simplicity we assume $J(0, \sigma_i) = J(\sigma_i, 0) = 0$ $\mu(0) = 0$ and rename $J(-1, -1) = J_{--}$, $J(+1, +1) = J_{++}$, $J(-1, +1) = J(+1, -1) = J_{+-}$ $\mu(-1) = \mu_-$, $\mu(+1) = \mu_+$. In addition, one can apply the BP method to compute the partition function approximately. In this ternary system, the recursive equations for the partition function of sub-systems along the branches will be

$$\begin{aligned} Z_{i \rightarrow j}(\sigma_i) &= e^{\beta[J(\sigma_i, -1) - \mu_-]} \prod_{k \in \partial_{ji}} Z_{j \rightarrow k}(-1) \\ &+ e^{\beta[J(\sigma_i, 1) - \mu_+]} \prod_{k \in \partial_{ji}} Z_{j \rightarrow k}(+1) \\ &+ \prod_{k \in \partial_{ji}} Z_{j \rightarrow k}(0) \end{aligned} \quad (\text{Equation 10})$$

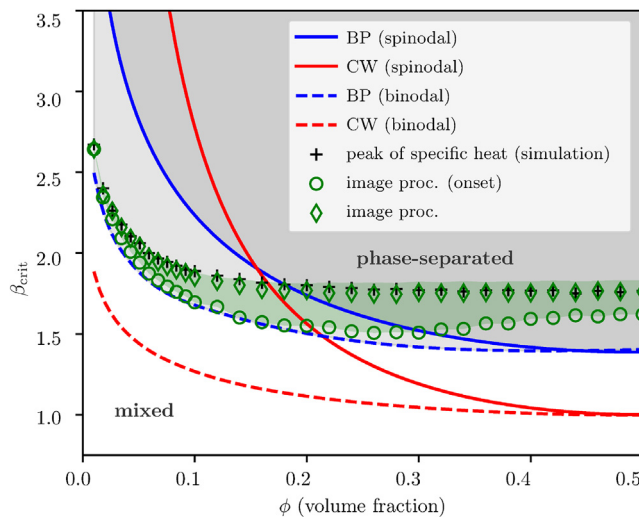


Figure 2. Transition lines for the binary system

Comparison of the transition lines (continuous: spinodal, dashed: binodal) of the mean-field regular solution model (Curie-Weiss-like, red curves), the mean-field finite connectivity cavity method (Bethe-Peierls, blue lines) and the numerical simulations on the nearest-neighbor 2D lattice model (points).

The black data points are obtained from the peak of the specific heat, whereas the green ones by means of the image processing method described in the text. The green shaded region depicts the transition from a metastable to a phase-separated state as detected by the image processing algorithm (see Figure S3 in the STAR Methods for more details).

Once again these equations can be written in exponential form singling out the dependence on the starting node σ in terms of the so-called message variables. Restricting ourselves to a homogeneous Cayley tree, we assume the homogeneity of the messages u^- , u^+ , and we obtain the equations

$$\begin{aligned} u^- &= \frac{1}{\beta} \log \left[\frac{e^{\beta(J_{--} - \mu_- + Ku^-)} + 1 + e^{\beta(J_{+-} - \mu_+ + Ku^+)}}{e^{\beta(-\mu_- + Ku^-)} + 1 + e^{\beta(-\mu_+ + Ku^+)}} \right], \\ u^+ &= \frac{1}{\beta} \log \left[\frac{e^{\beta(J_{+-} - \mu_- + Ku^-)} + 1 + e^{\beta(J_{++} - \mu_+ + Ku^+)}}{e^{\beta(-\mu_- + Ku^-)} + 1 + e^{\beta(-\mu_+ + Ku^+)}} \right] \end{aligned} \quad \text{(Equation 11)}$$

that together with the average densities

$$\begin{aligned} \varphi_- &= \frac{e^{-\beta\mu_- + \beta(K+1)u^-}}{e^{-\beta\mu_- + \beta(K+1)u^-} + 1 + e^{-\beta\mu_+ + \beta(K+1)u^+}} \\ \varphi_+ &= \frac{e^{-\beta\mu_+ + \beta(K+1)u^+}}{e^{-\beta\mu_- + \beta(K+1)u^-} + 1 + e^{-\beta\mu_+ + \beta(K+1)u^+}} \end{aligned} \quad \text{(Equation 12)}$$

provide the state equations of the system. The phase diagram can be drawn by checking if the matrix

$$H = \begin{pmatrix} \frac{\partial \mu_+}{\partial \varphi_+} & \frac{\partial \mu_+}{\partial \varphi_-} \\ \frac{\partial \mu_-}{\partial \varphi_+} & \frac{\partial \mu_-}{\partial \varphi_-} \end{pmatrix}$$

is positive definite (by the Routh-Hurwitz criterion $\det H > 0$ and $\text{tr}H > 0$). Results from numerical simulations and mean-field calculations are summarized in Figure 3, where depending on the interaction signs phase separation can be classified into three different kinds: i) associative, ii) segregative, and iii) counter-ionic de-mixing. Those three general types of phase behavior are well accounted for if cavity methods are applied, but more naive or direct mean-field models clearly fail to do so. A strong advantage of mean-field approximations is their low computational cost as compared to an explicit lattice model simulation, considering the fact that the system behavior can be assessed by solving a handful of nonlinear equations. In comparative terms, the time to reconstruct the phase diagrams for the ternary system, shown in Figure 3, differs by 6-7 orders of magnitude when we switch from the lattice model simulations to the resolution of the BP equations (hours vs ms in our implementation). This reduction of computational time enables a full inverse modeling approach to experimental data.

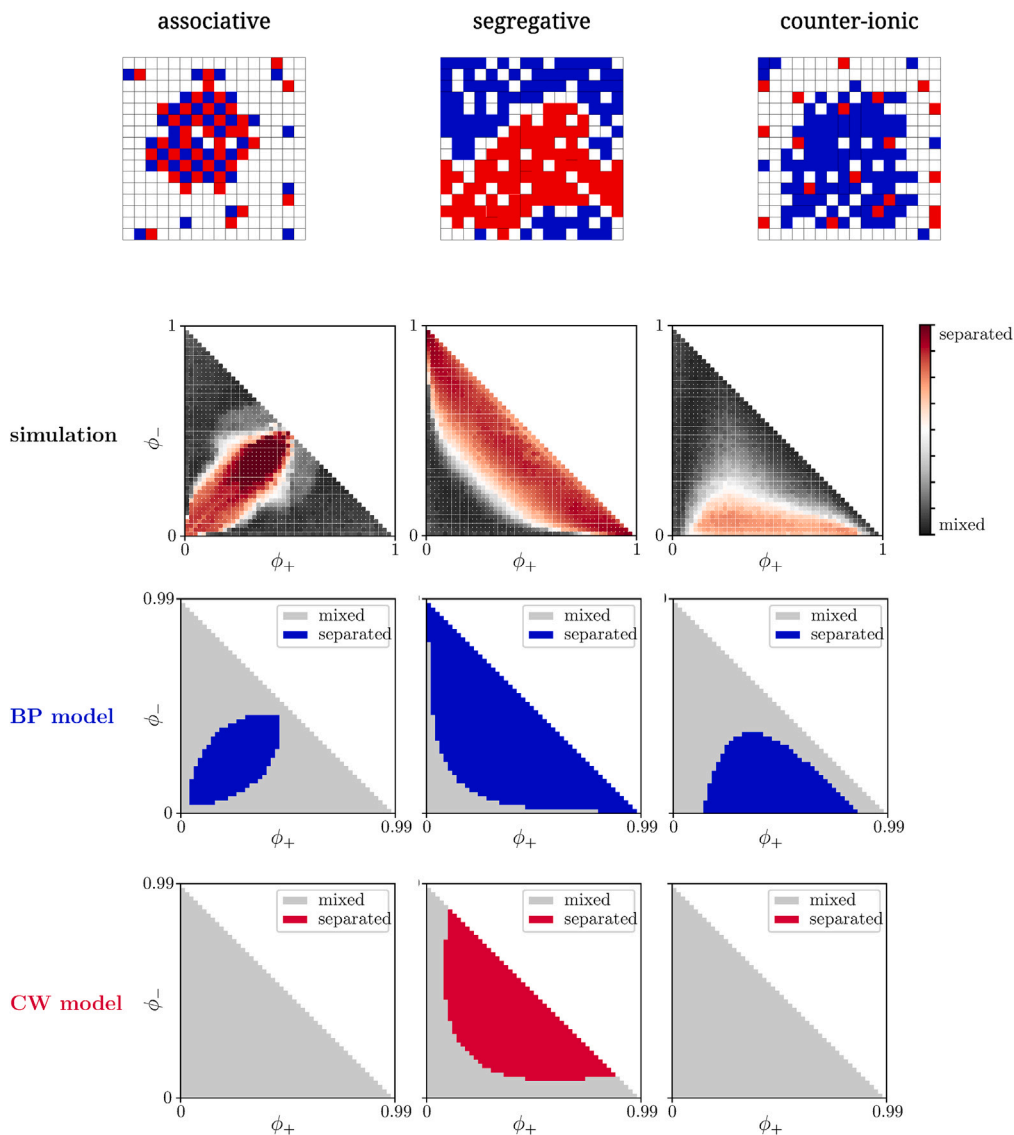


Figure 3. Phase diagrams for the ternary system

Comparison of the phase diagrams in the plane of solutes volume fractions (ϕ_+ , ϕ_-) for the ternary system obtained from lattice model simulations (top, obtained from image processing), BP mean-field (center, spinodal), and regular solution mean-field model (bottom, spinodal), for the associative (left $J_{--} = J_{++} = -1, J_{+-} = 3$), segregative (middle $J_{--} = J_{++} = 1, J_{+-} = -3$) and counter-ionic (right $J_{++} = 2, J_{--} = 0, J_{+-} = 0.5$) de-mixing cases (see text), respectively BP classifies simulation results with 88% accuracy.

Modeling experiments

We will consider here the experimental phase diagram of a system of poly-lysine and adenosine-diphosphate (ADP) in buffer solution, as obtained by microscopy imaging. This system, given the residual electrostatic charge of its components, is expected to show typical associative de-mixing behavior, which is not accounted for correctly by the regular mean-field solution model. We considered thus the task of inferring the parameters of the aforementioned ternary solution model that reproduce the experimental phase diagram.

This has been formally modeled as a binary classification problem and we implemented an algorithm for parameter inference based on heuristic optimization via differential evolution algorithms.⁴⁰ Results are reported in Figure 4, where we show the experimental phase diagram together with the simulations of the

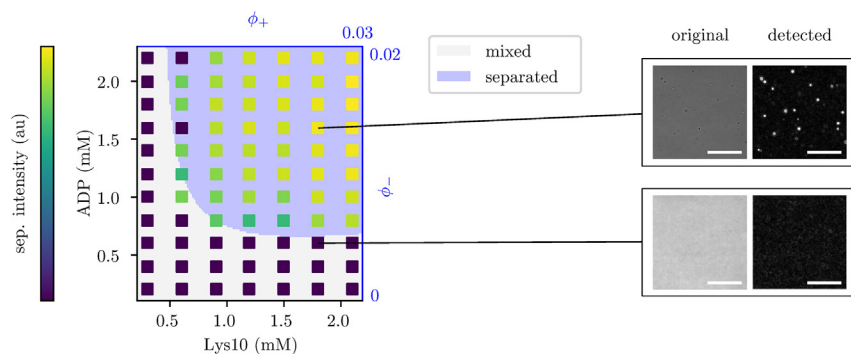


Figure 4. Experimental phase diagram reconstruction

Experimental phase diagram in the concentrations plane of a system of poly-lysine and ADP in the salt aqueous solution obtained by microscopic imaging, overlaid by a phase diagram of an inferred ternary system model on a 3D cubic lattice in the volume fractions plane

Microscope images of real samples (right) show the formation of the condensate in the form of droplets, scale bars represent $100 \mu\text{m}$. Separation intensity comes from an in-house developed method of automatic image processing of the microscopy images and corresponds to the logarithm of the area of the phase-separated region. The inferred model parameters are $J_{++} = -3.5$, $J_{--} = -2$, $J_{+-} = 3.8$. Experimental points are reproduced with 97% accuracy.

inferred models that are compatible with the associative de-mixing case. Although this provides a quantitative description, a small perturbation in the initial parameter values can lead to an equally well-inferred model, with different parameters. This hints at the presence of many local maxima for the likelihood of model parameters and calls for more refined experiments and/or an inference scheme going beyond simple binary classification. We then performed an inference calculation by constraining the model parameters to be in the region of segregative de-mixing ($J_{--} > 0$, $J_{++} > 0$, $J_{+-} < 0$, not shown), obtaining a consistently higher error rate (that is, the fraction of mismatch in binary phase classification, 15%, versus the 3% of the associative case). This shows that our simple setting is able to tell apart the different phase separation mechanisms.

Predicted re-entrance in the experimental phase diagram

As an additional piece of evidence, let us report here the confirmation of a successful prediction made by our inferred model. Our inferred binding interaction energies predict a form of the phase diagram that falls in the case of associative de-mixing, implying a re-entrance to the symmetric, well-mixed phase for high enough concentrations/volume fractions. Therefore, we performed experiments in that concentration range where such a phenomenon should occur and we found it, in nice agreement with our theoretical prediction (see Figure 5). We remark that the binding energies were inferred in the first part of the phase diagram, apparently showing no sign of re-entrance, so this can be seen as a bona fide prediction or a successful non-trivial generalization.

Image processing and simulations

Numerical simulations were performed via the Monte Carlo Kawasaki scheme,⁴¹ enforcing fixed volume fraction, and the inverse critical temperature was estimated independently, as the location of the peak of specific heat and the point where the free energy profile changes its concavity (see Figure 6, right). The results of Monte Carlo simulations were analyzed via a heuristic image processing algorithm to identify the occurrence of phase separation. Local particle densities were computed for each component in a lattice snapshot through 2 days convolution with a Gaussian kernel and periodic boundary condition. The logarithm of the distribution of the local density thus obtained was considered via the Gibbs equation as a bona-fide approximation of the free energy of de-mixing. An automated inspection of the number of minima of this reconstructed profile leads to the classification of systems into well-mixed and phase-separated. The procedure is inspired by statistical tests comparing non-parametric distributions and produces a separation confidence score depicted as the color scale in Figure 3. The method is illustrated in Figure 6, left, for two instances of the ternary system that are representative of the well-mixed and phase-separated systems, respectively. The method provided very accurate estimates, as can be seen in Figure 6, right where we show the scattering plot of the inverse critical temperature, at varying volume fractions, for the binary system obtained from the calculation of the peak of the specific heat (see STAR Methods S.5 and S.6).

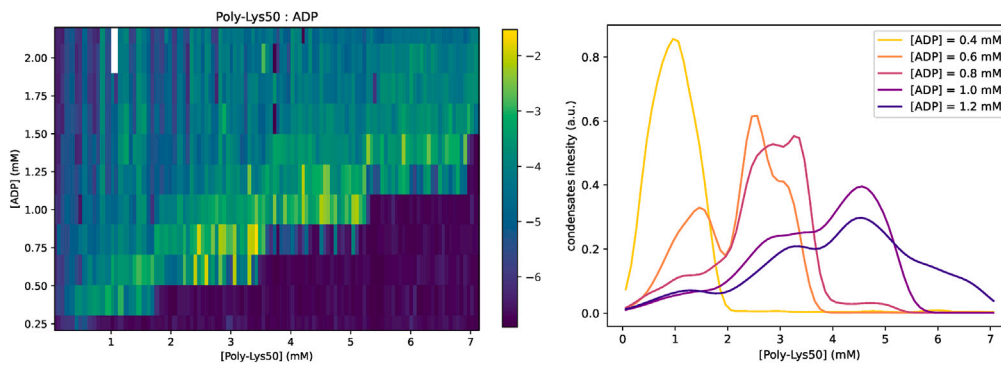


Figure 5. Predicted re-entrance in the phase diagram

The concentration of ADP ranged from 0.2 to 2.2 mM and was sampled in 11 steps while the concentration of poly-Lysine ranged from 0.06 to 7.14 mM and was sampled in 112 steps. **Left:** The resulting 11×112 matrix showing the level of phase separation was produced automatically with a custom image-processing method. **Right:** Cuts through the phase diagram at five different constant concentrations of ADP show the “re-entrance” in the phase diagram, i.e. with increasing concentration of poly-Lysine the phase-separated region appears and then disappears again. The cuts through the phase diagram have been smoothed using a Gaussian kernel-density estimator.

DISCUSSION

The study of how biological cells manage or fail to control the spatial/physical conditions of their internal milieu through mechanisms of phase separation is of paramount importance. This should benefit from the wealth of knowledge acquired in the field of the statistical physics of phase transitions in disordered systems, not only in terms of quantitative modeling, but also in data analysis and experiment design. In this article, we have illustrated an application of the mean-field Bethe-Peierls (BP) approximation in the context of heterogeneous phase separation. We have shown that the BP approach quantitatively reconstructs phase diagrams where the standard regular solution model fails even to give a qualitative description: more precisely, in a minimally heterogeneous ternary lattice microscopic model. In particular, the regular solution model^{21–25} can be seen as coming from an underlying hypothesis of a fully connected graph (see [STAR Methods S.2](#)) and this procedure is not guaranteed to provide a reliable approximation in presence of idiosyncratic interactions without appropriate ad-hoc gauge transformations (an elementary example being given by the Ising antiferromagnet³⁷).

An explicit derivation of the free energy from the partition function, as we do here with cavity methods, helps in the formulation of an appropriate ansatz for the order parameter form and underlying symmetries. This, apart from being a more adequate theoretical strategy to deal with condensation phenomena, opens a new way for quantitative modeling of experimental data. We actually provided an example reproducing the experimental phase diagram of the associative de-mixing of poly-lysine in the presence of nucleotides. The finding that many models lead to a quantitative description of experimental phase diagrams will deserve further investigation. In this respect, our approach, applied to synthetic data from lattice model simulations could shed light on the right experimental quantities to be measured, leading to well-defined descriptions of the system (i.e., optimal experimental protocols).

In dealing with real data, an interesting ingredient to analyze with our method would be the introduction of inner degeneracy for the basic degrees of freedom, in order to model complex mixtures, since it could also potentially trigger inverse behaviors.⁴² Yet, in order to reproduce the experimental phase diagram we did not need to consider this additional parameter. This could be ascribed to the simplicity of the data being analyzed and to the fact that the interpretation of single lattice points as single molecules is really for illustrative purposes and it should be revised in the light of quantitative comparison with experiments. This is well known for the Ising model, where lattice points are usually not identified with spins of single atoms or molecules but rather with aptly coarse-grained magnetic domains of a given scale.³¹ This aspect is non-trivial and its clarification would touch upon the establishment of a renormalization group approach for these systems, a task that we leave for further investigations.

Although our case study is relatively simple (in comparison with the biomolecular condensates that one can find in a cell’s cytoplasm), it is sufficient for our purposes here, and it is offered also as a relevant example for the origin of life research. Another interesting modification of the model would be to include explicit spatial

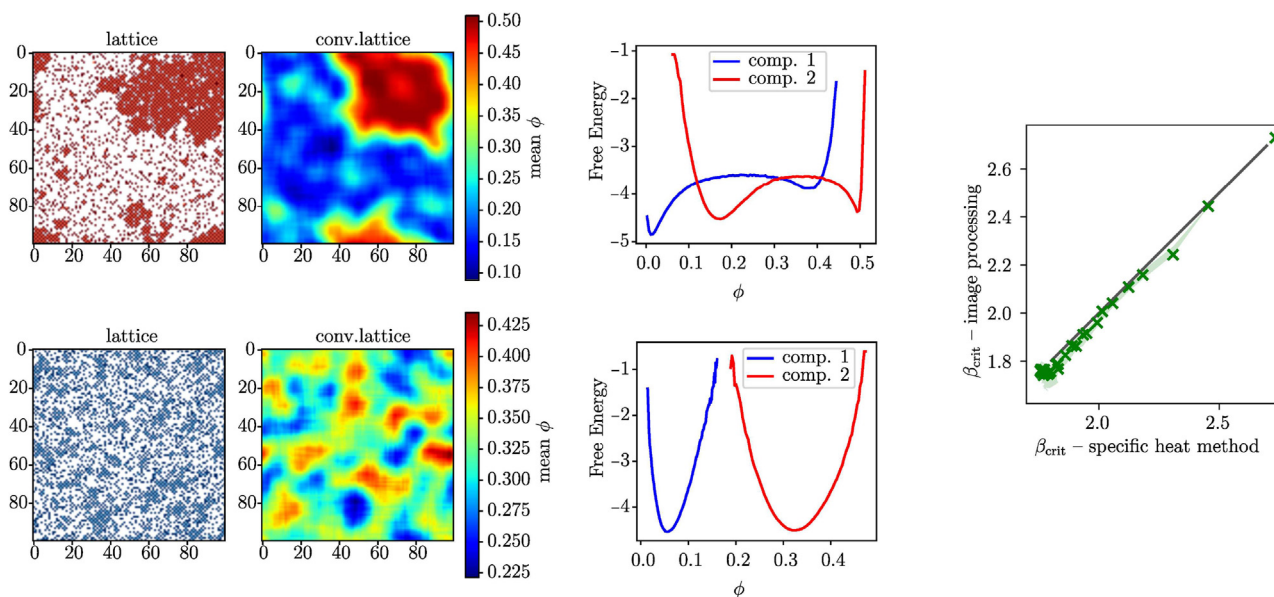


Figure 6. Automated detection of phase separation via image processing

Left: lattice model microscopic configuration and its convolution. Center: extracted free energy profile. Two instances of the ternary systems representative of well-mixed and phase-separated behavior respectively. A total of 100 snapshots sampled from the end of the simulation were processed independently. Right: comparison of the inverse critical temperature obtained from specific heat peak computation and image processing automated detection for the binary system.

dependence of the interactions in the form of quenched noise, like in spin glasses, to model in more complex cellular context interactions of the system with other degrees of freedom in a disordered matrix. This could potentially lead to a completely qualitatively different behavior, especially in the case of very heterogeneous idiosyncratic interactions. It is in fact known that the Potts model can be generalized to accommodate for glassy behavior.^{43,44}

One of the caveats of the cavity method consists in its mean-field character, thus limiting its predictive capabilities, in particular for the characterization of the phase transition (e.g. critical exponents³¹). Field theory methods could overcome this aspect and they have been applied successfully in this context recently to model phase separation of intrinsically disordered proteins^{45–47} see also the recent review.⁴⁸ Apart from that, a promising next step would be to use the BP approach to analyze strongly heterogeneous multi-component systems via replica methods, as it was originally the aim of the regular solution model. In contrast with the latter, BP will be not restricted to the case of mildly attracting interaction matrices. In fact, it could be used to explore any kind of interaction patterns, with idiosyncratic terms, since this approximation showed to be successful in attacking systems with a much more complex free energy landscape, such as spin glasses.

Limitations of the study

The nature of the proposed cavity method approximation is inherently mean field and thus semi-quantitative. Rigorously, it is known that the free energy calculated in a mean-field approximation (including our approach) is always an upper bound for the true free energy and that mean-field approximations fail to describe quantitatively critical behavior. More refined Monte Carlo numerical simulations or field theory-based methods would be needed e.g. for a quantitative assessment of critical exponents. The experimental system we successfully model is arguably rather simple, especially if compared to the complexity of a true cellular environment.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact

- Materials availability
- Data and code availability
- **METHOD DETAILS**
 - Formulation of the Bethe-Peierls (BP) model
 - Formulation of the regular solution model
 - Retrieving the phase diagrams for a binary-system using the Currie-Weiss (CW) model
 - Calculating the binodal-curve
 - Finite size scaling
 - Calibration of the image processing algorithm
 - Further details on numerical simulations
 - Further details on the comparison between the cavity method results and numerical simulations
 - Experimental methods

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106300>.

ACKNOWLEDGMENTS

This work has been supported by the Horizon 2020 Marie Curie ITN (“ProtoMet”—Grant Agreement no. 813873 with the European Commission), within which NL & RB obtained a PhD fellowship. NL thanks the John Templeton Foundation that funded part of his fellowship (grant #62220). Both NL and KRM acknowledge support from the Basque Government (IT1668-22) and the Spanish Ministry of Science and Innovation (PID2019-104576GB-I00). OT thanks the Biofisika Institute for kind hospitality during the development of this work. OT acknowledges the Faculty of Mathematics and Physics of the Charles University (Prague, Czech Republic) where he is enrolled as a PhD student. OT thanks for support to the European Regional Development Fund (project ChemBioDrug no. CZ.02.1.01/0.0/0.0/16 019/0000729). RB thanks the Technology Development Studio of MPI-CBG for their technical and infrastructural support. Many thanks to Martin Stöter setting up the automation of the experimental pipeline and Rico Barsacchi for assisting with the image acquisition.

AUTHOR CONTRIBUTIONS

Conceptualization DDM and KRM; methodology, DDM; investigation DDM, NL, OT, and RB; writing – original draft, DDM, NL, and OT; writing – review & editing, DDM, NL, OT, KRM, DT, CF, LL, and RB; funding acquisition, KRM and DT; resources RB, OT, and DT; supervision, DDM, KRM, DT, CF, and LL.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 31, 2022

Revised: December 23, 2022

Accepted: February 24, 2023

Published: March 6, 2023

REFERENCES

1. Harold, F.M. (2003). *The Way of the Cell: Molecules, Organisms, and the Order of Life* (Oxford University Press).
2. M Berg, J., Stryer, L., and Tymoczko, J.L. (2015). *Biochemistry* (Springer-Verlag).
3. Landau, L., Lifshitz, E.M., and Pitaevskii, L.P. (1980). *Course of theory physics. Statistical Physics vol. 9. chap. IX*.
4. Wurtz, J.D., and Lee, C.F. (2018). Stress granule formation via ATP depletion triggered phase separation. *New J. Phys.* 20, 045008. <https://doi.org/10.1088/1367-2630/aab549>.
5. Shin, Y., and Brangwynne, C.P. (2017). Liquid phase condensation in cell physiology and disease. *Science* 357, eaaf4382. <https://doi.org/10.1126/science.aaf4382>.
6. Conte, M., Irani, E., Chiariello, A.M., Abraham, A., Bianco, S., Esposito, A., and Nicodemi, M. (2022). Loop-extrusion and polymer phase-separation can co-exist at the single-molecule level to shape chromatin folding. *Nat. Commun.* 13, 4070. <https://doi.org/10.1038/s41467-022-31856-6>.
7. Sartori, P., and Leibler, S. (2020). Lessons from equilibrium statistical physics regarding the assembly of protein complexes. *Proc. Natl. Acad. Sci. USA* 117, 114–120. <https://doi.org/10.1073/pnas.1911028117>.
8. Frayn, K.N. (2009). *Metabolic Regulation: A Human Perspective* (John Wiley & Sons).
9. Zhang, Y., Bertulat, B., Tencer, A.H., Ren, X., Wright, G.M., Black, J., Cardoso, M.C., and Kutateladze, T.G. (2019). MORC3 forms nuclear condensates through phase separation. *iScience* 17, 182–189. <https://doi.org/10.1016/j.isci.2019.06.030>.
10. Petronilho, E.C., Pedrote, M.M., Marques, M.A., Passos, Y.M., Mota, M.F., Jakobus, B., de Sousa, G.D.S., Pereira da Costa, F., Felix, A.L., Ferretti, G.D.S., et al. (2021). Phase separation of p53 precedes aggregation and is affected by oncogenic mutations and ligands. *Chem. Sci.* 12, 7334–7349. <https://doi.org/10.1039/D1SC01739J>.

11. Lu, J., Qian, J., Xu, Z., Yin, S., Zhou, L., Zheng, S., and Zhang, W. (2021). Emerging roles of liquid-liquid phase separation in cancer: from protein aggregation to immune-associated signaling. *Front. Cell Dev. Biol.* 9, 631486. <https://doi.org/10.3389/fcell.2021.631486>.
12. Ahn, J.H., Davis, E.S., Daugird, T.A., Zhao, S., Quiroga, I.Y., Uryu, H., Li, J., Storey, A.J., Tsai, Y.H., Keeley, D.P., et al. (2021). Phase separation drives aberrant chromatin looping and cancer development. *Nature* 595, 591–595. <https://doi.org/10.1038/s41586-021-03662-5>.
13. Dora Tang, T.Y., Rohaida Che Hak, C., Thompson, A.J., Kuimova, M.K., Williams, D.S., Perriman, A.W., and Mann, S. (2014). Fatty acid membrane assembly on coacervate microdroplets as a step towards a hybrid protocell model. *Nat. Chem.* 6, 527–533. <https://doi.org/10.1038/nchem.1921>.
14. Zwicker, D., Seyboldt, R., Weber, C.A., Hyman, A.A., and Jülicher, F. (2017). Growth and division of active droplets provides a model for protocells. *Nat. Phys.* 13, 408–413. <https://doi.org/10.1038/nphys3984>.
15. Donau, C., Späth, F., Sosson, M., Kriebisch, B.A.K., Schnitter, F., Tena-Solsona, M., Kang, H.S., Salibi, E., Sattler, M., Mutschler, H., and Boekhoven, J. (2020). Active coacervate droplets as a model for membraneless organelles and protocells. *Nat. Commun.* 11, 5167. <https://doi.org/10.1038/s41467-020-18815-9>.
16. Seyboldt, R., and Jülicher, F. (2018). Role of hydrodynamic flows in chemically driven droplet division. *New J. Phys.* 20, 105010. <https://doi.org/10.1088/1367-2630/aae735>.
17. Al, O. (1924). *Proiskhozhdenie Zhizny, Moscow; Izd. In Moscovskii, Rabochii (Repr. in transl. Bernal 1967, S 1-214)*.
18. Lauber, N., Flamm, C., and Ruiz-Mirazo, K. (2021). “Minimal metabolism”: a key concept to investigate the origins and nature of biological systems. *Bioessays* 43, 2100103. <https://doi.org/10.1002/bies.202100103>.
19. Floris, E., Piras, A., Dall’Asta, L., Gamba, A., Hirsch, E., and Campa, C.C. (2021). Physics of compartmentalization: how phase separation and signaling shape membrane and organelle identity. *Comput. Struct. Biotechnol. J.* 19, 3225–3233. <https://doi.org/10.1016/j.csbj.2021.05.029>.
20. Milo, R., and Phillips, R. (2015). *Cell Biology by the Numbers* (Garland Science).
21. Sear, R.P., and Cuesta, J.A. (2003). Instabilities in complex mixtures with a large number of components. *Phys. Rev. Lett.* 91, 245701. <https://doi.org/10.1103/PhysRevLett.91.245701>.
22. Jacobs, W.M., and Frenkel, D. (2017). Phase transitions in biological systems with many components. *Biophys. J.* 112, 683–691. <https://doi.org/10.1016/j.bpj.2016.10.043>.
23. Jacobs, W.M. (2021). Self-assembly of biomolecular condensates with shared components. *Phys. Rev. Lett.* 126, 258101. <https://doi.org/10.1103/PhysRevLett.126.258101>.
24. Carugno, G., Neri, I., and Vivo, P. (2022). Instabilities of complex fluids with partially structured and partially random interactions. In *Physical Biology*.
25. Krishna, S., and Brenner, M.P. (2021). Phase separation in fluids with many interacting components. *Proc. Natl. Acad. Sci. USA* 118, e2108551118.
26. Flory, P.J. (1942). Thermodynamics of high polymer solutions. *J. Chem. Phys.* 10, 51–61. <https://doi.org/10.1063/1.1723621>.
27. Huggins, M.L. (1942). Some properties of solutions of long-chain compounds. *J. Phys. Chem.* 46, 151–158. <https://doi.org/10.1021/j150415a018>.
28. Overbeek, J.T.G., and Voorn, M.J. (1957). Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *J. Cell. Comp. Physiol.* 49, 7–26. <https://doi.org/10.1002/jcp.1030490404>.
29. Lin, Y.-H., Forman-Kay, J.D., and Chan, H.S. (2016). Sequence-specific polyampholyte phase separation in membraneless organelles. *Phys. Rev. Lett.* 117, 178101. <https://doi.org/10.1103/PhysRevLett.117.178101>.
30. Lin, Y.-H., Brady, J.P., Forman-Kay, J.D., and Chan, H.S. (2017). Charge pattern matching as a ‘fuzzy’ mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J. Phys.* 19, 115003. <https://doi.org/10.1088/1367-2630/aa9369>.
31. Cardy, J. (1996). *Scaling and Renormalization in Statistical Physics*, 5 (Cambridge university press).
32. Bethe, H.A. (1935). Statistical theory of superlattices. *Proc. R. Soc. London Series A Math. Phys. Sci.* 150, 552–575. <https://doi.org/10.1098/rspa.1935.0122>.
33. Peierls, R. (1936). Statistical theory of superlattices with unequal concentrations of the components. *Proc. R. Soc. London Series A Math. Phys. Sci.* 154, 207–222. <https://doi.org/10.1098/rspa.1936.0047>.
34. Mézard, M., and Parisi, G. (2003). The cavity method at zero temperature. *J. Stat. Phys.* 111, 1–34. <https://doi.org/10.1023/A:1022221005097>.
35. Yedidia J.S., Freeman W.T., Weiss Y. Understanding belief propagation and its generalizations. Exploring artificial Intelligence in the New Millennium, 8; 2003. p. 236–239.
36. Mezard, M., and Montanari, A. (2009). *Information, Physics, and Computation* (Oxford University Press).
37. Baxter, R.J. (2016). *Exactly Solved Models in Statistical Mechanics* (Elsevier).
38. Mertens, S. (2002). Computational complexity for physicists. *Comput. Sci. Eng.* 4, 31–47. <https://doi.org/10.1109/5992.998639>.
39. Duke, T., and Graham, I. (2009). Equilibrium mechanisms of receptor clustering. *Prog. Biophys. Mol. Biol.* 100, 18–24. <https://doi.org/10.1016/j.pbiomolbio.2009.08.003>.
40. Sarker, R., Mohammadian, M., and Yao, X. (2002). *Evolutionary Optimization*, 48 (Springer Science & Business Media).
41. Kawasaki, K. (1966). Diffusion constants near the critical point for time-dependent Ising models. I. *Phys. Rev.* 145, 224–230. <https://doi.org/10.1103/PhysRev.145.224>.
42. Schupper, N., and Shnerb, N.M. (2005). Inverse melting and inverse freezing: a spin model. *Phys. Rev. E* 72, 046107. <https://doi.org/10.1103/PhysRevE.72.046107>.
43. Gross, D., Kanter, I., and Sompolinsky, H. (1985). Mean-field theory of the Potts glass. *Phys. Rev. Lett.* 55, 304–307. <https://doi.org/10.1103/PhysRevLett.55.304>.
44. Krzakała, F., and Zdeborová, L. (2008). Potts glass on random graphs. *Europhys. Lett.* 81, 57005. <https://doi.org/10.1209/0295-5075/81/57005>.
45. McCarty, J., Delaney, K.T., Danielsen, S.P.O., Fredrickson, G.H., and Shea, J.E. (2019). Complete phase diagram for liquid-liquid phase separation of intrinsically disordered proteins. *J. Phys. Chem. Lett.* 10, 1644–1652. <https://doi.org/10.1021/acs.jpclett.9b00099>.
46. Pal, T., Wessén, J., Das, S., and Chan, H.S. (2021). Subcompartmentalization of polyampholyte species in organelle-like condensates is promoted by charge-pattern mismatch and strong excluded-volume interaction. *Phys. Rev. E* 103, 042406. <https://doi.org/10.1103/PhysRevE.103.042406>.
47. Wessén, J., Das, S., Pal, T., and Chan, H.S. (2022). Analytical formulation and field-theoretic simulation of sequence-specific phase separation of protein-like heteropolymers with short- and long-spatial-range interactions. *J. Phys. Chem. B* 126, 9222–9245. <https://doi.org/10.1021/acs.jpcc.2c06181>.
48. Lin, Y.-H., Wessén, J., Pal, T., Das, S., and Chan, H.S. (2023). Numerical techniques for applications of analytical theories to sequence-dependent phase separations of intrinsically disordered proteins. In *Phase-Separated Biomolecular Condensates* (Springer), pp. 51–94.
49. Binder, K., and Landau, D.P. (1984). Finite-size scaling at first-order phase transitions. *Phys. Rev. B* 30, 1477–1485.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Poly-L-lysine hydrochloride	Alamanda Polymers	PLKC10 (x=10)
Adenosine 5'-diphosphate sodium salt	Sigma Aldrich	A2754
Trizma base	Sigma Aldrich	T1503
Software and algorithms		
KNIME	Open Source	Version 4.3.2 https://www.knime.com/
Automation workflow for mixing 2 components	This paper	
Lattice model numerical simulations	This paper	https://github.com/ondrejtechacek/cavity-phase-sep
Parameter inference optimization procedure	This paper	https://github.com/ondrejtechacek/cavity-phase-sep
Custom image analysis	This paper	https://github.com/ondrejtechacek/cavity-phase-sep
Python 3.9, numpy, scipy, matplotlib, cython	Python Software Foundation	https://www.python.org
Differential Evolution	Storn & Price (1997)	https://doi.org/10.1023/A:1008202821328
Cavity method	Mezard & Montanari (2009) ³⁶	https://doi.org/10.1093/acprof:oso/9780198570837.001.0001
Kawasaki Monte Carlo	Kawasaki (1966) ⁴¹	https://doi.org/10.1103/PhysRev.145.224
Regular solution model	Jacobs & Fraenkel (2017) ²²	https://doi.org/10.1016/j.bpj.2016.10.043
Other		
1536-microwell plate	Greiner bio-one	783096

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Daniele De Martino (daniele.demartino@ehu.eu).

Materials availability

All reagents were obtained from commercial sources and directly used without any further purification. The polymers, poly-lysine of different lengths, were obtained from Alamanda polymers, Inc. Adenosine 5-diphosphate sodium salt (cat. no.: A2754), Adenosine 5-triphosphate disodium salt (cat. no.: A3377), reduced disodium salt hydrate (cat. no.: N8129) were obtained directly from Sigma Aldrich.

Data and code availability

Source code for the parameter inference as well as the Monte Carlo numerical simulation is available online: <https://github.com/ondrejtechacek/cavity-phase-sep>.

METHOD DETAILS

Formulation of the Bethe-Peierls (BP) model

To derive the BP mean field in order to calculate the phase-diagrams, one starts from the Hamiltonian for the system (Equation 1 in the main text). Furthermore, as depicted in Figure 1 in the main text, the lattice is approximated as a tree-graph that branches out from any fixed lattice-site. This allows for the lattice to be decomposed into sub-systems which are only connected through the lattice-site they emerge from, which furthermore means that the Hamiltonian of the system can be decomposed as well:

$$H(\vec{\sigma}) = - \sum_{i \in \mathcal{B}_1} J(\sigma_1, \sigma_i) + \mu(\sigma_1) + H_{-1}$$

where ∂_1 is the neighbourhood of the lattice-site-11 (see again Figure 1 in the main text) and H_{-1} is the Hamiltonian of the system without site-1. Substituting this expression into the partition function (Equation 2 in the main text) results in:

$$Z = \sum_{\vec{\sigma}} e^{-\beta H(\vec{\sigma})} = \sum_{\sigma_1} \sum_{\sigma_2, \dots, \sigma_N} e^{\beta \left[\sum_{i \in \partial_1} J(\sigma_1, \sigma_i) - \mu(\sigma_1) \right]} e^{-\beta H_{-1}}$$

Assuming that there are no interactions between sub-systems, one can write: $H_{-1} = \sum_{i \in \partial_1} H_i$, where H_i is the

Hamiltonian of the sub-system emerging from site- i . Thus the partition function can be further decomposed as:

$$\begin{aligned} Z &= \sum_{\sigma_1} \sum_{\sigma_2, \dots, \sigma_N} e^{\beta \sum_{i \in \partial_1} J(\sigma_1, \sigma_i)} e^{-\beta \mu(\sigma_1)} e^{-\beta \sum_{i \in \partial_1} H_i} \\ &= \sum_{\sigma_1} \sum_{\sigma_2, \dots, \sigma_N} e^{-\beta \mu(\sigma_1)} e^{\beta \sum_{i \in \partial_1} [J(\sigma_1, \sigma_i) - H_i]} \\ &= \sum_{\sigma_1} \sum_{\sigma_2, \dots, \sigma_N} e^{-\beta \mu(\sigma_1)} \prod_{i \in \partial_1} e^{\beta [J(\sigma_1, \sigma_i) - H_i]} \\ &= \sum_{\sigma_1} e^{-\beta \mu(\sigma_1)} \sum_{\sigma_2} \dots \sum_{\sigma_N} e^{\beta [J(\sigma_1, \sigma_2) - H_2]} \dots e^{\beta [J(\sigma_1, \sigma_M) - H_M]} \\ &= \sum_{\sigma_1} e^{-\beta \mu(\sigma_1)} \sum_{\text{sub}_2} e^{\beta [J(\sigma_1, \sigma_2) - H_2]} \dots \sum_{\text{sub}_M} e^{\beta [J(\sigma_1, \sigma_M) - H_M]} \\ &= \sum_{\sigma_1} e^{-\beta \mu(\sigma_1)} \prod_{i \in \partial_1} \underbrace{\sum_{\text{sub}_i} e^{\beta [J(\sigma_1, \sigma_i) - H_i]}}_{Z_{1 \rightarrow i}(\sigma_1)} \end{aligned}$$

Where sub_i is the i -th subsystem, it contains all sites σ_j that emerge from site σ_i . The partition function is then:

$$Z = \sum_{\sigma_1} e^{-\beta \mu(\sigma_1)} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(\sigma_1) \quad (\text{Equation s1})$$

One can further decompose the sub-partition function $Z_{1 \rightarrow i}(\sigma_1) = \sum_{\text{sub}_i} e^{\beta J(\sigma_1, \sigma_i)} e^{-\beta H_i}$. As above the Hamiltonian H_i of the i -th sub-system can be decomposed into further sub-systems (see again Figure 1 of the main text):

$$H_i = - \sum_{j \in \partial_{\lambda 1}} J(\sigma_i, \sigma_j) + \mu(\sigma_i) + H_{-i}$$

where $\partial_{\lambda 1}$ is the neighbourhood of the lattice-site- i without lattice-site-1 (and it's connected sub-systems) and H_{-i} is the Hamiltonian of the sub-system emerging from lattice-site- i . As above, one can assume that there no interactions between sub-systems, thus one can write $H_{-i} = \sum_{j \in \partial_{\lambda 1}} H_j$. The sub-partition function can then be further decomposed as:

$$\begin{aligned} Z_{1 \rightarrow i}(\sigma_1) &= \sum_{\sigma_i} \sum_{\text{sub}_i \setminus \sigma_i} e^{\beta J(\sigma_1, \sigma_i)} e^{\beta \sum_{j \in \partial_{\lambda 1}} J(\sigma_i, \sigma_j)} e^{-\beta \mu(\sigma_i)} e^{-\beta \sum_{j \in \partial_{\lambda 1}} H_j} \\ &= \sum_{\sigma_i} e^{\beta [J(\sigma_1, \sigma_i) - \mu(\sigma_i)]} \sum_{\text{sub}_i \setminus \sigma_i} e^{\beta \sum_{j \in \partial_{\lambda 1}} [J(\sigma_i, \sigma_j) - H_j]} \\ &= \sum_{\sigma_i} e^{\beta [J(\sigma_1, \sigma_i) - \mu(\sigma_i)]} \sum_{\text{sub}_i \setminus \sigma_i} \prod_{j \in \partial_{\lambda 1}} e^{\beta [J(\sigma_i, \sigma_j) - H_j]} \\ &= \sum_{\sigma_i} e^{\beta [J(\sigma_1, \sigma_i) - \mu(\sigma_i)]} \sum_{\text{sub}_i \setminus \sigma_i} \prod_{j \in \partial_{\lambda 1}} e^{\beta [J(\sigma_i, \sigma_j) - H_j]} \\ &= \sum_{\sigma_i} e^{\beta [J(\sigma_1, \sigma_i) - \mu(\sigma_i)]} \prod_{j \in \partial_{\lambda 1}} \underbrace{\sum_{\text{sub}_j} e^{\beta [J(\sigma_i, \sigma_j) - H_j]}}_{Z_{i \rightarrow j}(\sigma_i)} \end{aligned}$$

In other words, the sub-partition function for the sub-systems emerging from site-1 is:

$$Z_{1 \rightarrow i}(\sigma_1) = \sum_{\sigma_i} e^{\beta[J(\sigma_1, \sigma_i) - \mu(\sigma_i)]} \prod_{j \in \partial_{N1}} Z_{i \rightarrow j}(\sigma_i)$$

This can be generalized to:

$$Z_{i \rightarrow j}(\sigma_i) = \sum_{\sigma_j} e^{\beta[J(\sigma_i, \sigma_j) - \mu(\sigma_j)]} \prod_{k \in \partial_{Ni}} Z_{j \rightarrow k}(\sigma_j) \quad (\text{Equation s2})$$

which is Equation 3 in the main text. In other words $Z_{i \rightarrow j}(\sigma_i)$ is the sub-partition function of the sub-system starting from lattice-site- j given that lattice-site- j itself starts from lattice-site- i (i.e. lattice-site- i is fixed). In general we have: $Z_{i \rightarrow j}(\sigma_i) \neq Z_{j \rightarrow i}(\sigma_j)$. In the following sections this partition function will be solved for the special cases of a binary- and a ternary solution.

Binary system

As described in the main text, for a binary solution, i.e. $q = 2$ and $\sigma_i = 0, 1$, one can fix $J(\sigma_i, 0) = \mu(0) = 0$, as well as $J(1, 1) = J, \mu(1) = \mu$. Putting this into Equation s1 one can simplify the partition function:

$$\begin{aligned} Z &= \sum_{\sigma_i = 0,1} e^{-\beta\mu(\sigma_i)} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(\sigma_i) \\ &= \underbrace{e^{-\beta\mu(0)}}_{=1} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(0) + e^{-\beta \overbrace{\mu(1)}^{=\mu}} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(1) \end{aligned}$$

Therefore the partition function for a the binary solution is:

$$Z = \prod_{i \in \partial_1} Z_{1 \rightarrow i}(0) + e^{-\beta\mu} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(1) \quad (\text{Equation s3})$$

In a similar way one can simplify Equation s2:

$$\begin{aligned} Z_{i \rightarrow j}(\sigma_i) &= \sum_{\sigma_j = 0,1} e^{\beta[J(\sigma_i, \sigma_j) - \mu(\sigma_j)]} \prod_{k \in \partial_{Ni}} Z_{j \rightarrow k}(\sigma_j) \\ &= \underbrace{e^{\beta[J(\sigma_i, 0) - \mu(0)]}}_{=1} \prod_{k \in \partial_{Ni}} Z_{j \rightarrow k}(0) + e^{\beta \left[J(\sigma_i, 1) - \overbrace{\mu(1)}^{=\mu} \right]} \prod_{k \in \partial_{Ni}} Z_{j \rightarrow k}(1) \end{aligned}$$

which gives the sub-partition function as:

$$Z_{i \rightarrow j}(\sigma_i) = \prod_{k \in \partial_{Ni}} Z_{j \rightarrow k}(0) + e^{\beta[J(\sigma_i, 1) - \mu]} \prod_{k \in \partial_{Ni}} Z_{j \rightarrow k}(1) \quad (\text{Equation s4})$$

which is Equation 4 in the main text. As further described in the main text, one use the following parametrization: $Z_{i \rightarrow j}(\sigma_i) = A_{i \rightarrow j} e^{\beta u_{i \rightarrow j} \sigma_i}$. One can apply this to the RHS and the LHS of Equation s4 to get:

$$\begin{aligned} A_{i \rightarrow j} e^{\beta u_{i \rightarrow j} \sigma_i} &= \prod_{k \in \partial_{Ni}} A_{j \rightarrow k} e^{\beta u_{j \rightarrow k} \cdot 0} + e^{\beta[J(\sigma_i, 1) - \mu]} \prod_{k \in \partial_{Ni}} A_{j \rightarrow k} e^{\beta u_{j \rightarrow k} \cdot 1} \\ &= \prod_{k \in \partial_{Ni}} A_{j \rightarrow k} \left(1 + e^{\beta[J(\sigma_i, 1) - \mu]} e^{\beta \sum_{k \in \partial_{Ni}} u_{j \rightarrow k}} \right) \end{aligned}$$

Next one can take the fraction $Z_{i \rightarrow j}(1)/Z_{i \rightarrow j}(0)$:

$$\begin{aligned} \frac{A_{i \rightarrow j} e^{\beta u_{i \rightarrow j}}}{A_{i \rightarrow j}} &= \frac{\prod_{k \in \partial_{Ni}} A_{j \rightarrow k} \left(1 + e^{\beta \left[J(1,1) - \mu + \sum_{k \in \partial_{Ni}} u_{j \rightarrow k} \right]} \right)}{\prod_{k \in \partial_{Ni}} A_{j \rightarrow k} \left(1 + e^{\beta \left[J(0,1) - \mu + \sum_{k \in \partial_{Ni}} u_{j \rightarrow k} \right]} \right)} \\ e^{\beta u_{j \rightarrow k}} &= \frac{1 + e^{\beta \left(J - \mu + \sum_{k \in \partial_{Ni}} u_{j \rightarrow k} \right)}}{1 + e^{\beta \left(-\mu + \sum_{k \in \partial_{Ni}} u_{j \rightarrow k} \right)}} \end{aligned}$$

Taking the logarithm on both sides of the above equation results in the following set of self-consistent equations for the messages $u_{i \rightarrow j}$:

$$u_{i \rightarrow j} = \frac{1}{\beta} \log \left(\frac{1 + e^{\beta J - \beta \mu + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}}}{1 + e^{-\beta \mu + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}}} \right) \quad (\text{Equation s5})$$

which is Equation 5 from the main text. To further simplify this set of equations, one can assume that the tree-graph is a Caley-Graph with a branching of $C = K + 1$ and that there is homogeneity of the messages, i.e. $u_{i \rightarrow j} = u, \forall i, j$. Thus one has $\sum_{k \in \partial_{N_i}} u_{j \rightarrow k} = Ku$ as $|\partial_{N_i}| = K$. One therefore retrieves for Equation s5:

$$u = \frac{1}{\beta} \log \left[\frac{1 + e^{\beta(J - \mu + Ku)}}{1 + e^{\beta(-\mu + Ku)}} \right] \quad (\text{Equation s6})$$

which is Equation 6 from the main text. One can furthermore retrieve the probability $P(\sigma_1 = 1)$ of a lattice site being in state 1 (i.e. being occupied by a solute particle):

$$\begin{aligned} P(\sigma_1 = 1) &= \frac{e^{-\beta \mu(1)} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(1)}{Z} \\ &= \frac{e^{-\beta \mu} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(1)}{\prod_{i \in \partial_1} Z_{1 \rightarrow i}(0) + e^{-\beta \mu} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(1)} \equiv \varphi \end{aligned}$$

which is essentially the volume-fraction φ solute particles². Applying the same parametrization from above results in:

$$\begin{aligned} \varphi &= \frac{e^{-\beta \mu} \prod_{i \in \partial_1} A_{1 \rightarrow i} e^{\beta u_{1 \rightarrow i}}}{\prod_{i \in \partial_1} A_{1 \rightarrow i} + e^{-\beta \mu} \prod_{i \in \partial_1} A_{1 \rightarrow i} e^{\beta u_{1 \rightarrow i}}} \\ &= \frac{e^{-\beta \mu + \beta \sum_{i \in \partial_1} u_{1 \rightarrow i}}}{1 + e^{-\beta \mu + \beta \sum_{i \in \partial_1} u_{1 \rightarrow i}}} \\ &= \frac{e^{-\beta \mu + \beta(K+1)u}}{1 + e^{-\beta \mu + \beta(K+1)u}} \end{aligned}$$

Again the assumption that the tree-graph is a Caley-Graph, therefore $|\partial_1| = K + 1$, as well as the homogeneity $u_{1 \rightarrow i} = u, \forall i$, was used. This way one retrieves φ is depending on u and μ :

$$\varphi(u, \mu) = \frac{e^{-\beta \mu + \beta(K+1)u}}{1 + e^{-\beta \mu + \beta(K+1)u}} \quad (\text{Equation s7})$$

which is Equation 7 from the main text. In order to calculate the phase diagram it is necessary to have both φ and $\beta \mu$ parameterized in the following way:

$$\begin{cases} \beta \mu(u) \\ \varphi(u) \end{cases}$$

i.e. φ and $\beta \mu$ as functions of the common parameter u . Starting with $\beta \mu(u)$, one can use Equation s6:

$$\begin{aligned} e^{\beta u} &= \frac{1 + e^{\beta(J - \mu + Ku)}}{1 + e^{\beta(-\mu + Ku)}} \\ e^{\beta u} [1 + e^{\beta(-\mu + Ku)}] &= 1 + e^{\beta(J - \mu + Ku)} \\ e^{\beta u} + e^{\beta[-\mu + (K+1)u]} &= 1 + e^{\beta(J - \mu + Ku)} \\ e^{\beta u} - 1 &= e^{-\beta \mu} e^{\beta Ku} (e^{\beta J} - e^{\beta u}) \\ e^{-\beta \mu} &= \frac{e^{\beta u} - 1}{e^{\beta Ku} (e^{\beta J} - e^{\beta u})} \end{aligned}$$

Taking the logarithm on both sides results in:

$$\beta \mu(u) = -\log(e^{\beta u} - 1) + \beta Ku + \log(e^{\beta J} - e^{\beta u}) \quad (\text{Equation s8})$$

Similarly for $\varphi(u)$ one can use Equation s7:

$$\begin{aligned}\varphi &= \frac{e^{-\beta\mu + \beta(K+1)u}}{1 + e^{-\beta\mu + \beta(K+1)u}} \\ \varphi + \varphi e^{-\beta\mu} e^{\beta(K+1)u} &= e^{-\beta\mu} e^{\beta(K+1)u} \\ \varphi e^{\beta\mu} + \varphi e^{\beta(K+1)u} &= e^{\beta(K+1)u} \\ \varphi e^{\beta\mu} &= (1 - \varphi) e^{\beta(K+1)u} \\ e^{\beta(K+1)u} &= \left(\frac{\varphi}{\varphi - 1}\right) e^{\beta\mu} \\ \beta u &= \frac{1}{K+1} \left[\log\left(\frac{\varphi}{\varphi - 1}\right) + \beta\mu \right]\end{aligned}$$

substituting $\beta\mu$ with Equation s8 gives:

$$\begin{aligned}\beta u &= \frac{1}{K+1} \left[\log\left(\frac{\varphi}{\varphi - 1}\right) - \log(e^{\beta u} - 1) + \beta K u + \log(e^{\beta J} - e^{\beta u}) \right] \\ (K+1)\beta u &= \log\left[\frac{\left(\frac{\varphi}{\varphi - 1}\right)(e^{\beta J} - e^{\beta u})}{e^{\beta u} - 1}\right] + \beta K u \\ e^{K\beta u} e^{\beta u} &= \frac{\left(\frac{\varphi}{\varphi - 1}\right)(e^{\beta J} - e^{\beta u})}{(e^{\beta u} - 1)} e^{\beta K u}\end{aligned}$$

from this one can retrieve:

$$\left(\frac{\varphi}{\varphi - 1}\right) = \frac{e^{\beta u}(e^{\beta u} - 1)}{e^{\beta J} - e^{\beta u}} \quad \text{(Equation s9)}$$

Calculating the spinodal-curve for a binary solution

In order to get the spinodal-curve one usually determines where the first derivative of the chemical potential μ with respect to the volume fraction φ (i.e. second derivative of the free energy F) is zero. As here $\beta\mu(u)$ and $\varphi(u)$ are functions of the parameter u one needs to take:

$$\frac{\partial(\beta\mu)}{\partial\varphi} = \frac{\partial(\beta\mu)}{\partial u} \frac{\partial u}{\partial\varphi} = \frac{\partial(\beta\mu)}{\partial u} \left(\frac{\partial\varphi}{\partial u}\right)^{-1} = \frac{\frac{\partial(\beta\mu)}{\partial u}}{\frac{\partial\varphi}{\partial u}} = 0$$

which basically describes the Nullcline of the curve. This means that it is sufficient to look at the numerator being 0, i.e. $\frac{\partial(\beta\mu)}{\partial u} \rightarrow 0$. Therefore one can use Equation s8:

$$\frac{\partial(\beta\mu)}{\partial u} = -\frac{\beta e^{\beta u}}{e^{\beta u} - 1} + \beta K - \frac{\beta e^{\beta u}}{e^{\beta J} - e^{\beta u}} = 0$$

substituting $w = e^{\beta u}$ one can solve this equation:

$$\begin{aligned}\frac{-\beta w(e^{\beta J} - w) + \beta K(w - 1)(e^{\beta J} - w) - \beta w(w - 1)}{(w - 1)(e^{\beta J} - w)} &= 0 \\ K(w - 1)(e^{\beta J} - 1) &= w(e^{\beta J} - w) + w(w - 1) \\ K(w - 1)(e^{\beta J} - 1) &= w(e^{\beta J} - 1) \\ Kwe^{\beta J} - Ke^{\beta J} - Kw^2 + Kw = we^{\beta J} - w - Kw^2 + (Ke^{\beta J} + k - e^{\beta J} + 1)w - Ke^{\beta J} &= 0\end{aligned}$$

The solution to this quadratic equation yields the following expressions:

$$w_{1,2} = \frac{-(Ke^{\beta J} + K - e^{\beta J} + 1) \pm \sqrt{(Ke^{\beta J} + K - e^{\beta J} + 1)^2 - 4K^2 e^{\beta J}}}{-2K}$$

Collection everything together results in

$$\begin{aligned} \left(\frac{\varphi}{\varphi-1}\right) &= \frac{e^{\beta u}(e^{\beta u}-1)}{e^{\beta J}-e^{\beta u}} = \frac{w(w-1)}{e^{\beta J}-w} \\ w_{1,2} &= \frac{-b \pm \sqrt{b^2 - 4K^2 e^{\beta J}}}{-2K} \\ b &= Ke^{\beta J} + K - e^{\beta J} + 1 \end{aligned} \quad (\text{Equation s10})$$

which is Equation 8 from the main text. In conclusion, to get a spinodal-curve one can chose values for βJ in order to get values for $w_{1,2}$. One then substitutes them into Equation s10, to get values for φ . This way one thus has a function $\varphi(\beta J)$, in order to get $\beta J(\varphi)$ one simply takes the inverse $(\varphi(\beta J))^{-1} = \beta J(\varphi)$.

Ternary system

For a ternary solution with $\sigma_i \in \{-1, 0, +1\}$ one can assume, like in the binary solution, that $J(\sigma_i, 0) = \mu(0) = 0$. Furthermore, as described in the main text, one can do the following relabeling: $J(-1, -1) = J_{--}, J(+1, +1) = J_{++}, J(-1, +1) = J(+1, -1) = J_{+-}, \mu(-1) = \mu_-, \mu(+1) = \mu_+$. Putting this into Equation s1 for the partition function gives:

$$\begin{aligned} Z &= \sum_{\sigma_1 = -1, 0, +1} e^{-\beta \mu(\sigma_1)} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(\sigma_1) \\ &= e^{-\beta \mu(-1)} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(-1) + e^{-\beta \mu(0)} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(0) + e^{-\beta \mu(+1)} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(+1) \end{aligned}$$

Thus one retrieves the partition function for ternary solution as:

$$Z = e^{-\beta \mu_-} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(-1) + \prod_{i \in \partial_1} Z_{1 \rightarrow i}(0) + e^{-\beta \mu_+} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(+1) \quad (\text{Equation s11})$$

In a similar way we one can modify Equation s2:

$$\begin{aligned} Z_{i \rightarrow j}(\sigma_i) &= \sum_{\sigma_j = -1, 0, 1} e^{\beta [J(\sigma_i, \sigma_j) - \mu(\sigma_j)]} \prod_{k \in \partial_{N_i}} Z_{j \rightarrow k}(\sigma_j) \\ &= e^{\beta [J(\sigma_i, -1) - \mu(-1)]} \prod_{k \in \partial_{N_i}} Z_{j \rightarrow k}(-1) + e^{\beta [J(\sigma_i, 0) - \mu(0)]} \prod_{k \in \partial_{N_i}} Z_{j \rightarrow k}(0) \\ &\quad + e^{\beta [J(\sigma_i, +1) - \mu(+1)]} \prod_{k \in \partial_{N_i}} Z_{j \rightarrow k}(+1) \end{aligned}$$

Which gives the sub-partition function for the sub-systems in the ternary lattice as:

$$\begin{aligned} Z_{i \rightarrow j}(\sigma_i) &= e^{\beta [J(\sigma_i, -1) - \mu_-]} \prod_{k \in \partial_{N_i}} Z_{j \rightarrow k}(-1) + \prod_{k \in \partial_{N_i}} Z_{j \rightarrow k}(0) \\ &\quad + e^{\beta [J(\sigma_i, +1) - \mu_+]} \prod_{k \in \partial_{N_i}} Z_{j \rightarrow k}(+1) \end{aligned} \quad (\text{Equation s12})$$

which is Equation 10 in the main text. Like in the binary solution one can introduce a parametrization:

$Z_{i \rightarrow j}(\sigma_i) = A_{i \rightarrow j} e^{\beta (v_{i \rightarrow j} \sigma_i + w_{i \rightarrow j} \sigma_i^2)}$. Applying this to the RHS and the LHS of Equation s12 gives:

$$\begin{aligned} A_{i \rightarrow j} e^{\beta (v_{i \rightarrow j} \sigma_i + w_{i \rightarrow j} \sigma_i^2)} &= e^{\beta [J(\sigma_i, -1) - \mu_-]} \prod_{k \in \partial_{N_i}} A_{j \rightarrow k} e^{\beta (-v_{j \rightarrow k} + w_{j \rightarrow k})} + \prod_{k \in \partial_{N_i}} A_{j \rightarrow k} \\ &\quad + e^{\beta [J(\sigma_i, +1) - \mu_+]} \prod_{k \in \partial_{N_i}} A_{j \rightarrow k} e^{\beta (v_{j \rightarrow k} + w_{j \rightarrow k})} \\ &= \prod_{k \in \partial_{N_i}} A_{j \rightarrow k} \left(e^{\beta [J(\sigma_i, -1) - \mu_-]} e^{\beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^-} + 1 \right. \\ &\quad \left. + e^{\beta [J(\sigma_i, +1) - \mu_+]} e^{\beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+} \right) \end{aligned}$$

As in the binary solution one can take the fraction $Z_{i \rightarrow j}(-1)/Z_{i \rightarrow j}(0)$:

$$\frac{A_{i \rightarrow j} e^{\beta u_{j \rightarrow i}^-}}{A_{i \rightarrow j}} = \frac{\prod_{k \in \partial_{N_i}} A_{j \rightarrow k} \left(e^{\beta \left[J_{--} - \mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right]} + 1 + e^{\beta \left[J_{+-} - \mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right]} \right)}{\prod_{k \in \partial_{N_i}} A_{j \rightarrow k} \left(e^{\beta \left[-\mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right]} + 1 + e^{\beta \left[-\mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right]} \right)}$$

$$e^{\beta u_{j \rightarrow i}^-} = \frac{e^{\beta \left(J_{--} - \mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right)} + 1 + e^{\beta \left(J_{+-} - \mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right)}}{e^{\beta \left(-\mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right)} + 1 + e^{\beta \left(-\mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right)}}$$

The same way one can get for $Z_{i \rightarrow j}(+1)/Z_{i \rightarrow j}(0)$:

$$\frac{A_{i \rightarrow j} e^{\beta u_{j \rightarrow i}^+}}{A_{i \rightarrow j}} = \frac{\prod_{k \in \partial_{N_i}} A_{j \rightarrow k} \left(e^{\beta \left[J_{+-} - \mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right]} + 1 + e^{\beta \left[J_{++} - \mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right]} \right)}{\prod_{k \in \partial_{N_i}} A_{j \rightarrow k} \left(e^{\beta \left[-\mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right]} + 1 + e^{\beta \left[-\mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right]} \right)}$$

$$e^{\beta u_{j \rightarrow i}^+} = \frac{e^{\beta \left(J_{+-} - \mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right)} + 1 + e^{\beta \left(J_{++} - \mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right)}}{e^{\beta \left(-\mu_- + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^- \right)} + 1 + e^{\beta \left(-\mu_+ + \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+ \right)}}$$

These are again the messages for the ternary system. Taking the logarithm on both sides of the above equations one can retrieve get the formulas:

$$u_{i \rightarrow j}^- = \frac{1}{\beta} \log \left(\frac{e^{\beta J_{--} - \beta \mu_- + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^-} + 1 + e^{\beta J_{+-} - \beta \mu_+ + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+}}{e^{-\beta \mu_- + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^-} + 1 + e^{-\beta \mu_+ + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+}} \right) \quad \text{(Equation s13)}$$

$$u_{i \rightarrow j}^+ = \frac{1}{\beta} \log \left(\frac{e^{\beta J_{+-} - \beta \mu_- + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^-} + 1 + e^{\beta J_{++} - \beta \mu_+ + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+}}{e^{-\beta \mu_- + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^-} + 1 + e^{-\beta \mu_+ + \beta \sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^+}} \right)$$

Again, one can assume a Caley-Graph with a branching of $C = K + 1$ and $u_{i \rightarrow j}^\pm = u^\pm, \forall i, j$. Thus one has $\sum_{k \in \partial_{N_i}} u_{j \rightarrow k}^\pm = K u^\pm$ as $|\partial_{N_i}| = K$. Equation s13 then simplifies to:

$$u^- = \frac{1}{\beta} \log \left[\frac{e^{\beta(J_{--} - \mu_- + K u^-)} + 1 + e^{\beta(J_{+-} - \mu_+ + K u^+)}}{e^{\beta(-\mu_- + K u^-)} + 1 + e^{\beta(-\mu_+ + K u^+)}} \right] \quad \text{(Equation s14)}$$

$$u^+ = \frac{1}{\beta} \log \left[\frac{e^{\beta(J_{+-} - \mu_- + K u^-)} + 1 + e^{\beta(J_{++} - \mu_+ + K u^+)}}{e^{\beta(-\mu_- + K u^-)} + 1 + e^{\beta(-\mu_+ + K u^+)}} \right]$$

which is Equation 11 in the main text. As in the binary solution one can calculate the probabilities $P(\sigma_1 = -1), P(\sigma_1 = +1)$ of a lattice site being in the state $-1, +1$ (i.e. being occupied by a solute particle of the $-$ type or $+$ type):

$$P(\sigma_1 = -1) = \frac{e^{-\beta \mu_-} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(-1)}{e^{-\beta \mu_-} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(-1) + \prod_{i \in \partial_1} Z_{1 \rightarrow i}(0) + e^{-\beta \mu_+} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(+1)} \equiv \varphi_-$$

$$P(\sigma_1 = +1) = \frac{e^{-\beta \mu_+} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(+1)}{e^{-\beta \mu_-} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(-1) + \prod_{i \in \partial_1} Z_{1 \rightarrow i}(0) + e^{-\beta \mu_+} \prod_{i \in \partial_1} Z_{1 \rightarrow i}(+1)} \equiv \varphi_+$$

Which are essentially the volume-fractions φ_+ , φ_- for the solute particles³. Applying the parametrization from above results in:

$$\begin{aligned}\varphi_{\pm} &= \frac{e^{-\beta\mu_{\pm}} \prod_{i \in \partial_1} A_{1i} e^{\beta u_{1 \rightarrow i}^{\pm}}}{e^{-\beta\mu_-} \prod_{i \in \partial_1} A_{1i} e^{\beta u_{1 \rightarrow i}^-} + \prod_{i \in \partial_1} A_{1i} + e^{-\beta\mu_+} \prod_{i \in \partial_1} A_{1i} e^{\beta u_{1 \rightarrow i}^+}} \\ &= \frac{e^{-\beta\mu_{\pm}} \prod_{i \in \partial_1} e^{\beta u_{1 \rightarrow i}^{\pm}}}{e^{-\beta\mu_-} \prod_{i \in \partial_1} e^{\beta u_{1 \rightarrow i}^-} + 1 + e^{-\beta\mu_+} \prod_{i \in \partial_1} e^{\beta u_{1 \rightarrow i}^+}} \\ &= \frac{e^{-\beta\mu_{\pm}} e^{\beta \sum_{i \in \partial_1} u_{1 \rightarrow i}^{\pm}}}{e^{-\beta\mu_-} e^{\beta \sum_{i \in \partial_1} u_{1 \rightarrow i}^-} + 1 + e^{-\beta\mu_+} e^{\beta \sum_{i \in \partial_1} u_{1 \rightarrow i}^+}} \\ &= \frac{e^{-\beta\mu_{\pm}} e^{\beta(K+1)u^{\pm}}}{e^{-\beta\mu_-} e^{\beta(K+1)u^-} + 1 + e^{-\beta\mu_+} e^{\beta(K+1)u^+}} \equiv \varphi_+\end{aligned}$$

assuming again a Caley-Graph with $|\partial_1| = K + 1$ and $u_{1 \rightarrow i}^{\pm} = u^{\pm}, \forall i$. This way one retrieves φ_-, φ_+ in a ternary-solution as:

$$\begin{aligned}\varphi_- &= \frac{e^{-\beta\mu_- + \beta(K+1)u^-}}{e^{-\beta\mu_- + \beta(K+1)u^-} + 1 + e^{-\beta\mu_+ + \beta(K+1)u^+}} \\ \varphi_+ &= \frac{e^{-\beta\mu_+ + \beta(K+1)u^+}}{e^{-\beta\mu_- + \beta(K+1)u^-} + 1 + e^{-\beta\mu_+ + \beta(K+1)u^+}}\end{aligned} \quad \text{(Equation s15)}$$

which is Equation 12 from the main text.

Calculating the spinodal-curve for a ternary solution

In order to get the spinodal-curve for a ternary system a small “workaround” is needed. First, one can define the following variable transformation

$$m_{\pm} = -\mu_{\pm} + Cu^{\pm}$$

(recall that for our tree-graph one has $C = K + 1$). One thus gets for Equation s15:

$$\begin{aligned}\varphi_{\pm} &= \frac{e^{\beta m_{\pm}}}{1 + e^{\beta m_-} + e^{\beta m_+}} \\ m_{\pm} &= \log[\varphi_{\pm} (1 + e^{\beta m_-} + e^{\beta m_+})]\end{aligned}$$

Given φ_{\pm} one can numerically solve this self-consistent equation to get an estimator for m_{\pm} . In a similar way one can use the variable transformation from above on Equation s14 (recall $K = C - 1$):

$$u^{\pm} = \frac{1}{\beta} \log \left[\frac{e^{\beta(J_{\pm\pm} + m_{\pm} - u^{\pm})} + e^{\beta(J_{+-} + m_{\mp} - u^{\mp})} + 1}{e^{\beta(m_- - u^-)} + e^{\beta(m_+ - u^+)} + 1} \right]$$

Together with the estimate for m_{\pm} one can again numerically solve this self consistent equation to get an estimate for u^{\pm} . Finally one can get μ_{\pm} from the variable transformation together with the estimates for m_{\pm} and u^{\pm} :

$$\mu_{\pm} = -m_{\pm} + Cu^{\pm}$$

This way one retrieves an estimate for the external fields μ_{\pm} in every point (φ_+, φ_-) . to get the spinodal-curve one then needs to take the Hessian-matrix:

$$H = \begin{pmatrix} \frac{\partial \mu_+}{\partial \varphi_+} & \frac{\partial \mu_+}{\partial \varphi_-} \\ \frac{\partial \mu_-}{\partial \varphi_+} & \frac{\partial \mu_-}{\partial \varphi_-} \end{pmatrix}$$

(recall that $\mu_{\pm} = \frac{\partial F}{\partial \varphi_{\pm}}$, so $\frac{\partial \mu_{\pm}}{\partial \varphi_{\pm}} = \frac{\partial^2 F}{\partial \varphi_{\pm} \partial \varphi_{\pm}}$). One can compute the derivatives numerically. With the hessian one can use the property that if $\text{tr}(H) > 0$ and $\det(H) > 0$ the matrix is positive-definite meaning that the free energy F is concave-upward which means that there is no phase-separation. Thus for all points (φ_+, φ_-) where this is not the case we know that there is phase-separation happening.

Formulation of the regular solution model

To derive the mean field equations of the regular solution model one again starts from the Hamiltonian of the system (Equation 1 in the main text). Again the lattice is approximated as a graph, however in contrast to the BP-model one assumes a fully connected graph, rather than a tree. One can furthermore substitute the interaction function $J(\sigma_i, \sigma_j)$ and the chemical potential $\mu(\sigma_i)$ with a symmetric matrix and a vector respectively, i.e.: $J(\sigma_i, \sigma_j) = J_{\sigma_i \sigma_j}, \mu(\sigma_i) = \mu_{\sigma_i}$. As the lattice is now fully connected the pairwise sum runs now over all lattice sites. This means one can rewrite the Hamiltonian to:

$$H(\vec{\sigma}) = -\frac{1}{2N} \sum_{i=1}^N \sum_{j=1}^N J_{\sigma_i \sigma_j} + \sum_{i=1}^N \mu_{\sigma_i}$$

One can further rewrite $J_{\sigma_i \sigma_j}$ and μ_{σ_i} to:

$$J_{\sigma_i \sigma_j} = \sum_{k=0}^{q-1} \sum_{l=0}^{q-1} J_{kl} \delta_{k\sigma_i} \delta_{l\sigma_j}, \mu_{\sigma_i} = \sum_{k=0}^{q-1} \mu_k \delta_{k\sigma_i}$$

with $0, \dots, q-1$ being the possible states for σ_i (q possible states overall) and $\delta_{k\sigma_i}$ being the Kronecker-delta. Substituting this back into the Hamiltonian yields:

$$\begin{aligned} H(\vec{\sigma}) &= -\frac{1}{2N} \sum_{i,j=1}^N \sum_{k,l=0}^{q-1} J_{kl} \delta_{k\sigma_i} \delta_{l\sigma_j} + \sum_{i=1}^N \sum_{k=0}^{q-1} \mu_k \delta_{k\sigma_i} \\ &= \frac{1}{2N} \sum_{k,l=0}^{q-1} J_{kl} \sum_{i,j=1}^N \delta_{k\sigma_i} \delta_{l\sigma_j} + \sum_{k=0}^{q-1} \mu_k \sum_{i=1}^N \delta_{k\sigma_i} \\ &= \frac{1}{2N} \sum_{k,l=0}^{q-1} J_{kl} \underbrace{\left(\sum_{i=1}^N \delta_{k\sigma_i} \right)}_{= N_k} \underbrace{\left(\sum_{j=1}^N \delta_{l\sigma_j} \right)}_{= N_l} + \sum_{k=0}^{q-1} \mu_k \underbrace{\sum_{i=1}^N \delta_{k\sigma_i}}_{= N_k} \end{aligned}$$

with N_k being the number lattice sites that are in state k , i.e. the number of type- k molecules. Thus the Hamiltonian of the regular solution model is:

$$H(\vec{\sigma}) = -\frac{1}{2N} \sum_{k=0}^{q-1} \sum_{l=0}^{q-1} J_{kl} N_k N_l + \sum_{k=0}^{q-1} \mu_k N_k \quad (\text{Equation s16})$$

Again one can insert Equation s16 into the partition function (Eq. 2 from the main text) to get:

$$Z = \sum_{\sigma_1, \dots, \sigma_N} e^{\beta \left[\frac{1}{2N} \sum_{k,l} J_{kl} N_k N_l - \sum_k \mu_k N_k \right]}$$

This sum is running over the state vector $\vec{\sigma} = (\sigma_1, \dots, \sigma_i, \dots, \sigma_N)$ (i.e. the vector representing all the state of all the lattice-sites). Having transformed the exponent to be dependent on N_k rather than σ_i one would also want the sum to be running over the state vector $\vec{N} = (N_0, \dots, N_k, \dots, N_{q-1})$ which is the vector representing the amount of lattice-sites that are in a certain state k . In general a certain state vector \vec{N} can be represented by multiple vectors $\vec{\sigma}$. This multiplicity is accounted for by further introducing a multinomial coefficient in the sum. One can therefore rewrite the partition function to:

$$Z = \sum_{N_0, \dots, N_q} \binom{N}{N_0 \dots N_{q-1}} e^{\beta \left[\frac{1}{2N} \sum_{k,l} J_{kl} N_k N_l - \sum_k \mu_k N_k \right]}$$

One can use Stirling-approximation, i.e. $N! \approx e^{N(\log N - 1)}$, to do the following simplification:

$$\binom{N}{N_0 \dots N_{q-1}} = \frac{N!}{N_0! \dots N_{q-1}!} \approx e^{N(\log N - 1) - \sum_k N_k (\log N_k - 1)} = e^{N \log N - \sum_k N_k \log N_k}$$

Substituting this back into the partition function yields:

$$Z = \sum_{\{N_k\}} \exp \left[\beta \left(\frac{1}{2N} \sum_{k,l=0}^{q-1} J_{kl} N_k N_l - \sum_{k=0}^{q-1} \mu_k N_k \right) + N \log N - \sum_{k=0}^{q-1} N_k \log N_k \right] = \sum_{\{N_k\}} e^{f(\vec{N})}$$

which is a partition sum over the state space $\{N_k\}$ with the constraint that $\sum N_k = N$. Since N_k basically represents the amount of type- k molecules in the system, one has: $\varphi_k = \frac{N_k}{N}$, with φ_k being the volume fraction of type- k molecules, with $\sum \varphi_k = 1$. Thus the state space $\{\varphi_k\}$ is essentially defined by a simplex. One can generalize the above partition sum to an integral over this simplex:

$$Z = \int_{\{\varphi_k\}} e^{-\beta NF(\vec{\varphi})} d\vec{\varphi}$$

With $F(\vec{\varphi})$ being the free energy of the system. Comparing this integral with the partition sum one can see that $f(\vec{N}) = -\beta NF(\vec{\varphi})$. Therefore one can rewrite the exponent in the partition sum from above in order to get $F(\vec{\varphi})$:

$$\begin{aligned} f(\vec{N}) &= \beta \left(\frac{1}{2N} \sum_{k,l=0}^{q-1} J_{kl} N_k N_l - \sum_{k=0}^{q-1} \mu_k N_k \right) + N \log N - \sum_{k=0}^{q-1} N_k \log N_k \\ &= -\beta N \left[-\frac{1}{2N^2} \sum_{k,l=0}^{q-1} J_{kl} N_k N_l + \frac{1}{N} \sum_{k=0}^{q-1} \mu_k N_k - \frac{1}{\beta} \log N + \frac{1}{\beta N} \sum_{k=0}^{q-1} N_k \log N_k \right] \\ &= -\beta N \left[-\frac{1}{2} \sum_{k,l=0}^{q-1} J_{kl} \underbrace{\frac{N_k}{N} \frac{N_l}{N}}_{=\varphi_k = \varphi_l} + \sum_{k=0}^{q-1} \mu_k \underbrace{\frac{N_k}{N}}_{\varphi_k} + \frac{1}{\beta} \left(\sum_{k=0}^{q-1} \frac{N_k}{N} \log N_k - \log N \right) \right] \\ &= -\beta N \left[-\frac{1}{2} \sum_{k,l=0}^{q-1} J_{kl} \varphi_k \varphi_l + \sum_{k=0}^{q-1} \mu_k \varphi_k + \frac{1}{\beta} \left(\sum_{k=0}^{q-1} \varphi_k \log N_k - \underbrace{\log N}_{=\sum_k \frac{N_k}{N} \log N} \right) \right] \\ &= -\beta N \left[-\frac{1}{2} \sum_{k,l=0}^{q-1} J_{kl} \varphi_k \varphi_l + \sum_{k=0}^{q-1} \mu_k \varphi_k + \frac{1}{\beta} \left(\sum_{k=0}^{q-1} \varphi_k (\log N_k - \underbrace{\log N}_{=\log \frac{N_k}{N}}) \right) \right] \\ &= -\beta N \left[-\frac{1}{2} \sum_{k,l=0}^{q-1} J_{kl} \varphi_k \varphi_l + \sum_{k=0}^{q-1} \mu_k \varphi_k + \frac{1}{\beta} \sum_{k=0}^{q-1} \varphi_k \log \varphi_k \right] \end{aligned}$$

Using the relation $f(\vec{N}) = -\beta NF(\vec{\varphi})$ from above one retrieves:

$$F(\vec{\varphi}) = -\frac{1}{2} \sum_{k,l=0}^{q-1} J_{kl} \varphi_k \varphi_l + \sum_{k=0}^{q-1} \mu_k \varphi_k + \frac{1}{\beta} \sum_{k=0}^{q-1} \varphi_k \log \varphi_k \quad (\text{Equation s17})$$

which is the free energy of the regular solution model.

Binary system

In a binary system one has $\sigma_i = 0, 1$ i.e. $q = 2$. One can assume for the interactions and the chemical potential: $J_{00} = J_{01} = J_{10} = 0, \mu_0 = 0$ as well as $J_{11} = J, \mu_1 = \mu$. Furthermore for the volume fraction one has: $\varphi_1 = \varphi$ and $\varphi_0 = 1 - \varphi$. This means that Equation s17 simplifies to:

$$F = -\frac{J\varphi^2}{2} + \mu\varphi + \frac{1}{\beta} [\varphi \log \varphi + (1 - \varphi) \log(1 - \varphi)] \quad (\text{Equation s18})$$

One can get the chemical potential by taking the derivative $\frac{\partial F}{\partial \varphi} = 0$:

$$\frac{\partial F}{\partial \varphi} = -J\varphi + \mu + \frac{1}{\beta} [\log \varphi - \log(1 - \varphi)] = 0$$

From this one retrieves the chemical potential as:

$$\beta\mu = \beta J\varphi - \log \varphi + \log(1 - \varphi) \quad (\text{Equation s19})$$

Ternary system

In a ternary system one has $\sigma_i = -1, 0, +1$ i.e. $q = 3$. One can assume for the interactions and the chemical potential: $J_{0,0} = J_{0,-1} = J_{-1,0} = J_{0,+1} = J_{+1,0} = 0, \mu_0 = 0$ as well as $J_{+1,+1} = J_{++}, J_{+1,-1} = J_{-1,+1} = J_{+-}, J_{-1,-1} = J_{--}, \mu_{-1} = \mu_-, \mu_{+1} = \mu_+$. Furthermore for the volume fraction one has: $\varphi_{-1} = \varphi_-, \varphi_{+1} = \varphi_+$ and $\varphi_0 = 1 - \varphi_- - \varphi_+$. This means that Equation s17 simplifies to:

$$F = -\frac{J_{--}\phi_-^2}{2} - J_{+-}\phi_-\phi_+ - \frac{J_{++}\phi_+^2}{2} + \mu_-\phi_- + \mu_+\phi_+ + \frac{1}{\beta}[\phi_-\log\phi_- + \phi_+\log\phi_+ + (1 - \phi_- - \phi_+)\log(1 - \phi_- - \phi_+)] \quad (\text{Equation s20})$$

One can get the chemical potentials by taking the derivative $\frac{\partial F}{\partial \varphi_{\pm}} = 0$:

$$\frac{\partial F}{\partial \varphi_{\pm}} = -J_{\pm\pm}\varphi_{\pm} - J_{+-}\varphi_{\mp} + \mu_{\pm} + \frac{1}{\beta}[\log\varphi_{\pm} - \log(1 - \varphi_- - \varphi_+)] = 0$$

From this one retrieves the chemical potentials as:

$$\begin{aligned} \beta\mu_- &= \beta J_{--}\varphi_- + \beta J_{+-}\varphi_+ - \log\varphi_- + \log(1 - \varphi_- - \varphi_+) \\ \beta\mu_+ &= \beta J_{++}\varphi_+ + \beta J_{+-}\varphi_- - \log\varphi_+ + \log(1 - \varphi_- - \varphi_+) \end{aligned} \quad (\text{Equation s21})$$

Retrieving the phase diagrams for a binary-system using the Currie-Weiss (CW) model

The Currie-Weiß (CW) model is another approach in order to calculate the phase-diagram⁶. Like with the BP-model one starts with the Hamiltonian of the system (Equation 1 in the main text). In a binary-solution with $\sigma_i = 0, 1$ one can substitute $J(\sigma_i, \sigma_j) = J\sigma_i\sigma_j$ and $\mu(\sigma_i) = \mu\sigma_i$ which results in the Hamiltonian:

$$H(\vec{\sigma}) = -J\sum_{\langle i,j \rangle}\sigma_i\sigma_j + \mu\sum_i\sigma_i \quad (\text{Equation s22})$$

For the CW-model one make the following approximation:

$$\sigma_i = \langle \sigma \rangle + \delta_i$$

with $\delta_i \ll 1$. In other words, the state at lattice site- i is approximated with an average value $\langle \sigma \rangle$ and a small deviation δ_i . This way one can simplify the pairwise-sum from Equation s22:

$$\begin{aligned} \sum_{\langle i,j \rangle}\sigma_i\sigma_j &= \sum_{\langle i,j \rangle}(\langle \sigma \rangle + \delta_i) \cdot (\langle \sigma \rangle + \delta_j) \\ &= \sum_{\langle i,j \rangle} \left[\langle \sigma \rangle^2 + \left(\underbrace{\delta_i}_{=\sigma_i - \langle \sigma \rangle} + \underbrace{\delta_j}_{=\sigma_j - \langle \sigma \rangle} \right) \langle \sigma \rangle + \underbrace{\delta_i\delta_j}_{\approx 0} \right] \\ &\approx \sum_{\langle i,j \rangle} [\langle \sigma \rangle^2 + (\sigma_i - \langle \sigma \rangle + \sigma_j - \langle \sigma \rangle) \langle \sigma \rangle] \\ &= \sum_{\langle i,j \rangle} [\sigma_i \langle \sigma \rangle + \sigma_j \langle \sigma \rangle - \langle \sigma \rangle^2] \\ &= \langle \sigma \rangle \sum_{\langle i,j \rangle} \sigma_i + \langle \sigma \rangle \sum_{\langle i,j \rangle} \sigma_j - \langle \sigma \rangle^2 \sum_{\langle i,j \rangle} 1 \\ &= \langle \sigma \rangle \frac{1}{2} \sum_{i=1}^N \sigma_i \underbrace{\sum_{j \in N_i} 1}_{=k+1} + \langle \sigma \rangle \frac{1}{2} \sum_{j=1}^N \sigma_j \underbrace{\sum_{i \in N_j} 1}_{=k+1} - \langle \sigma \rangle^2 \frac{1}{2} \sum_{i=1}^N \sum_{j \in N_i} 1 \\ &= \langle \sigma \rangle \frac{k+1}{2} \sum_i \sigma_i + \langle \sigma \rangle \frac{k+1}{2} \sum_j \sigma_j - \langle \sigma \rangle^2 \frac{N(k+1)}{2} \\ &= \langle \sigma \rangle (k+1) \sum_i \sigma_i - \langle \sigma \rangle^2 \frac{N(k+1)}{2} \end{aligned}$$

Substituting this back into the Hamiltonian gives:

$$\begin{aligned} H &= -J \left[\langle \sigma \rangle (k+1) \sum_i \sigma_i - \langle \sigma \rangle^2 \frac{N(k+1)}{2} \right] + \mu \sum_i \sigma_i \\ &= -J \langle \sigma \rangle (k+1) \sum_i \sigma_i + J \langle \sigma \rangle^2 \frac{N(k+1)}{2} + \mu \sum_i \sigma_i \end{aligned}$$

This way one can retrieve the Hamiltonian for the CW-model as:

$$H(\vec{\sigma}) = J \langle \sigma \rangle^2 \frac{N(k+1)}{2} - [J \langle \sigma \rangle (k+1) - \mu] \sum_i \sigma_i \quad (\text{Equation s23})$$

Inserting Equation s23 into the partition function (Equation 2 from the main text) yields:

$$\begin{aligned} Z &= \sum_{\sigma_1 \dots \sigma_N} e^{-\beta J \langle \sigma \rangle^2 \frac{N(k+1)}{2}} e^{\beta [J \langle \sigma \rangle (k+1) - \mu] \sum_i \sigma_i} \\ &= e^{-\beta J \langle \sigma \rangle^2 \frac{N(k+1)}{2}} \sum_{\sigma_1 \dots \sigma_N} \prod_i e^{\beta [J \langle \sigma \rangle (k+1) - \mu] \sigma_i} \\ &= e^{-\beta J \langle \sigma \rangle^2 \frac{N(k+1)}{2}} \sum_{\sigma_1} \dots \sum_{\sigma_N} e^{\beta [J \langle \sigma \rangle (k+1) - \mu] \sigma_1} \dots e^{\beta [J \langle \sigma \rangle (k+1) - \mu] \sigma_N} \\ &= e^{-\beta J \langle \sigma \rangle^2 \frac{N(k+1)}{2}} \underbrace{\sum_{\sigma_1=0,1} e^{\beta [J \langle \sigma \rangle (k+1) - \mu] \sigma_1}}_{(1 + e^{\beta [J \langle \sigma \rangle (k+1) - \mu]})} \dots \underbrace{\sum_{\sigma_N=0,1} e^{\beta [J \langle \sigma \rangle (k+1) - \mu] \sigma_N}}_{(1 + e^{\beta [J \langle \sigma \rangle (k+1) - \mu]})} \end{aligned}$$

Thus resulting in the partition function for the CW-model:

$$Z = e^{-\beta J \langle \sigma \rangle^2 \frac{N(k+1)}{2}} (1 + e^{\beta [J \langle \sigma \rangle (k+1) - \mu]})^N \quad (\text{Equation s24})$$

Taking the logarithm on both sides yields:

$$\log Z = N \left[\log(1 + e^{\beta [J \langle \sigma \rangle (k+1) - \mu]}) - \beta J \langle \sigma \rangle^2 \frac{(k+1)}{2} \right] \quad (\text{Equation s25})$$

Looking at the derivative of $\log Z$ with respect to $\beta\mu$ for the general expression $Z = \sum_{\sigma_i} e^{\beta f_{int}} e^{-\beta\mu \sum \sigma_i}$:

$$\begin{aligned} \frac{\partial \log Z}{\partial(\beta\mu)} &= \frac{1}{Z} \frac{\partial Z}{\partial(\beta\mu)} = -\frac{1}{Z} \sum_{\sigma_i} e^{\beta f_{int}} \sum_i \sigma_i e^{-\beta\mu \sum \sigma_i} \\ &= -\frac{1}{Z} \sum_i \sum_{\sigma_i} \sigma_i e^{\beta f_{int}} e^{-\beta\mu \sum \sigma_i} \\ &= -\sum_i \underbrace{\langle \sigma_i \rangle}_{=\langle \sigma \rangle} = N \langle \sigma \rangle \end{aligned}$$

In other words one retrieves the following relation:

$$\langle \sigma \rangle = -\frac{1}{N} \frac{\partial \log Z}{\partial(\beta\mu)}$$

In general one can substitute $\langle \sigma \rangle = \varphi$, i.e. the average site occupancy corresponds to the volume fraction of the solute. Applying the above relation to the logarithm of Equation s25 gives:

$$\begin{aligned} \varphi &= -\frac{1}{N} \frac{\partial}{\partial(\beta\mu)} N \left[\log(1 + e^{\beta [J \varphi (k+1) - \mu]}) - \beta J \varphi^2 \frac{(k+1)}{2} \right] \\ &= -\frac{\partial}{\partial(\beta\mu)} \log(1 + e^{\beta [J \varphi (k+1) - \mu]}) \\ &= \frac{e^{\beta [J \varphi (k+1) - \mu]}}{1 + e^{\beta [J \varphi (k+1) - \mu]}} \end{aligned}$$

Thus one retrieves for the volume fraction φ

$$\varphi(\beta\mu) = \frac{e^{\beta J\varphi(k+1)} e^{-\beta\mu}}{1 + e^{\beta J\varphi(k+1)} e^{-\beta\mu}}$$

One can solve this equation now for $\beta\mu$

$$\begin{aligned} \varphi(1 + e^{\beta J\varphi(k+1)} e^{-\beta\mu}) &= e^{\beta J\varphi(k+1)} e^{-\beta\mu} \\ \varphi + \varphi e^{\beta J\varphi(k+1)} e^{-\beta\mu} &= e^{\beta J\varphi(k+1)} e^{-\beta\mu} \\ e^{\beta J\varphi(k+1)} e^{-\beta\mu} (1 - \varphi) &= \varphi \\ e^{\beta J\varphi(k+1)} e^{-\beta\mu} &= \frac{\varphi}{1 - \varphi} \\ \beta J\varphi(k+1) - \beta\mu &= \log \varphi - \log(1 - \varphi) \end{aligned}$$

Thus one retrieves the following formula for $\beta\mu$:

$$\beta\mu = \log(1 - \varphi) - \log \varphi + \beta J\varphi(k+1) \quad (\text{Equation s26})$$

Calculating the spinodal-curve

As with the BP-model, the spinodal-curve can be determined by setting $\frac{\partial(\beta\mu)}{\partial\varphi} = 0$, i.e.:

$$\begin{aligned} \frac{\partial(\beta\mu)}{\partial\varphi} &= -\frac{1}{1 - \varphi} - \frac{1}{\varphi} + \beta J(k+1) = 0 \\ \beta J(k+1) &= \frac{1}{\varphi} + \frac{1}{1 - \varphi} \\ \beta J(k+1) &= \frac{1}{\varphi(1 - \varphi)} \end{aligned}$$

Thus one has for the spinodal-curve:

$$\beta J = \frac{1}{(k+1)} \frac{1}{\varphi(1 - \varphi)} \quad (\text{Equation s27})$$

which is [Equation 9](#) from the main text.

Calculating the binodal-curve

The binodal curve has been calculated through the Maxwell construction. We report an example here below in [Figure S1](#) (BP for the binary system with $K = 3, \beta = 3$). This is obtained by the standard procedure of finding the point where the chemical potentials of the two phases are equal. Upon looking in [Figure S1](#) (top) this geometrically corresponds to find the horizontal line whose intersection with the mean-field isothermal curve gives regions of equal area. We report here in [Figure S1](#) (bottom) the binodal curves for the binary system with both approximations (CW and BP) alongside with the spinodals.

Finite size scaling

The systems we are considering here have short-range interactions and finite number of components (we are not dealing with a disordered systems) and thus, in regard to the study of the peak of specific heat as a function of the temperature, finite size effects shall follow established results of classical finite size scaling theory.⁴⁹ In particular, in our case, the rounding and shifting of the transition point is ruled by terms of the order $O(L^{-d})$ (L being the lattice size and d the dimension), as we show below in [Figure S2](#)

Calibration of the image processing algorithm

[Figures 2](#) and [3](#) (main text) employ a novel image processing algorithm to detect phase separation in snapshots of numerical simulations. The algorithm is explained in the text and illustrated by the [Figure 6](#) (main text). Here we elaborate on the algorithm and describe the required calibration procedure.

The image processing algorithm defines a measure of phase separation ξ , a continuous variable from 0 (well mixed) to 1 (phase separated). A threshold value needs to be calibrated based on comparison

with numerical simulations of different systems along a simulated annealing pathway (β slowly increasing). Then, the method can be used for binary classification (mixed/separated). The [Figure S3](#) shows the measure of phase separation ξ compared to the specific heat, as a part of the calibration procedure.

Further details on numerical simulations

In all simulations, the lattice was initialized randomly with a uniform distribution of all components. In case of the results presented in the [Figure 2](#) (main text), the system was slowly evolved along a simulation pathway with the inverse thermodynamic temperature β increasing linearly from 0 to 8 over 10^4 iterations (simulated annealing). During each iteration, the system was equilibrated for 10^3 steps and the final snapshot was analysed for phase-separation. The total number of steps per system (defined by volume fraction ϕ) was therefore 10^7 , each step consisting of L^2 Kawasaki swaps, where L is the system size. For the [Figure 3](#) (main text), the following procedure was employed. During an initial annealing period (10^4 steps), β was increased from 0 to 1, which was used for the rest of the simulation. Afterwards, the system was simulated for 10^6 steps in order to reach equilibrium, which has been verified by observing a plateau in the measured total energy of the system. Finally, the simulation was advanced further 10^6 steps during which a total of 10^3 uniformly distributed snapshots were collected for present analysis.

Further details on the comparison between the cavity method results and numerical simulations

For sake of clarity, we report in [Figure S4](#) phase diagrams for the ternary system where we superimpose the results from BP with the ones from numerical simulations: in all cases we are above 88% accuracy.

Experimental methods

General experimental conditions

Experiments were carried at room temperature in a 1536 well microplate (Greiner bio-one, item no.: 783096) where the coacervate forming components, i.e. Poly-L-lysine hydrochloride of length 10mer (Alamanda polymers, item: PLKC10) and Adenosine 5'-diphosphate sodium salt (Sigma Aldrich, cat. no.: A2754), were distributed, from 10X stock solutions, using Labcyte Echo 550 acoustic liquid dispenser. The phase diagram was obtained in 20mM Tris-HCl buffer at pH 9.0 which was dispensed using Fluidx XRD-384 reagent dispenser. After mixing the solutions, bright-field images of each of the wells were acquired using Yokogawa CellVoyager™ CV7000 high-throughput cytological discovery system at 60X magnification (Olympus objective UPLSAPO60XW, product no.: N1480800). The formation of coacervate was detected by visual inspection of the acquired images and by automated detection of phase separation (through image processing). All the chemicals were prepared in water at 200 mM concentrations. The buffer used here is Tris-HCl, which was prepared by dissolved 25 mM Trizma base (cat. no.: T1503) and was adjusted to pH 9.0 using hydrochloric acid.

Automated liquid handling

Automated liquid handling was performed using two different instruments. The Fluidx XRD-384 reagent dispenser was using to fill 3.2 μ L of buffer solution in a 1536 well microplate (Greiner bio-one, item no.: 783096). The positive and the negative chemical species were dispensed using Labcyte Echo 550 acoustic liquid dispenser in a volume of 4 nL each. The remaining volume up to 4 μ L was filled up with Milli-Q water. The .csv file that served as an input for the Labcyte Echo 550 was prepared using KNIME. The KNIME workflow takes an excel sheet with desired concentrations of positive and negative chemical species and performs required transformations in order to produce a .csv file that is Echo 550 readable. Once the dispensing is over the 1536 well microplate is vortexed in order to mix the two components. The microwell plate is then proceeded for microscopy. The [Figure S5](#) provides a comprehensive view of the complete experimental pipeline.