Phase transitions in biological systems with many components

William M. Jacobs* and Daan Frenkel†
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom
(Dated: September 11, 2016)

Biological mixtures, such as the cytosol and fluid membranes, often consist of thousands of distinct components. There is growing evidence that, under physiological conditions, these mixtures can phase separate into spatially distinct regions with differing compositions. To exploit phase separation as a mechanism to create spatial organization, a biological system must navigate a phase diagram that may, in principle, be exceedingly complex. Here we show that, under certain general conditions, the formation of phase-separated domains is remarkably robust. To elucidate the conditions for 'functional' phase separation, we study the phase behavior of a model of a multicomponent mixture in which the interactions between the components are chosen randomly. Simulations show that such mixtures either demix into multiple phases, each involving only a small number of distinct components, or they behave as if all components were alike, forming two similar-composition phases. We find a sharp transition between these two classes of phase behavior that depends on the variance of the inter-molecular interactions and the number of components. Despite the tendency of mixtures with many components to exhibit binary phase behavior, our model suggests that biological mixtures are naturally poised to undergo a small number of compositionally distinct phase transitions.

INTRODUCTION

Biological systems carry out complex chemical reactions involving a large number of interacting components. However, it would be misleading to view the cell as a well-mixed reaction vessel. Rather, many of the molecular species in the cell exist as phase-separated domains (1, 2), and there is considerable evidence that such phase separation plays a functional role. Examples include the formation of cytoplasmic granules (3–6), nucleoli (7, 8) and amorphous clusters of proteins involved in signaling pathways (9). These structures typically appear as liquid-like droplets that are selectively enriched in some components, resulting in spontaneous compartmentalization (10). Similarly, in some lipid membranes, interactions between embedded components are believed to drive the formation of phase-separated rafts (11–13). For many of these examples, there is strong evidence that phase separation is essentially an equilibrium phenomenon: active processes, such as ATP hydrolysis, are not necessary to observe spatial segregation (14, 15).

If the formation of compositionally distinct domains is important for the normal functioning of a cell, it is logical to ask what constraints the condition of multiphase coexistence puts on the inter-molecular interactions. Thermodynamic coexistence among bulk phases is necessary to achieve a heterogeneous spatial distribution of components in the absence of membranes, such as the component segregation observed in the nucleolus (16). For spontaneous compartmentalization to be useful, it is likely that each phase should be enriched in a relatively small number of functionally related components. Furthermore, the stabilization of these phases should entail only local tuning of the abundances and inter-molecular interactions that contribute to the segregated phases: the desired phase behavior should be robust to fluctuations in the interactions between functionally unrelated components.

To understand the physical requirements for meeting these criteria, we first consider what happens in a multicomponent mixture with *random* pairwise interactions. We assume that the mixture contains N distinct components. Adopting a model first proposed by Sear and Cuesta (17), we treat the pair interactions between the molecules in the mixture as pairwise additive and isotropic, i.e., independent of the relative orientations (18). As a result, a single random variable can be used to describe the average strength of the interaction between any pair of components. Assuming no detailed knowledge of the molecular interactions, we draw each of the $N(N+1)/2$ random interactions independently from the same distribution (that we shall specify later).

We find that this random-mixture model generates phase diagrams that meet the criteria for biological functionality under certain general conditions. The phase diagrams of such mixtures fall into two classes, depending on the distribution of interactions and the number of components: one class that supports multiple demixed phases with very different compositions and a much simpler case where only two compositionally similar phases coexist. As predicted by a previous simulation study (19), increasing the number of components tends to favor the simpler phase behavior. Here we show that the disappearance of multiple demixed phases occurs suddenly and without a smooth transition to the simpler phase behavior. However, we predict that the location of this cross-over depends relatively weakly on the number of components. Thus, although multiphase coexistence is not simply a consequence of having a large number of components, our model suggests that this behavior is nevertheless a

* Present address: Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA, 02138, USA; wjacobs@fas.harvard.edu
† df246@cam.ac.uk

likely outcome in biological mixtures and that the selection of specific phases requires only local tuning of the inter-molecular interactions. We provide an intuitive explanation for these results and then discuss the implications for biological mixtures.

THEORY AND MONTE CARLO SIMULATIONS

Phase diagrams in many dimensions

Classical thermodynamics predicts that in a mixture with N components, it is possible to observe as many as $N+2$ simultaneously coexisting equilibrium phases (20). The precise number of coexisting phases that appear in any particular system depends on the interactions between the components. Yet before describing the phase behavior of mixtures with many components, it is useful to examine the generic features of the phase diagrams of solutions with two solutes. In what follows, we shall use an 'implicit solvent' picture in which the solvent is not considered to be a component but is merely the medium in which the various molecular species move. Moreover, we shall assume that both the temperature and pressure are constant; in general, the compositions of the coexisting phases are temperature and pressure-dependent.

Because the two components may have dissimilar pairwise interactions, phase separation can drive the formation of phases with differing compositions. It is convenient to characterize an N-component mixture by the concentration vector $\boldsymbol{\rho} \equiv \{\rho_1, \rho_2, \cdots, \rho_N\}$, where the total concentration of a phase is $\phi \equiv \sum_{i=1}^{N} \rho_i$. We use dimensionless units for the component concentrations so that all concentration vectors lie within a unit $(N+1)$-dimensional simplex, $0 \leq \phi \leq 1$. The composition of a phase refers to the mole fraction of each component, $x_i \equiv \rho_i/\phi$. Consequently, the composition describes the direction (but not the magnitude) of $\boldsymbol{\rho}$.

During a phase transition, both the magnitude and the orientation of $\boldsymbol{\rho}$ can change. Let us imagine that, prior to phase separation, the mixture is described by the parent concentration vector $\boldsymbol{\rho^{(0)}}$. After phase separation, coexistence is established between phases with concentration vectors $\boldsymbol{\rho^{(1)}}$ and $\boldsymbol{\rho^{(2)}}$ and associated total concentrations $\phi_1 < \phi_2$. In the phase diagrams shown in Figure 1a–c, the shaded areas denote coexistence regions, and the circles indicate pairs of daughter phases $\boldsymbol{\rho^{(1)}}$ and $\boldsymbol{\rho^{(2)}}$; the parent concentration vector lies on the dashed tie line connecting the two daughter phases. To characterize the nature of a particular phase transition, it is useful to define an 'angle of phase separation,' the angle between the parent concentration vector and the tie line connecting the two coexisting phases:

$$\theta \equiv \cos^{-1}\left[\frac{\boldsymbol{\rho^{(0)}} \cdot \left(\boldsymbol{\rho^{(2)}} - \boldsymbol{\rho^{(1)}}\right)}{\left\|\boldsymbol{\rho^{(0)}}\right\| \left\|\boldsymbol{\rho^{(2)}} - \boldsymbol{\rho^{(1)}}\right\|}\right]. \qquad (1)$$

While this angle does not indicate the composition of the new phase, it serves as an order parameter that reports the similarity in composition between the parent phase and the coexisting daughter phases.

Multicomponent phase separation can result in phases with equal compositions (Figure 1a), the formation of a high-concentration phase that is enriched in one component (Figure 1b), or some combination thereof. In the case where $\theta \to 0$, the coexisting phases differ only in their total concentrations. This scenario is referred to as a *condensation* phase transition. Alternatively, if phase separation at an initially low concentration is driven by the *demixing* of a single component, then $\theta \to \theta_N \equiv \cos^{-1}(N^{-1/2})$. In principle, in mixtures with more than two components, 'intermediate' phase behavior can occur if θ takes a value between 0 and θ_N; this scenario implies the selective phase separation of many, but not all, components. Figure 1c also shows a three-phase coexistence region, where the mixture phase separates into three coexisting phases with concentrations corresponding to the vertices of the dark shaded triangle. For parent concentrations lying within this region, multiple demixing transitions occur at the same temperature and pressure. However, as shown in Figure 1c, only one phase transition is typically encountered at the boundary of the homogeneous phase, unless the interaction matrix is degenerate.

Characterization of phase diagrams via simulation

Simulations allow us to study the nature of phase separation in the limit of very many components. Unlike the canonical phase diagrams shown in Figure 1a–c, we perform simulations in the grand-canonical ensemble, where the chemical potential of each component is held fixed and the number of particles of each type is allowed to fluctuate. The grand-canonical ensemble is ideal for studying phase behavior because it allows the formation of pure phases of variable concentration in a single simulation. This feature allows us to observe coexistence between bulk phases without the formation of an interface.

Because the component concentrations fluctuate, grand-canonical simulations sample a free-energy landscape in which low-free-energy basins correspond to the phases on a canonical phase diagram. For each representative phase diagram shown in Figure 1a–c, a corresponding free-energy landscape is shown below in Figure 1d–f. In each case, the illustration shows the free-energy landscape for the component chemical potentials that correspond to any parent concentration vector on the highlighted tie line. Importantly, simulations in the grand-canonical ensemble reveal metastable free-energy basins even when the corresponding phase is unstable in the canonical phase diagram. For example, the metastable basin in the asymmetric free-energy landscape shown in Figure 1f corresponds to a third phase that must be stable at a nearby set of chemical potentials, indicating that

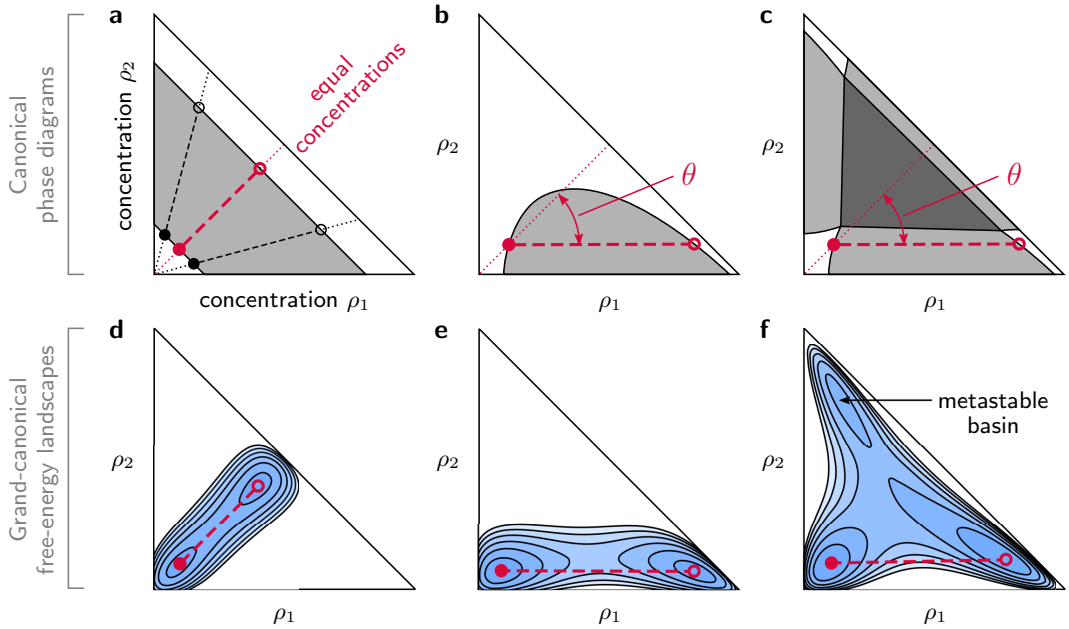

FIG. 1. Representative phase diagrams and free-energy landscapes of two-component mixtures. In the constant-temperature and pressure phase diagrams shown in panels **a**–**c**, single phase regions are shown in white, two-phase coexistence regions in light gray and a three-phase coexistence region in dark gray. The components have concentrations ρ_1 and ρ_2, dotted lines indicate constant compositions, and dashed lines indicate example tie lines connecting coexisting phases (circles). The red dotted line indicates an equimolar parent composition, while the red dashed line indicates the tie line at the boundary of the equimolar homogeneous phase. The angle of phase separation, θ, is the angle between the highlighted parent composition and the tie line at the cloud point (filled circle). For each phase diagram, panels **d**–**f** depict corresponding free-energy landscapes for parent concentrations lying on the highlighted tie lines. In these landscapes, the chemical potentials of the components are fixed, and the component concentrations fluctuate around two or more free-energy basins. In panels **c** and **f**, the highlighted tie line is in a two-phase region, but the appearance of a metastable phase on the free-energy landscape indicates close proximity to a three-phase coexistence region.

the tie line on the canonical phase diagram in Figure 1c is close to a three-phase coexistence region.

In order to to determine whether a matrix of intermolecular interactions supports multiphase coexistence, we follow a simulation strategy that allows us to characterize the canonical phase diagram. Here we assume that the parent phase has equal concentrations for all components. Starting from the low-concentration homogeneous phase, we located the lowest-concentration phase boundary that intersects the parent composition vector, i.e., $\boldsymbol{\rho^{(1)} = \rho^{(0)}}$, which is commonly called the *cloud point*. This point is indicated by the filled circles in Figure 1a–c. Then, we sampled the free-energy landscape at the corresponding chemical potentials to determine the proximity of this phase boundary to a multiphase coexistence region. These two sets of simulations are described below.

Monte Carlo simulations of binary phase coexistence

To locate the cloud point associated with an equimolar parent composition, we carried out grand-canonical Monte Carlo (GCMC) simulations (21) of a three-dimensional lattice model. In this model, lattice sites may either be vacant, representing the non-interacting solvent, or occupied by one of the N components. The energy, U, of a lattice configuration is calculated by summing the interactions between all pairs of particles on nearest-neighbor (n.n.) lattice sites:

$$U = -\sum_{(u,v)\in\{\text{n.n.}\}} \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \delta_{C_u i}\, \epsilon_{ij}\, \delta_{C_v j} \right]. \qquad (2)$$

The tuples (u, v) run over all distinct nearest-neighbor pairs, C_u is the component index of the particle at lattice site u ($C_u = 0$ if the site is vacant), and δ is the Kronecker delta.

We employed multicanonical biasing (22) to facilitate rapid crossing of the free-energy barrier that separates low and high-concentration lattice configurations. In order to locate the cloud point in these simulations, we tuned the component chemical potentials to achieve equal free energies in the low and high-concentration phases, $F^{(1)} = F^{(2)}$, while satisfying the imposed composition constraint on the low-concentration phase, $\boldsymbol{x^{(1)} = x^{(0)}}$. This strategy has been shown to minimize finite-size effects (23). We note that this approach does not constrain the simulation to sample only from the two phases of interest if other free-energy basins are present. For computational efficiency, we used a $L \times L \times L$ cubic lattice

with periodic boundaries and $L = 6$. Complete details of the simulation method are provided in Ref. 19.

Free-energy landscapes near multiphase coexistence

For each realization of the interaction matrix, we generated approximately 1000 random lattice configurations, with concentration vectors chosen uniformly from the unit simplex of component concentrations. We then allowed each initial configuration to evolve via the unbiased GCMC algorithm, with the component chemical potentials fixed at the previously determined cloud point, until the concentration fluctuations stabilized. These GCMC trajectories tend to travel 'downhill' on the free-energy landscape, and, without the multicanonical biasing that was used in the previous set of simulations, the probability of escaping from a free-energy basin is extremely small. The endpoints of these trajectories therefore lie close to the minima of the free-energy basins, which correspond to stable or metastable thermodynamic phases. We clustered these endpoints in order to determine the number of distinct free-energy basins, and then we found the mean concentration vector of each of these (meta)stable phases by averaging the endpoint concentration vectors within each basin.

RESULTS

Simulations of condensation and demixing in mixtures with random pairwise interactions

To study an ensemble of mixtures with random interactions, we generated many independent realizations of the pairwise interaction matrices $\{\epsilon_{ij}\}$. We assume that the interactions between pairs of components are independent random variables drawn from a Gaussian distribution with mean $\bar{\epsilon}$ and variance σ^2. (We shall discuss the effects of correlated interactions in the context of a mean-field model later.) We limited the number of components to 64 for tractability and then chose the variance of the Gaussian interaction distributions in order to observe both the demixing and condensation limits of the phase behavior. The mean interaction strength was fixed at $\bar{\epsilon} = 1.07\epsilon_c$, where $-\epsilon_c = -0.87k_BT$ is the critical bond energy of the one-component lattice gas, k_B is the Boltzmann constant and T is the absolute temperature.

For each realization of $\{\epsilon_{ij}\}$, we used our simulation strategy to calculate the component chemical potentials at the lowest-concentration phase boundary and to determine the number of free-energy basins at this point. We also computed the average composition of the coexisting high-concentration phase at this set of chemical potentials in order to calculate θ. As noted above, it is highly unusual to observe multiple coincident transitions precisely at this first phase boundary. When we do observe more than two free-energy basins, multiphase coexistence must occur at very similar component chemical potentials. In this case, a stronger driving force — either higher concentrations or stronger average interactions — is necessary to push the mixture into the multiphase region. This scenario is consistent with our expectations for functional multiphase coexistence, because phase transitions that are close in chemical-potential space may be easily manipulated by small changes to the component abundances and inter-molecular interactions.

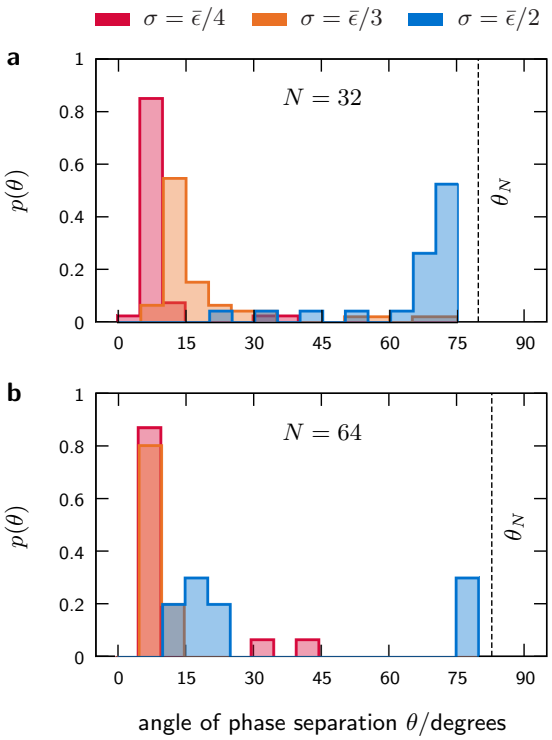

FIG. 2. Histograms showing the bimodal distributions of the angle of phase separation in simulations with (a) 32 and (b) 64 components. The histogram for each random-mixture ensemble was constructed from all high-concentration free-energy basins at the cloud point (see text). Condensation into two phases with equal compositions is associated with a small angle of phase separation, θ, whereas demixing into phases with dissimilar compositions occurs when θ approaches θ_N, the angle of phase separation corresponding to the demixing of a single component. In panel **b**, the $\sigma = \bar{\epsilon}/2$ ensemble has nearly equal probabilities of condensation and demixing.

Bimodal distribution of phase behaviors

Our simulations show that random interactions result in two distinct types of phase behavior: either a handful of components will demix or the mixture will condense into two phases with nearly identical compositions. The bimodal distribution of the angle of phase separation can be seen in the histograms shown in Figure 2a–b, where we have included all high-concentration phases that were

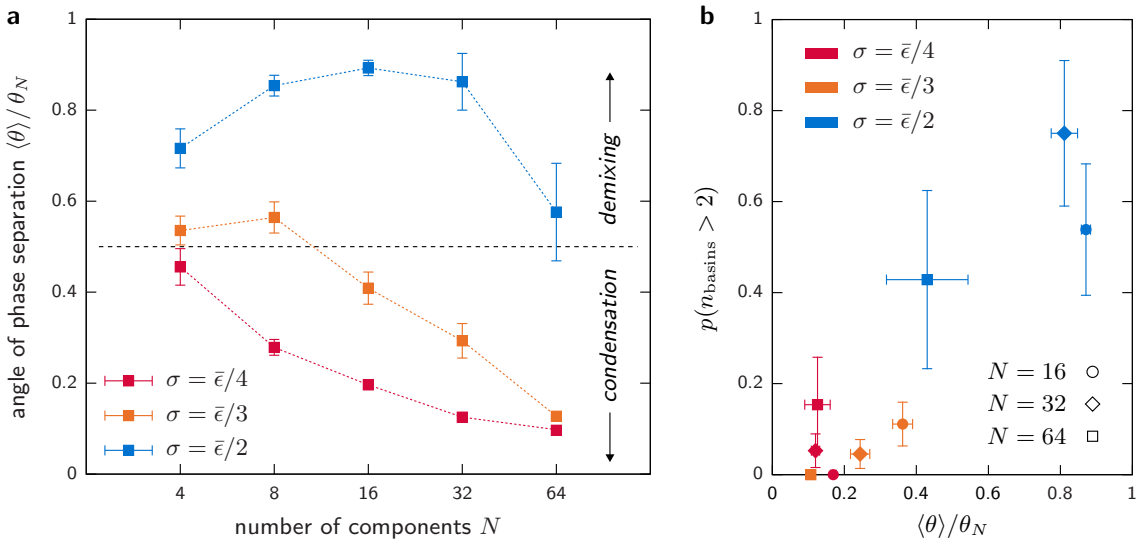

FIG. 3. The probability of multiphase coexistence correlates with the angle of phase separation at the cloud point. (a) Increasing the number of components in a solution suppresses demixing and leads to a single condensation phase transition, regardless of the variance of the interactions. Each point represents an average over an ensemble of random mixtures; error bars indicate the standard deviation of the distribution of angles for each ensemble. The dashed line roughly indicates where cross-over from demixing to condensation occurs. (b) The probability of observing more than two free-energy basins at the cloud point, indicating close proximity to a multiphase coexistence region. Demixing transitions are associated with the presence of multiple coexisting phases. Vertical error bars indicate the error due to finite sampling.

identified from the free-energy landscapes of hundreds of random mixtures with 32 or 64 components. Low values of θ correspond to condensation phase transitions. In this limit, a multicomponent mixture behaves as if all the components were alike. In contrast, the higher-θ peak corresponds to demixing phase transitions. In this case, the maximum angle of phase separation is given by θ_N. Importantly, in simulations with 8 or more components, we rarely find phase behavior that is intermediate between condensation and demixing.

The average phase behavior of a random mixture ensemble depends on both the width of the distribution of inter-molecular interactions, σ, and the number of components, N. Condensation transitions dominate for large N and small σ, while demixing transitions are more likely in mixtures with fewer components and a broader distribution of interactions. In fact, the average angle of phase separation tends toward zero at large N, regardless of the variance of the random interactions. This ensemble-averaged phase behavior is shown in Figure 3a, where θ is shown relative to the largest possible angle of phase separation for a given number of components, θ_N. The cross-over between demixing and condensation, where $\langle\theta\rangle$ starts to tend towards 0, occurs at larger N in random-mixture ensembles with greater values of σ. In Figure 3, the cross-over occurs near $N \simeq 4$, $N \simeq 16$ and $N \simeq 64$ for ensembles with $\sigma = \bar{\epsilon}/4$, $\sigma = \bar{\epsilon}/3$ and $\sigma = \bar{\epsilon}/2$, respectively. The variance in θ across realizations of interaction matrix also decreases as $\langle\theta\rangle$ approaches zero, while the larger variance in the ensemble-averaged angle of phase separation near $\langle\theta\rangle \simeq 1/2$ is a consequence of the bimodal distribution shown in Figure 2. We shall discuss the origin of bimodality and the scaling of N, σ and $\bar{\epsilon}$ at the cross-over between demixing and condensation using mean-field calculations below.

Multiphase coexistence correlates with the angle of phase separation

Free-energy landscapes calculated at the cloud point indicate that multiphase coexistence is most likely to occur in concert with demixing transitions. In Figure 2c, we plot the probability, calculated separately for each random-mixture ensemble, of finding more than one high-concentration phase in the vicinity of the cloud point. The correlation of multiphase coexistence near the cloud point with the ensemble-averaged angle of phase separation implies that multiphase coexistence can be predicted by the angle of phase separation at the cloud point. This observation allows us to classify the likely phase diagram of a random mixture ensemble based on the statistical distribution of the inter-molecular interactions in a random mixture ensemble. Furthermore, near the cross-over from demixing to condensation, we find many instances where a stable, high-concentration homogeneous phase is accompanied by one or more demixed phases. This feature indicates that the cross-over from demixing to condensation is a sharp, first-order transition that depends on the relative stability of competing high-concentration phases; the smooth transition in Figure 3a is the result of averaging over the bimodal distribution of angles of phase separation.

Mean-field model of multicomponent phase separation

We can gain further insight into the transition between these two opposing types of phase behavior by analyzing a mean-field model. In particular, this approach reproduces the bimodality of the phase behavior observed in our simulations. This mean-field model also allows us to predict the scaling of the condensation–demixing cross-over as a function of N, σ and $\bar{\epsilon}$.

Ignoring all spatial correlations between the various particles and vacancies on the lattice, the Helmholtz free energy of a mixture with concentration vector $\boldsymbol{\rho}$ and lattice coordination number z is

$$\beta F = \sum_i \rho_i \ln \rho_i + (1-\phi)\ln(1-\phi) - \frac{\beta z}{2}\sum_{ij}\rho_i \epsilon_{ij}\rho_j, \quad (3)$$

where $\phi \equiv \sum_i \rho_i \leq 1$ and $\beta \equiv (k_{\mathrm{B}}T)^{-1}$ is the inverse temperature. In Eq. 3, the second term accounts for the mixing entropy of vacancies. Because calculating the total concentration of each phase at coexistence is challenging in the context of a multicomponent mixture, we shall make the simplifying assumption that all high-concentration phases have the same total concentration ϕ. With this approximation, we can write the free-energy difference, ΔF, between a phase with an arbitrary composition $\boldsymbol{x}$ and the equal-composition phase as

$$\frac{\beta \Delta F}{\phi} = \left(\sum_i x_i \ln x_i + \ln N\right) - \frac{\beta z \phi}{2}\sum_{ij} x_i \Delta\epsilon_{ij} x_j, \quad (4)$$

where $\Delta\epsilon_{ij} \equiv \epsilon_{ij} - \sum_{ij}\epsilon_{ij}/N^2$. The first and second terms in Eq. 4 represent the mixing entropy and the average energy differences between the two phases, respectively.

Now let us assume that the components in an arbitrary phase interact via a specific subset of n^2 interaction energies and that each of these interactions is equally likely to appear at any bond between nearest-neighbor particles. Choosing such a subset entails reshuffling the rows and columns of the interaction matrix in order to select a $n \times n$ sub-matrix of interaction energies, as illustrated in Figure 4a. For example, a phase with $n = 1$ could correspond to a single-component phase, if the chosen interaction is on the diagonal of the matrix, or a two-component phase in which the components occupy alternating lattice sites, if the chosen interaction is off-diagonal. The lowest free energy of a phase with n^2 interactions, relative to the equal-composition phase with N^2 interactions, is

$$\frac{\Delta F(n)}{\beta^{-1}\phi} = \ln\left(\frac{N}{n}\right) - \beta' \max_{\{\boldsymbol{\tau}^{(n)},\boldsymbol{\upsilon}^{(n)}\}}\left[\frac{1}{n^2}\sum_{ij}\tau_i^{(n)}\Delta\epsilon_{ij}\upsilon_j^{(n)}\right], \quad (5)$$

where $\beta' \equiv \beta z\phi/2$. The vectors $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{\upsilon}^{(n)}$, which have precisely n ones and $N-n$ zeros, implement the reshuffling of rows and columns (Figure 4a). The second term in Eq. 5 is proportional to the maximum value (over all possible vectors $\boldsymbol{\tau}^{(n)}$ and $\boldsymbol{\upsilon}^{(n)}$) of the *average* of the n^2 interaction energies.

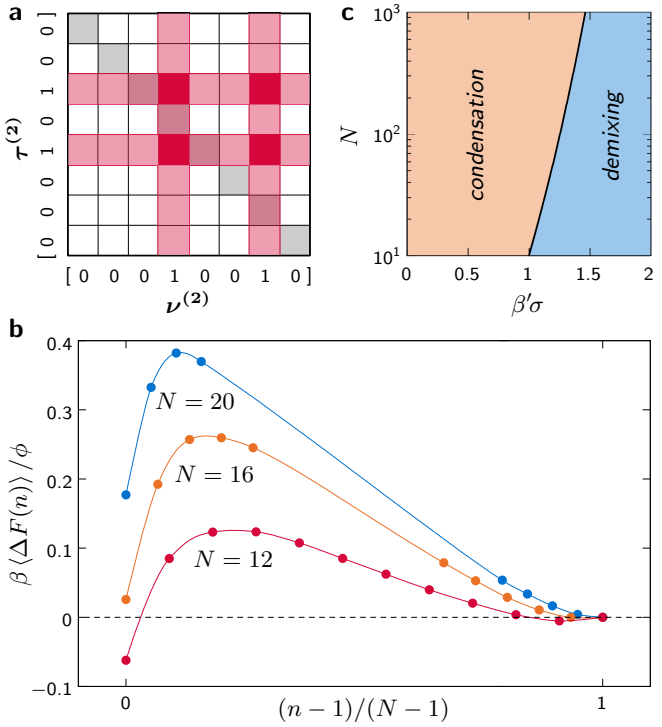

FIG. 4. A mean-field model predicts bimodal phase behavior for random mixtures. (a) An illustration of a 2×2 sub-matrix (red squares) of $\boldsymbol{\Delta\epsilon}$ selected by the vectors $\boldsymbol{\tau}^{(2)}$ and $\boldsymbol{\upsilon}^{(2)}$; see Eq. 5. (b) The mean-field free-energy difference between the most stable phase comprising n^2 distinct interaction energies and the equal-composition phase with N components, under the constraint that both phases have the same overall concentration ϕ. Numerical calculations of Eq. 5 are averaged over many realizations of the random matrix $\boldsymbol{\Delta\epsilon}$; for this illustration, we have chosen $\beta'\sigma = 1.05$. The conditions for the cross-over regime can be found by tuning N and σ to achieve equal free energies in the most stable demixed phase ($n = 1$) and the most stable condensed phase ($n \simeq N$). (c) The mean-field phase diagram for the control parameters N and $\beta'\sigma$. The coexistence curve between condensation and demixing scales approximately as $\sqrt{\ln N} \sim \beta'\sigma$.

The composite parameter β' has units of inverse energy and sets the scale against which the standard deviation of the interaction distribution should be measured. This parameter is an implicit function of the mean interaction strength, $\bar{\epsilon}$, which plays an important role in determining the total concentration, ϕ, at the cloud point. The parameter β' can also be related to the critical interaction energy, ϵ_{c}, in a single-component system. In the analogous one-component mean-field model, $\epsilon_{\mathrm{c}} = 4/\beta z$ and $\phi_{\mathrm{c}} = 1/2$, meaning that $\beta' = \epsilon_{\mathrm{c}}^{-1}$. Clearly, $\bar{\epsilon}$ must be at least as great as this critical bond strength in order to observe a condensation phase transition in a multicomponent system. However, the dependence of β' on $\bar{\epsilon}$ is, in general, more complicated for $\bar{\epsilon} > \epsilon_{\mathrm{c}}$.

Numerical analysis shows that Eq. 5 reproduces the bimodal phase behavior observed in our simulations. We evaluated Eq. 5 for many realizations of the random interaction matrix and plotted the results in Figure 4b. The second term in Eq. 5 is minimized when $n = 1$, in which case $\boldsymbol{\tau}^{(\mathbf{1})}$ and $\boldsymbol{\upsilon}^{(\mathbf{1})}$ pick out the largest entry in the interaction matrix. The average of the n^2 interactions regresses toward the mean, $\sum_{ij} \Delta\epsilon_{ij}/N^2 = 0$, as more interaction energies are incorporated. However, this reduction in the average interaction strength is not perfectly balanced by the increase in mixing entropy. As a result, Figure 4b shows that the free-energy difference as a function n is bistable: either a demixing transition with $n = 1$ or a condensed phase with $n \simeq N$ is thermodynamically favored. Equal probabilities for these two scenarios, corresponding to coexistence between demixing and condensation phase transitions, can be obtained by tuning either the total number of components N or the standard deviation of the interaction energies such that $\Delta F = 0$ in the $n = 1$ phase.

The free-energy curves plotted in Figure 4b resemble a classical first-order phase transition between stable demixed and condensed phases, where the total number of components and the standard deviation of the interaction strengths are the control parameters. In random mixtures with uncorrelated interaction energies, high-concentration phases with an intermediate number of components — corresponding to phases with intermediate values of θ — are unlikely to be observed. The emergence of a condensation phase transition, where $n \simeq N$, can be attributed to the greater mixing entropy of the equal-composition, high-total-concentration phase. In contrast, if $n \simeq 1$, the thermodynamic driving force for demixing results from the extreme values of the distribution of random interactions. Only a small number of distinct demixed phases are therefore likely to be stable. We find that the fluctuations in the extreme values of $\boldsymbol{\Delta\epsilon}$ are large in comparison to the standard deviation of the distribution of interaction energies, leading to significant variance in $\Delta F(n)$ across realizations of the interaction matrix. As a result, the mean-field model predicts a smooth cross-over between demixing and condensation behavior when averaged over an ensemble of random mixtures, as found in our simulation results.

We can predict the scaling of the condensation–demixing cross-over by determining the conditions for coexistence between condensation and demixing transitions from the extreme-value statistics of the interaction matrix. When $n = 1$, the expected value of the second term in Eq. 5 is $-\beta'\sigma(\sqrt{4.73 \ln N} - 1)$, where σ has units of $k_{\mathrm{B}}T$. It follows that the coexistence curve between condensation and demixing is $\sqrt{\ln N} = a\beta'\sigma + \sqrt{(a\beta'\sigma)^2 - \beta'\sigma}$, where $a \simeq 1.09$. This relationship between N and $\beta'\sigma$ is considerably steeper than the $\sqrt{N} \sim \sigma/\bar{\epsilon}$ scaling determined from a stability analysis of the homogeneous parent phase (17). Consequently, the N–$\beta'\sigma$ phase diagram shown in Figure 4c suggests that, in mixtures characterized by large values of $\beta'\sigma$, the number of components required to promote a condensation phase transition may be extremely large. The simulation results shown in Figure 3 are also incompatible with $\sqrt{N} \sim \sigma/\bar{\epsilon}$ scaling, although our simulation data is insufficient to test the mean-field prediction.

This mean-field model can also be used to predict the phase behavior of multicomponent mixtures under more general assumptions regarding the parent-phase composition and the inter-molecular interactions. In mixtures with unequal component concentrations, the dimensionless Gibbs entropy of the parent phase $-\sum_i x_i^{(0)} \ln x_i^{(0)}$ measures the effective number of components in the mixture. The entropy difference between the condensed and demixed phases (the first term in Eq. 5) will, in general, be different for each component. We have also verified that the predicted bistability is robust with respect to weak correlations among the inter-molecular interactions. In particular, we checked this result for mixtures where the interaction matrix follows the Lorentz–Berthelot mixing rule (24) and thus has the form $\epsilon_{ij} \sim \sqrt{s_i s_j}$, where the $\{s_i\}$ are independent Gaussian random variables with the constraint $s_i \geq 0$, and the largest value of ϵ_{ij} is always found on the diagonal of the interaction matrix. In cases where the interaction distribution is bimodal, our predictions apply to the subset of strongly interacting components, which may condense or demix independently from the other weakly interacting components.

DISCUSSION

For phase separation to be a viable mechanism of spatial organization, it must be possible to tune the stability of multiple coexisting phases via a limited number of local parameters: the phase behavior of a mixture with many components should be insensitive to fluctuations in the interactions between each and every pair of components. We have shown that a model in which the inter-molecular interactions are chosen randomly produces phase diagrams that satisfy this requirement under quite general conditions. The most probable phase behavior of such a mixture falls into one of two classes, depending on the number of components and the statistical distribution of inter-molecular interactions. At the point where a homogeneous mixture first becomes unstable, either a few components will demix, or all components will condense together. The former type of phase behavior typically supports coexistence among multiple compositionally distinct phases given small changes in the component chemical potentials. This means that, when demixing occurs in a random mixture, the conditions for multiphase coexistence are both accessible and easily tunable.

Our results have a number of important implications for biological mixtures. One consequence of bimodal phase behavior is that only a small number of interactions need to be tuned to achieve coexistence among multiple demixed phases. In the absence of strong correlations

in the inter-molecular interactions, our mean-field model predicts that each demixed phase is unlikely to be enriched in more than a handful components. The stabilities of these phases can therefore be tuned by optimizing a few specific chemical potentials. To manipulate phase behavior within the lifetime of a single cell, such control is likely achieved through the regulation of macromolecular abundances, while on evolutionary timescales, the phase behavior may be tuned via local changes to the inter-molecular interactions. In an intracellular mixture, multiphase demixing results in the formation of compositionally inhomogeneous droplets, such as spontaneous nucleolar compartmentalization (16). However, the spatial ordering of the phases within a droplet depends on the interfacial free energies, which we have not examined here.

Functional multiphase coexistence furthermore requires that demixing transitions are not hidden by a dominating condensation instability. Our mean-field model predicts that the required number of components for condensation grows rapidly with the variance of the interaction-energy distribution. This scaling implies that demixing transitions can be observed in mixtures that contain a few thousand distinct components with an interaction-distribution standard deviation on the order of a few $k_{\mathrm{B}}T$. In our simulations, we have chosen the variance of the interaction strengths in order to explore the behavior of the model over a computationally tractable number of components. However, these parameters could instead be obtained from high-throughput protein–protein interaction assays. A previous analysis of Yeast 2-Hybrid experiments (25) found that the proteome-wide distribution of nonspecific interactions is indeed quite broad, with an estimated mean and standard deviation of $-4k_{\mathrm{B}}T$ and $2.5k_{\mathrm{B}}T$, respectively.

Our model also shows that optimization of all inter-component interactions is not necessary to stabilize a multicomponent system. In contrast to the complexity that might be expected from the Gibbs phase rule, phase separation into two phases with similar compositions is a common outcome for mixtures with very many components. This result may help to explain the condensation-like phase behavior that has been observed for components embedded in lipid membranes (1). Although such simple phase behavior might look like that of a single-component solution, this is not a result of the interacting components being indistinguishable. On the contrary, this behavior is a consequence of the mixing entropy that stabilizes homogeneous-composition phases in mixtures with a large number of distinguishable components.

Owing to the simplicity of our model and the minimal constraints imposed on the form of the interaction matrix, our results can be applied to a wide variety of biological systems, including both two-dimensional membranes and three-dimensional fluids. However, additional constraints on the inter-molecular interactions are required before quantitative conclusions can be drawn for any specific system. Whereas we have assumed uncorrelated pairwise interactions in our simulations, some classes of biomolecules, such as intrinsically disordered proteins, are more likely to interact promiscuously. In particular, experiments have implicated proteins with low-complexity sequences in some examples of *in vivo* phase separation (26). These considerations imply a more structured interaction matrix than the uncorrelated Gaussian ensemble. In addition, protein expression levels are typically anticorrelated with the propensity of a protein to form nonspecific interactions (25, 27, 28). Our mean-field calculations indicate that bimodal phase behavior is robust to physically realistic correlations among the random inter-molecular interactions; however, very strong correlations may affect this result. Further investigation is warranted to examine the effects of correlated interactions and unequal concentrations on the phase behavior of multicomponent mixtures.

In summary, we have described the phase behavior of a minimal model of a biological mixture with many components. We have shown that the general features of these phase diagrams depend on a small number of parameters that describe the distribution of interactions between pairs of components. With a sufficiently broad distribution of uncorrelated inter-molecular interactions, such mixtures typically exhibit tunable multiphase coexistence. Our results therefore suggest that the natural behavior of multicomponent mixtures with random interactions is similar to that required for phase separation to be an organizing principle in biological mixtures. While the study of high-dimensional phase diagrams has long been a topic of theoretical interest (29, 30), recent explorations of the organization of complex biological systems — which are truly multicomponent mixtures — have brought new relevance to this problem. The general principles developed here will guide future studies in this emerging field.

ACKNOWLEDGMENTS

W.M.J. acknowledges support from the Gates Cambridge Trust and the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1143678.

REFERENCES

[1] Hyman, A. A., and K. Simons, 2012. Beyond oil and water – Phase transitions in cells. *Science* 337:1047–1049.

[2] Weber, S. C., and C. P. Brangwynne, 2012. Getting RNA and protein in phase. *Cell* 149:1188–1191.

[3] Brangwynne, C. P., C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, 2009. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324:1729–1732.

[4] Elbaum-Garfinkle, S., Y. Kim, K. Szczepaniak, C. C.-H. Chen, C. R. Eckmann, S. Myong, and C. P. Brangwynne,

2015. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 201504822.

[5] Wippich, F., B. Bodenmiller, M. G. Trajkovska, S. Wanka, R. Aebersold, and L. Pelkmans, 2013. Dual specificity kinase DYRK3 couples stress granule condensation/dissolution to mTORC1 signaling. *Cell* 152:791–805.

[6] Molliex, A., J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, 2015. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163:123–133.

[7] Brangwynne, C. P., T. J. Mitchison, and A. A. Hyman, 2011. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proc. Natl. Acad. Sci. U.S.A.* 108:4334–4339.

[8] Nott, T. J., E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, J. D. Forman-Kay, and A. J. Baldwin, 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* 57:936–947.

[9] Sear, R. P., 2007. Dishevelled: A protein that functions in living cells by phase separating. *Soft Matter* 3:680–684.

[10] Keating, C. D., 2012. Aqueous phase separation as a possible route to compartmentalization of biological molecules. *Acc. Chem. Res.* 45:2114–2124.

[11] Lingwood, D., and K. Simons, 2010. Lipid rafts as a membrane-organizing principle. *Science* 327:46–50.

[12] Lingwood, D., J. Ries, P. Schwille, and K. Simons, 2008. Plasma membranes are poised for activation of raft phase coalescence at physiological temperature. *Proc. Natl. Acad. Sci. U.S.A.* 105:10005–10010.

[13] Kaiser, H.-J., D. Lingwood, I. Levental, J. L. Sampaio, L. Kalvodova, L. Rajendran, and K. Simons, 2009. Order of lipid phases in model and plasma membranes. *Proc. Natl. Acad. Sci. U.S.A.* 106:16645–16650.

[14] Berry, J., S. C. Weber, N. Vaidya, M. Haataja, and C. P. Brangwynne, 2015. RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. U.S.A.* 112:E5237–E5245.

[15] Weber, S. C., and C. P. Brangwynne, 2015. Inverse size scaling of the nucleolus by a concentration-dependent phase transition. *Curr. Biol.* 25:641–646.

[16] Feric, M., N. Vaidya, T. S. Harmon, D. M. Mitrea, L. Zhu, T. M. Richardson, R. W. Kriwacki, R. V. Pappu, and C. P. Brangwynne, 2016. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* 165:1686–1697.

[17] Sear, R. P., and J. Cuesta, 2003. Instabilities in complex mixtures with a large number of components. *Phys. Rev. Lett.* 91:245701.

[18] Jacobs, W. M., D. W. Oxtoby, and D. Frenkel, 2014. Phase separation in solutions with specific and nonspecific interactions. *J. Chem. Phys.* 140:204109.

[19] Jacobs, W. M., and D. Frenkel, 2013. Predicting phase behavior in multicomponent mixtures. *J. Chem. Phys.* 139:024108.

[20] Gibbs, J. W., 1906. The scientific papers of J. Willard Gibbs, volume 1. Longmans, Green and Company, New York.

[21] Frenkel, D., and B. Smit, 2001. Understanding molecular simulation: From algorithms to applications. Academic Press.

[22] Berg, B. A., and T. Neuhaus, 1992. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* 68:9.

[23] Buzzacchi, M., P. Sollich, N. B. Wilding, and M. Müller, 2006. Simulation estimates of cloud points of polydisperse fluids. *Phys. Rev. E* 73:046110.

[24] Berthelot, D., 1898. Sur le mélange des gaz. *Compt. Rendus* 126:1703–1706.

[25] Zhang, J., S. Maslov, and E. I. Shakhnovich, 2008. Constraints imposed by non-functional protein–protein interactions on gene expression and proteome size. *Mol. Sys. Biol.* 4:210.

[26] Brangwynne, C. P., P. Tompa, and R. V. Pappu, 2015. Polymer physics of intracellular phase transitions. *Nat. Phys.* 11:899–904.

[27] Heo, M., S. Maslov, and E. I. Shakhnovich, 2011. Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108:4258–4263.

[28] Johnson, M. E., and G. Hummer, 2011. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. *Proc. Natl. Acad. Sci. U.S.A.* 108:603–608.

[29] Griffiths, R. B., and J. C. Wheeler, 1970. Critical points in multicomponent systems. *Phys. Rev. A* 2:1047.

[30] Griffiths, R. B., 1975. Phase diagrams and higher-order critical points. *Phys. Rev. B* 12:345.