SEE CORRECTION FOR THIS ARTICLE
RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Evolved interactions stabilize many coexisting phases in multicomponent liquids

David Zwicker[a,1] and Liedewij Laan[b]

Edited by Daan Frenkel, University of Cambridge, Cambridge, United Kingdom; received January 27, 2022; accepted April 21, 2022

Phase separation has emerged as an essential concept for the spatial organization inside biological cells. However, despite the clear relevance to virtually all physiological functions, we understand surprisingly little about what phases form in a system of many interacting components, like in cells. Here we introduce a numerical method based on physical relaxation dynamics to study the coexisting phases in such systems. We use our approach to optimize interactions between components, similar to how evolution might have optimized the interactions of proteins. These evolved interactions robustly lead to a defined number of phases, despite substantial uncertainties in the initial composition, while random or designed interactions perform much worse. Moreover, the optimized interactions are robust to perturbations, and they allow fast adaption to new target phase counts. We thus show that genetically encoded interactions of proteins provide versatile control of phase behavior. The phases forming in our system are also a concrete example of a robust emergent property that does not rely on fine-tuning the parameters of individual constituents.

statistical physics | droplets | biomolecular condensates | optimization

Significance

Biological cells need to orchestrate thousands of different biomolecules for reliable function. To do this, they partly rely on phase separation as a passive physical mechanism to form membraneless compartments of distinct composition. We study how molecular interactions determine phase behavior by introducing a numerical method for multicomponent liquids. We find that random and structured interaction patterns typically do not lead to a precise number of phases. In contrast, evolutionarily optimized interactions can solve this task perfectly and are robust to internal and external perturbations. We thus demonstrate that adjusting microscopic interactions leads to stable emergent behaviors in these complex systems.

Biological cells are incredibly complex and consist of thousands of different biomolecules that move and react rapidly. Yet, cells display robust behavior, partly because they separate molecules into distinct compartments. One important class of compartments is biomolecular condensates, which have now been identified in eukaryotes (1–3), prokaryotes (4–6), and plants (7, 8). In all systems, multiple different condensates coexist, and some condensates, like the nucleolus (9) and nuclear speckles (10), even possess subcompartments. The collective organization of biomolecules into condensates is explained by phase separation (11), which is a physical mechanism where a gain in enthalpic interactions offsets the entropy loss when molecules are confined. Since all proteins interact weakly by various mechanisms (12), phase separation is widely expected in the proteome (13) and transcriptome (14). However, it is still mysterious how cells regulate phase separation.

Biomolecular condensates need to form robustly, despite internal and external uncertainties that cells cannot control. Having the right condensates, in the right situation, at the right time is crucial since condensates participate in almost all cellular processes (15), they affect the fitness of prokaryotes (16), and malfunctioning is implicated in many diseases (17). It is particularly mysterious how cells reliably form many different kinds of condensates in a common cytosol, despite copy number fluctuations of all components. Are the interactions between components tuned such that the right condensates form reliably? It is conceivable that multiple driving forces of phase separation (12) have been adjusted over evolutionary time scales. Indeed, theoretical studies (18, 19), numerical simulations (20, 21), and in vitro experiments (22, 23) demonstrated that small modifications of the sequence of a protein can have profound impact on its phase separation behavior. However, it is not clear whether these results on single components can be transferred to multicomponent mixtures.

While the theoretical basis of phase separation is well understood in binary mixtures (11, 24–26), even predicting equilibrium states is challenging in multicomponent mixtures. This is due to enormous variability in heterotypic interactions, which leads to complex phase diagrams (27). We can now construct complete phase diagrams for up to five components (28) and predict the associated phase morphology (29). This showed that the number of coexisting phases typically depends on the overall composition of the system, but it is unclear how this phase count depends on the component count and the specific interaction matrix. Answering this question is critical since typical biological condensates consist of many components (27, 30–32), and the scaffold-client picture (33), where a single scaffold component dominates the phase behavior, might not always apply (34). State-of-the-art numerical techniques can simulate spatially resolved compositions of up to 16 components (35, 36), but these techniques are often too costly to truly explore the

Author affiliations: [a]Max Planck Institute for Dynamics and Self-Organisation, 37077 Göttingen, Germany; and [b]Department of Bionanoscience, Delft University of Technology, 2629 HZ Delft, The Netherlands

Author contributions: D.Z. and L.L. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: david.zwicker@ds.mpg.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2201250119/-/DCSupplemental.

Published July 6, 2022.

space of possible interactions. The formation of many different phases can be studied elegantly when molecules are provided by reservoirs (37–39), but phase coexistence cannot be directly examined with this method. Instead, random matrix theory has been used to unveil parameter regimes of multicomponent fluids that lead to many coexisting phases (36, 40–43). However, it is unclear how well random interactions capture real proteins, which have evolved for millions of generations. In fact, it is unclear what properties of interacting proteins need to be conserved during evolution for robust phase separation behavior.

Results

We here present an approach to analyze multiphase equilibrium states of multicomponent liquids that is based on relaxation dynamics. We then use this model to investigate how components need to interact such that a given number of phases form reliably.

A Simplified Physical Model Reveals Equilibrium States. We consider an isothermal, incompressible liquid composed of N different components and an inert solvent. In equilibrium, such a system can in principle form $N+1$ liquid phases (44), which are homogeneous regions with distinct composition. However, in typical realistic systems, fewer phases form since some components might be miscible. To reveal how the number of phases formed depends on the interactions of the components, we consider the general case of M coexisting phases with volumes $V^{(n)}$ for $n=1,\dots,M$. Since phases are homogeneous, their composition is fully described by the particle counts $N_i^{(n)}$ for each component $i=1,\dots,N$ or the associated volume fractions $\phi_i^{(n)}=\nu N_i^{(n)}/V^{(n)}$, where we consider equal molecular volumes ν for simplicity. Note that the fraction of the inert solvent, $\phi_0^{(n)}=1-\sum_{i=1}^{N}\phi_i^{(n)}$, is not an independent variable. Multiple phases can coexist when the associated free energy $F=\sum_{n=1}^{M}V^{(n)}f(\{\phi_i^{(n)}\})$ is minimal, where f is the free energy density that depends on the local composition. We here consider regular solution theory (45),

$$f(\{\phi_i\})=\frac{k_\mathrm{B}T}{\nu}\left[\phi_0\ln(\phi_0)+\sum_{i=1}^{N}\phi_i\ln(\phi_i)+\sum_{i,j=1}^{N}\frac{\chi_{ij}}{2}\phi_i\phi_j\right], \quad [1]$$

where $k_\mathrm{B}T$ is the thermal energy scale and the first two terms capture the entropic contributions of the solvent and all other components, respectively. Conversely, the last term quantifies the enthalpic interaction between all components. The elements of the interaction matrix χ_{ij} can, for instance, be derived from the interaction energies w_{ij} between components i and j on a lattice, $\chi_{ij}=z(2w_{ij}-w_{ii}-w_{jj})/(2k_\mathrm{B}T)$, where z is the lattice coordination number (28, 46). This implies that the diagonal entries vanish, $\chi_{ii}=0$, while the off-diagonal entries capture the relevant balance between heterotypic and homotypic interactions (Fig. 1*A*). Note that effective repulsion ($\chi_{ij}>0$) can originate not only from heterotypic repulsion ($w_{ij}>0$) but also from homotypic attraction that outweighs the heterotypic interaction ($w_{ii}+w_{jj}<2w_{ij}$).

The multicomponent liquid reaches equilibrium when F is minimal, implying that the chemical potentials $\mu_i=\nu\partial f/\partial\phi_i$ and the pressures $P=\sum_i\phi_i\partial f/\partial\phi_i-f$ are equal between all phases (26). We express these quantities in nondimensional form, $\hat{\mu}_i=\mu_i/k_\mathrm{B}T$ and $\hat{P}=P\nu/k_\mathrm{B}T$, for each phase n,

$$\hat{\mu}_i^{(n)}=\ln(\phi_i^{(n)})-\ln(\phi_0^{(n)})+\sum_{j=1}^{N}\chi_{ij}\phi_j^{(n)} \quad [2a]$$

$$\hat{P}^{(n)}=-\ln(\phi_0^{(n)})+\sum_{i,j=1}^{N}\frac{\chi_{ij}}{2}\phi_i^{(n)}\phi_j^{(n)}. \quad [2b]$$

The equilibrium conditions for the system then read

$$\hat{\mu}_i^{(1)}=\hat{\mu}_i^{(2)}=\cdots=\hat{\mu}_i^{(M)} \quad \text{and} \quad [3a]$$

$$\hat{P}^{(1)}=\hat{P}^{(2)}=\cdots=\hat{P}^{(M)}, \quad [3b]$$

for $i=1,\dots,N$, which are $(M-1)N$ and $M-1$ nonlinear equations, respectively. Additionally, there are N equations for the conservation of particles, $\sum_n N_i^{(n)}=\text{const}$, and an equation for volume conservation, $\sum_n V^{(n)}=\text{const}$. Taken together, these equations can in principle be solved for the M volumes $V^{(n)}$ and NM particle counts $N_i^{(n)}$, although this is generally challenging (28).

The equilibrium conditions [**3a**] and [**3b**] describe the local coexistence of phases of potentially different composition $\phi_i^{(n)}$. Since these conditions only involve the intensive quantities $\phi_i^{(n)}$, coexisting volume fractions can be determined without specifying the extensive volumes $V^{(n)}$, similar to the Maxwell construction in a binary system (26). To extend this idea to multicomponent liquids, we exploit that all systems relax to equilibrium by exchanging particles between phases until Eqs. **3a** and **3b** are satisfied. In the simplest case, this exchange is described by

$$\partial_{\hat{t}}\phi_i^{(n)}=\phi_i^{(n)}\sum_{m=1}^{M}\left[\phi_i^{(m)}\left(\hat{\mu}_i^{(m)}-\hat{\mu}_i^{(n)}\right)+\hat{P}^{(m)}-\hat{P}^{(n)}\right], \quad [4]$$

where $\hat{t}$ is a nondimensional time (*SI Appendix*, SI Text). Clearly, Eq. **4** is at a stationary state, $\partial_{\hat{t}}\phi_i^{(n)}=0$, when the equilibrium conditions [**3a**] and [**3b**] are obeyed. We show in *SI Appendix*, SI Text, that the converse is also true, so the relaxation dynamics given by Eq. **4** lead us to equilibrium states whose composition we can then analyze further (*SI Appendix*, Fig. S1). These states depend on the interaction matrix χ_{ij} and the initial compositions $\phi_i^{(n)}(\hat{t}=0)$. Consequently, if we sampled all initial compositions, we would discover all possible equilibrium configurations and thus recover the phase diagram associated with a particular interaction matrix χ_{ij} (28).

Practically, we sample phase space by initializing the M phases with random compositions chosen uniformly over all allowed volume fractions (*SI Appendix*, SI Text). This accounts for the typical cellular situation where concentrations fluctuate widely and where initial compositions of small phases depend on the details of nucleation (47, 48). This initialization also precludes abnormal situations where a component is present only in trace amounts in all phases (*SI Appendix*, Fig. S2). Random initial conditions imply that the dynamics described by Eq. **4** can lead to various stationary states, even for identical interaction matrices. Fig. 1*D* shows trajectories for two representative cases, revealing the typical situation that some phases reach identical composition. We determine the number K of distinct, coexisting phases by clustering all M phases based on the similarity of their final composition (*Materials and Methods*). Repeating this procedure over many random initial conditions, we can estimate the distribution $P(K)$,

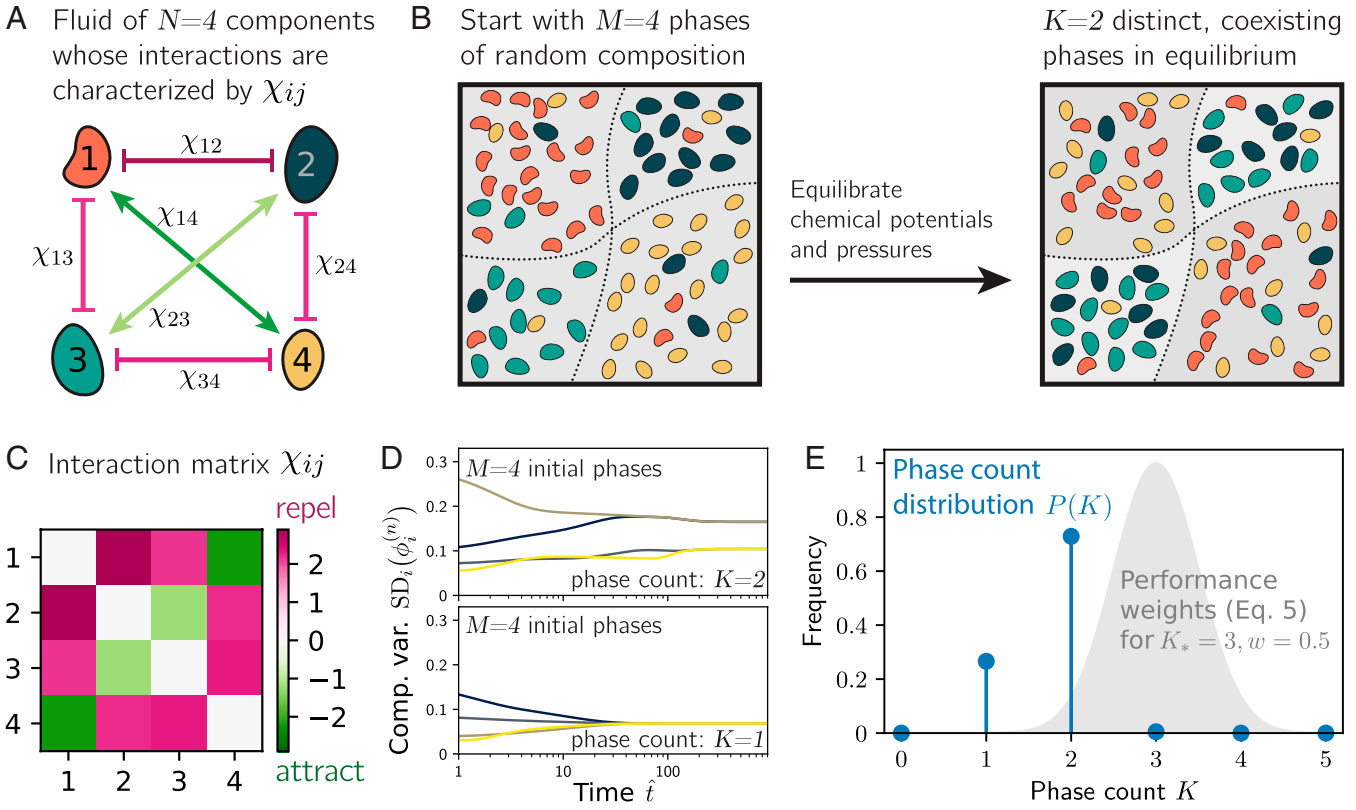

Fig. 1. A dynamical system recovers coexisting phases of multicomponent liquids. (*A*) Schematic of $N = 4$ components with attractive ($\chi_{ij} < 0$, green arrows between orange/red and teal/gray components) and repulsive ($\chi_{ij} > 0$, pink lines between remaining pairs) interactions. (*B*) Schematic showing how the N-component liquid is initially split into $M = 4$ phases of random composition. After equilibrating chemical potentials and pressures, only $K = 2$ phases of distinct composition remain. (*C*) Interaction matrix χ_{ij} corresponding to *A*. (*D*) Two representative simulations with different initial composition for χ_{ij} of *C*. The composition variation $[\langle\langle\phi_i^{(n)}\rangle^2\rangle_i - \langle\phi_i^{(n)}\rangle_i^2]^{1/2}$ is shown as a function of time $\hat{t}$. (*E*) Frequencies $P(K)$ of phase counts K for random initial conditions. The performance g follows from a convolution of $P(K)$ with weights (gray area) (Eq. **5**).

which corresponds to the frequency with which K different condensates form simultaneously inside a cell. While cells surely also control compositions of these condensates, we focus on their number since this is a more fundamental requirement, e.g., to prevent formation of aberrant condensates.

Random Interactions Do not Lead to Reliable Phase Counts. To gain intuition for the behavior of the multicomponent system, we first consider random interaction matrices χ_{ij}. To compare with the literature (36, 40–42), we draw entries independently from a normal distribution with mean $\bar{\chi}$ and variance σ_χ^2 (Fig. 2*A*). For each parameter pair $(\bar{\chi}, \sigma_\chi^2)$, we investigate 10^4 realizations of χ_{ij} and initial compositions and summarize the resulting distribution $P(K)$ by its mean and SD. Fig. 2*B* shows that only a single phase forms when interactions are generally weak (low $\bar{\chi}$ and σ_χ), consistent with an ideal solution where entropy favors mixing. When interactions are increased without strong variations (larger $\bar{\chi}$, low σ_χ), a demixing transition happens at $\bar{\chi} \approx \bar{\chi}_{\mathrm{crit}}$, and $K \approx N + 1$ phases are typical at large $\bar{\chi}$. Here all components segregate from each other and form separate phases, each enriched in a single component. Fig. 2*C* shows that the width of the phase count distribution, $\mathrm{SD}(K) = \langle(K - \langle K\rangle)^2\rangle^{1/2}$, is largest in the transition zone, indicating that the actually observed K strongly depends on the chosen interaction matrix and initial composition. The critical value $\bar{\chi}_{\mathrm{crit}}$, where the demixing transition takes places, increases with the component count N (Fig. 2*D*), which confirms a trend that was observed in earlier work (40, 42). Fig. 2*B* also shows that the width of the transition zone is generally broader for larger σ_χ, consistent with the fact that interactions are more variable. Interestingly, the statistics of the phase count K become independent of $\bar{\chi}$ for large variations σ_χ. In this case, we observe $K \approx N/2$, which was previously conjectured for $\bar{\chi} = 0$ (36). Taken together, our simplified dynamics are consistent with known results for random matrices.

The results shown in Fig. 2*B* indicate that random interactions of N components can lead to approximately 1, $N/2$, and N phases in large regions of the parameter space, while other values require fine-tuning. Even if it is possible to find parameters $\bar{\chi}$ and σ_χ that on average lead to a desired phase count K_*, it will not always be reached since the actual distribution of the number of phases, $P(K)$, possesses a significant width (Fig. 2*C*). To quantify how well the system reaches a target phase count K_*, we define the performance

$$g = \sum_{K=1}^{K_{\max}} P(K) \exp\left[-\frac{(K - K_*)^2}{2w^2}\right], \qquad [5]$$

which is constructed such that $0 < g \leq 1$ and $g = 1$ if and only if all initial conditions lead to K_* phases. Here w controls how strongly deviations from the target K_* are punished (Fig. 1*E*). Fig. 2*E* shows that the maximal performance of the random ensemble is $g \approx 0.3$, even though the choice $K_* = 5$ is close to $N/2$, so large σ_χ leads to $\langle K\rangle \approx K_*$. Indeed, *SI Appendix*, Fig. S3, shows that random ensembles perform even worse for other targets K_*. Taken together, it is thus not sufficient to vary the two parameters $\bar{\chi}$ and σ_χ of the random interactions to obtain a particular phase count reliably.

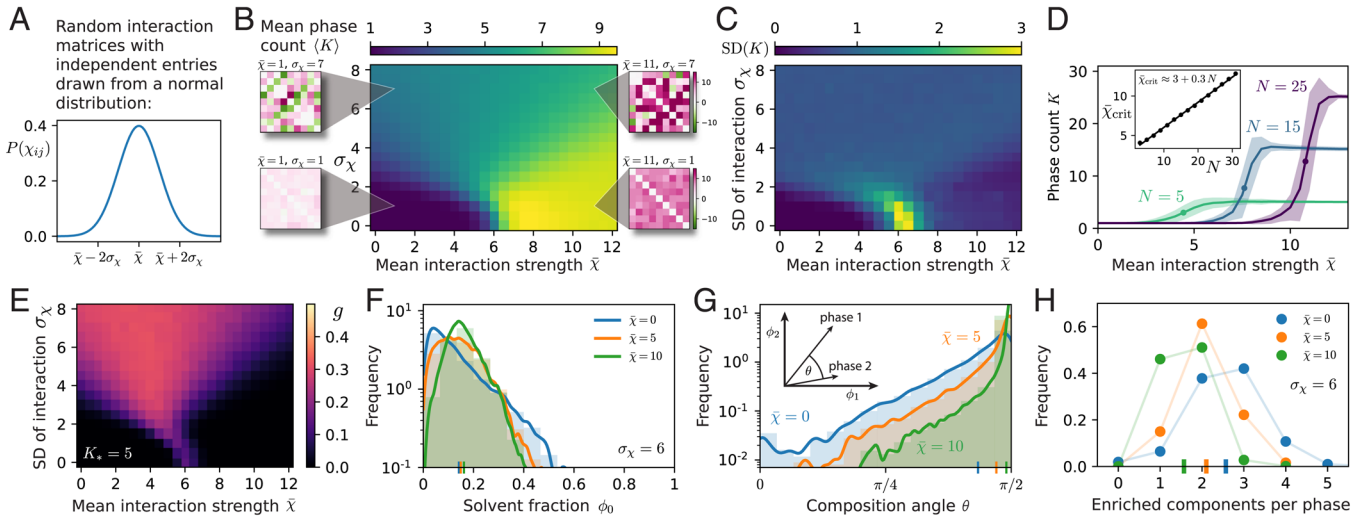

Fig. 2. Random interaction matrices cannot target specific phase counts K reliably. (*A*) Distribution $P(\chi_{ij})$ of the entries of the interaction matrix. (*B* and *C*) $\langle K\rangle$ and SD(K) as a function of the mean $\bar{\chi}$ and SD σ_χ of the distribution for the interactions χ_{ij} for $N = 9$ components (*Insets* show example matrices). The SD is calculated over initial conditions and averaged over the ensemble of χ_{ij}. (*D*) $\langle K\rangle \pm$ SD(K) as a function of $\bar{\chi}$ at $\sigma_\chi = 1$ for $N = 5$, 15, 25. The dot indicates the demixing transition point $\bar{\chi}_{crit}$. (*Inset*) $\bar{\chi}_{crit}$ as a function of the component count N (linear fit, $\bar{\chi}_{crit} = 3 + 0.3\,N$). (*E*) Performance g associated with data from *B* for $K_* = 5$ and $w = 1$. (*F* and *G*) Distribution of the solvent fraction ϕ_0 (*F*) and composition angles θ (*G*) shown as histograms and using kernel density estimation (lines) for $N = 9$, $\sigma_\chi = 6$, and several $\bar{\chi}$. (*H*) Distribution of the number of components enriched in phases for $N = 9$, $\sigma_\chi = 6$, and several $\bar{\chi}$. In *B*–*H*, averages are over 10^4 realizations, and distribution means are indicated as vertical bars on the horizontal axes.

Equilibrium phases resulting from random interactions also show strong composition variations. For instance, the solvent fraction ϕ_0 varies between 0 and 0.5 (Fig. 2*F*). We quantify differences of phase compositions using the composition angle $\theta_{nm} = \arccos\left(\vec{\phi}_n \cdot \vec{\phi}_m / |\vec{\phi}_n| \cdot |\vec{\phi}_m|\right)$, which is simply the angle between the composition vectors $\vec{\phi}_n = (\phi_1^{(n)}, \ldots, \phi_N^{(n)})$ of two phases n and m (36) (Fig. 2 *G*, Inset). Note that θ is zero when phases have identical composition (but not necessarily the same total concentration), while $\theta = \frac{\pi}{2}$ when compositions are orthogonal, i.e., when they have no components in common. While compositions of initial phases are similar (*SI Appendix*, Fig. S2*F*), they typically become very different after equilibration (Fig. 2*G*). In particular, the mean difference increases with stronger repulsion (larger $\bar{\chi}$). However, even for the strongest repulsion, there is significant overlap between phases, indicating that components are not cleanly sorted into distinct phases. To quantify this, we count for each phase how many components have a fraction larger than 1.5 times the average fraction. The number of such enriched components is smaller for stronger interactions, although it varies widely (Fig. 2*H*). Taken together, typical random interaction matrices cannot provide a reliable phase count K, so some additional structure is required.

Structured Interactions Can Lead to Reliable Phase Counts. To elucidate what structure in interaction matrices reliably leads to a desired phase count K_*, we next group the N interacting components in K_* clusters. We impose a repulsive interaction χ_+ between components belonging to different clusters, while components within a cluster exhibit a weak interaction χ_-. We expect that components in the same cluster cosegregate, so the system behaves as if it consisted of K_* effective components that all repel each other with strength χ_+. Indeed, Fig. 3*A* shows that demixing into many phases happens when χ_+ is sufficiently large, while the intracluster interaction χ_- has a weaker effect. Cosegregation even takes place when the intracluster interaction is slightly repulsive ($\chi_- > 0$). These designed matrices display expected behavior, and the resulting performance is high, particularly for large χ_+ (Fig. 3*B*). However, the segregation into distinct phases is not perfect, which is also visible in the distribution of the composition angles θ shown in Fig. 3*C*: even for strong repulsion (large χ_+), there is a significant fraction of phases with similar composition ($\theta \approx \pi/4$), even though exactly two components are enriched in each phase (Fig. 3*D*). It seems as if weakly concentrated components, including the solvent, prevent reliable cosegregation of clustered components. Taken together,

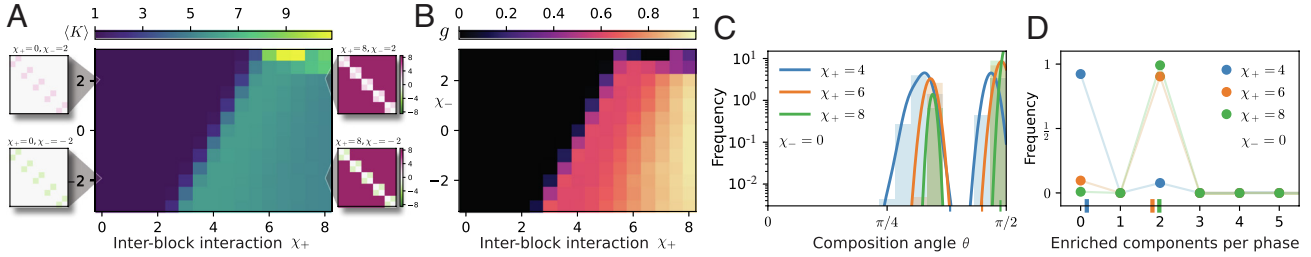

Fig. 3. Interaction matrices with block structure can target specific phase counts K reliably. (*A* and *B*) Mean phase count $\langle K\rangle$ and performance g (for $w = 1$) as functions of the interaction χ_+ between different blocks and the interaction χ_- within blocks for $N = 10$ components arranged in $K_* = 5$ equal blocks (*Insets* in *A* show example matrices). (*C*) Distribution of composition angles θ shown as histograms and using kernel density estimation (lines) for several χ_+ at $\chi_- = 0$. (*D*) Distribution of the number of components enriched in phases for several χ_+ at $\chi_- = 0$. *A*–*D* show averages over 10^4 initial compositions. Distribution means are indicated by vertical bars.

structured interaction matrices perform much better than random matrices, but they can still exhibit significant variations in phase counts.

Evolutionarily Optimized Interactions Lead to Reliable Phase Counts. Neither completely random nor fully structured interaction matrices are very realistic in biology since the interaction energies χ_{ij} summarize complex interactions of proteins (12), which change continuously during evolution (22). We thus next ask whether an evolutionary optimization of interaction matrices can reliably lead to mixtures with a particular target phase count K_*.

To mimic biology, we evolve an ensemble of individuals, characterized by interaction matrices χ_{ij}, over multiple generations. We initialize a population of 32 individuals with randomly chosen interaction matrices χ_{ij} using $\overline{\chi} = \overline{\chi}_{\text{crit}}$ and $\sigma_\chi = 1$. For each individual, we numerically determine $P(K)$ and the associated performance g (Eq. **5**), which will now play the role of a fitness. In the selection step, we remove the 30% of the population with lowest performance, replacing them by randomly chosen high-performance individuals to maintain population size. We then mutate the interactions χ_{ij} of all individuals by adding independent random numbers from a normal distribution with zero mean and SD σ_e. Repeating this procedure for many generations improves the performance of all individuals, so that they reliably reach the target K_*. However, we also noticed that this naive optimization results in very large interaction magnitudes (*SI Appendix*, Fig. S4), which might be unrealistic. To prevent such unphysical behavior, we additionally scale the interaction matrix χ_{ij} by $\chi_{\text{bound}} / \langle |\chi_{ij}| \rangle$ if its mean absolute value $\langle |\chi_{ij}| \rangle$ exceeds the threshold χ_{bound}. This limits the average interaction magnitude, $\langle |\chi_{ij}| \rangle \leq \chi_{\text{bound}}$, but the evolutionary optimization still discovers interaction matrices with a precise phase count (Fig. 4*A*) and perfect performance (Fig. 4*B*) (*SI Appendix*, Fig. S5). Optimized interaction matrices thus vastly outperform random matrices and allow targeting specific phase counts K_* despite strong fluctuations in initial composition.

The outstanding performance of evolved interaction matrices χ_{ij} is surprising since we limited the interaction magnitude (Fig. 4*C*) and use highly variable initial compositions. What properties of χ_{ij} lead to the excellent performance g? Simply visualizing optimized interaction matrices (Fig. 4*D*) does not reveal any obvious structure. The distribution of the entries χ_{ij} of optimized interaction matrices is very broad (Fig. 4*E*), although its width is directly limited by our constraint of $\langle |\chi_{ij}| \rangle$. For sufficiently large N, the distribution is well described by a normal distribution (dotted blue line). This is surprising since unstructured random matrices chosen from such a normal distribution did not perform well (Fig. 2). The similarity to the random ensemble also shows in the distribution of the solvent fraction ϕ_0 (compare Fig. 4*F* to Fig. 2*F*) and the number of enriched components (compare Fig. 4*H* to Fig. 2*H*). In contrast, the distribution of composition angles θ is slightly different (compare Fig. 4*G* to Fig. 2*G*). However, $\langle \theta \rangle$ is larger for the random ensemble with large $\overline{\chi}$, implying more distinct phases. Taken together, optimized matrices share many similarities with random matrices, although minute differences apparently lead to a much improved performance.

The evolutionary optimization quickly discovered interaction matrices that lead to a reliable phase count, and these matrices evolve continuously (Movie S1). This raises the question of whether this task is actually difficult; how frequent are optimal matrices in the space of all matrices? Our analysis of random matrices clearly showed that matrices must fulfill some basic requirements to have a phase count K close to the target K_*. In particular, the average magnitude of the entries χ_{ij} and the associated SD need to be chosen such that $\langle K \rangle \approx K_*$ (Fig. 2*B*). While we showed that the ensemble of random matrices with these properties does not work optimally (Fig. 2*E*), individual matrices from the ensemble might perform well. To quantify this, we determined the performance g for 64 random matrices characterized by a particular choice of $\overline{\chi}$ and σ_χ. *SI Appendix*, Fig. S6 shows that we easily discover matrices with high performance and that a large fraction of all matrices with suitable statistics, determined by $\langle \chi_{ij} \rangle$ and SD(χ_{ij}), performs optimally. Taken together, this implies that subtle properties of the interaction matrices govern their performance.

Performance of Evolutionarily Optimized Interactions Is Robust. We showed that interaction matrices leading to exactly K_* phases can be obtained through random trial and error or by

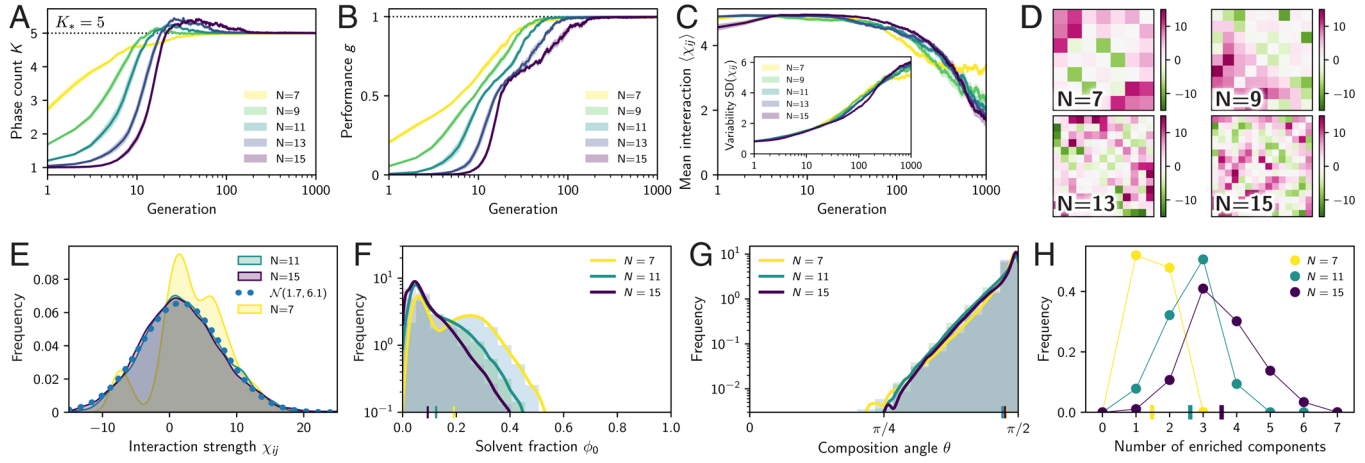

Fig. 4. Evolutionary optimized interactions result in reliable phase counts K. (*A* and *B*) Phase count K and performance g as functions of generation for different number of components N. The target phase count $K_* = 5$ and the maximal performance $g = 1$ are indicated by dotted lines. (*C*) Interaction strength $\langle \chi_{ij} \rangle$ and SD(χ_{ij}) as a function of generation. (*D*) Examples of optimized interaction matrices for various N. Components have been clustered by similarity. (*E*) Distribution of interaction strengths in the final matrices. The blue dotted line indicates a normal distribution that best fits the case $N = 15$. (*F* and *G*) Distribution of the solvent fraction ϕ_0 (*F*) and composition angles θ (*G*) shown as histograms and using kernel density estimation (lines) for several N. The corresponding means are indicated as vertical bars. (*H*) Distribution of the number of components enriched in phases for several N with means indicated as vertical bars. In *A*–*H*, model parameters are $K_* = 5$, $w = 1$, $\sigma_e = 0.3$, and $\chi_{\text{bound}} = 5$. Each population contains 32 individuals, which were initialized with random matrices. *A*–*C* for each N show ensemble mean and associated SE (SEM) of eight independent repetitions. The statistics of the resulting 256 optimized matrices are shown in *E*–*H*.

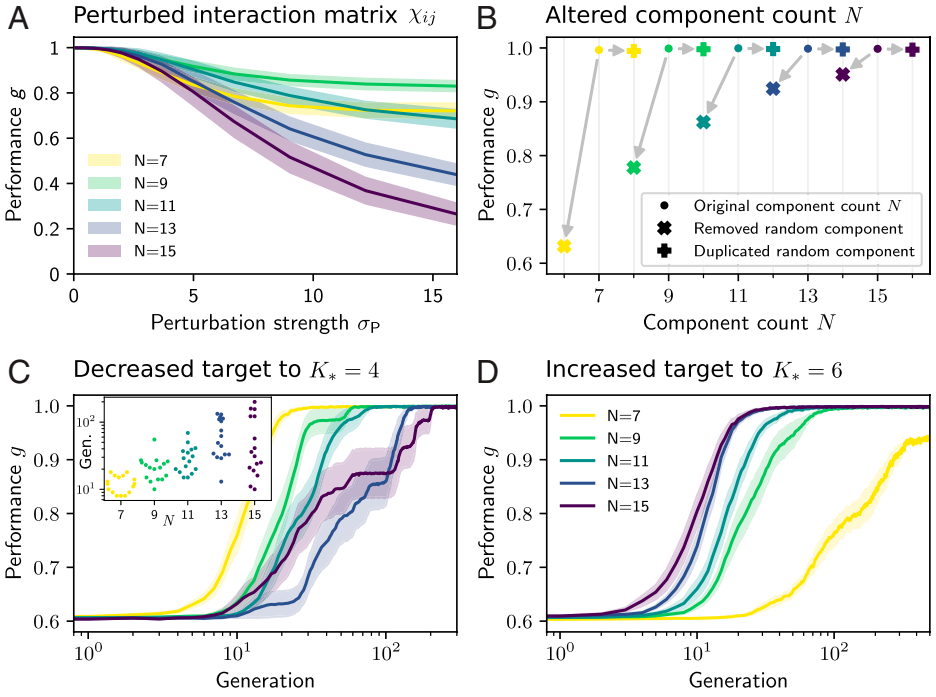

Fig. 5. Mixtures with more components are typically more robust. (*A*) Performance $\langle g \rangle \pm \mathrm{SD}(g)$ of optimized interaction matrices when one row (and column) is perturbed by normally distributed random numbers with SD σ_p for several component counts N at $K_* = 5$. Averages are over all 256 optimized matrices from Fig. 4. For each matrix, g is determined from an ensemble of 64 perturbations and random initial conditions. (*B*) Performance g when a mixture optimized with N components loses (cross symbols) or duplicates (plus symbols) a component for various N at $K_* = 5$. SD is smaller than symbol size. (*C* and *D*) Evolution of $\langle g \rangle \pm \mathrm{SEM}(g)$ as a function of generation when the target phase count is decreased from $K_* = 5$ to $K_* = 4$ (*C*) or increased from $K_* = 5$ to $K_* = 6$ (*D*) for several N. Each of the eight optimized populations from Fig. 4 is evolved twice. The *Inset* in *C* shows the generation at which individual trajectories exceed a performance of 0.8 for various N.

evolutionary optimization. This situation corresponds to maximizing the performance g in a fixed environment without any fluctuations beyond the initial composition. However, biological systems constantly face additional fluctuations, both internally (e.g., changes of the component count) and externally (e.g., changing environment). Such systems not only need to work in a particular case, but they need to be robust to these fluctuations, too. To see how evolution of multicomponent phase separation fares in such challenging situations, we next study the dynamics when interaction matrix χ_{ij}, the number N of components, or the target phase count K_* varies.

We start by perturbing a single component in the evolutionarily optimized interaction matrices χ_{ij} by choosing a random row (and column) to which we add uncorrelated random numbers from a normal distribution of vanishing mean and SD σ_p. Fig. 5*A* shows that the performance of optimized matrices is only weakly affected for $\sigma_\mathrm{p} \lesssim 2$, while larger perturbations reduce the performance significantly. Mixtures with more components are more sensitive to these perturbations, presumably because our procedure modifies the interaction between the chosen component and all other ones, so larger mixtures exhibit more perturbations. Taken together, we find that optimized mixtures still form phases reliably even when interaction energies are perturbed by $\sim k_\mathrm{B}T$.

We next test the robustness of the system against changes in component count N itself, which captures gene loss and duplication in real systems. We quantify the effect of changing N by measuring the performance g when one of the components of the optimized interactions matrices is removed or duplicated. Fig. 5*B* shows that removing a component reduces the performance substantially, although the reduction is smaller for larger N. Conversely, duplicating a component has hardly any effect on performance. Taken together, this suggests that using more components to form a fixed number of phases is more robust to internal fluctuations, like variations in component count.

We next consider external fluctuations of the environment. Since we do not model the environment explicitly, we consider changes of the target phase count K_*, assuming that the environment changes such that organisms need to form fewer or more phases. Changing K_* by 1 will necessarily reduce the performance from the optimal value $g \approx 1$ to $g = \exp(-\frac{1}{2}w^{-2}) \approx 0.6$ (Eq. **5**). To see how well different systems adapt to new environments, we study how quickly the performance recovers under the evolutionary dynamics. Fig. 5*C* shows that individuals quickly adjust to a lower target count, although the generation at which this happens varies widely (Fig. 5 *C*, *Inset*). This adaptation tends to be a bit slower for more components, presumably because more interactions have to be adjusted. Conversely, adaptation to an increased target count K_* is easier with more components (Fig. 5*D*). Note that the smallest system with $N = 7$ does not succeed to meet the target $K_* = 6$ reliably due to the constraint on $\langle|\chi_{ij}|\rangle$. Taken together, this suggests that there is a larger flexibility in the phase composition at larger N, which allows us to quickly find an interaction matrix resulting in an additional phase. Reducing the phase count is more complicated, likely because many interactions have to be adjusted. In fact, there is a tradeoff between robustly reaching a constant phase count K_* despite perturbations and using the same perturbations to flexibly adjust to new environments.

Discussion

Understanding equilibrium properties of biomolecules is crucial before we can tackle the more challenging problem of a living system. We here proposed a method to study how the many interacting constituents of a cell spontaneously segregate into

different phases. This method recovers the demixing transition that was previously observed when many components exhibit random interactions (36, 42). We also find a variable phase count for a given set of interactions, which is a signature of the complex phase diagrams (28). Beyond these limiting cases, our method can efficiently handle arbitrary interactions involving several tens of different components, thus increasing the range of systems that can be studied.

We use our method to optimize interaction matrices to yield a precise phase count. These optimized interactions are also robust to perturbations and allow a fast adaption to an increased target phase count, particularly if many components are involved. In contrast, forming fewer phases seems to be more challenging for larger mixtures, presumably because these mixtures are actually robust to perturbations. It will be interesting to study this trade-off between robustness and evolvability in more detail in the future.

Optimal interaction matrices are surprisingly easy to discover, and even random matrices have a high chance of yielding a robust number of phases, which is independent of the initial composition. On the contrary, other random matrices from the same ensemble perform much worse. What are properties that separate the optimal matrices from generic ones? Answering this question is directly relevant to biomolecular condensates, where hidden structures in intrinsically disordered regions might strongly affect the phase behavior of proteins (49). Another question concerns the composition of phases resulting from optimal interactions. So far, we focused on maintaining a robust phase count to ensure all relevant condensates, but no aberrant phases, form. However, the large space of optimal matrices might additionally allow the control of partitioning of key components. Our method can also be extended to describe more complex behavior of biomolecular condensates, including response to external cues (19, 50), active regulation (51–53), and noise buffering (54, 55). Ultimately, our predictions could be tested using engineered condensates (56) and quantitative reconstitution (32). Besides these concrete applications for biomolecular condensates, our method might also answer more fundamental questions about evolving systems: how can a cell exhibit robust functions while its proteins evolve (57–59)? We hope that our abstract model and method for determining equilibrium states will illuminate the fundamental problem of how variable microscopic interactions lead to robust collective properties.

Materials and Methods

We solve Eq. **4** using an explicit scheme with adaptive time stepping (60). Since the simulation typically converges exponentially, we conclude that a stationary state has been reached when all $\partial_t \phi_i^{(n)} < 10^{-4}$. For each choice of χ_{ij}, we run 64 simulations with random initial conditions (*SI Appendix*, SI Text) for $M = N + 2$ initial phases to estimate the distribution of the phase count K, which is the minimal number of points $\vec{x}_m$ so that $\min_m \left(\|\vec{x}_m - \vec{\phi}^{(n)}\| \right) \leq 10^{-2}$ for all phases n. The kernel density estimates shown in Figs. 2–4 are based on Gaussian kernels whose SD is 10 % of the range of the x axis (61).

Data Availability. Source code of the numerical method has been deposited in GitHub (https://github.com/zwicker-group/paper-multicomponent-evolution).

ACKNOWLEDGMENTS. We thank Evan Spruijt for a critical review of the manuscript and helpful discussions. D.Z. acknowledges funding by the Max Planck Society. L.L. gratefully acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement 758132) and funding from the Netherlands Organization for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) through a Vidi grant (016.Vidi.171.060).

1. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
2. M. Feric *et al.*, Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686–1697 (2016).
3. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
4. C. A. Azaldegui, A. G. Vecchiarelli, J. S. Biteen, The emergence of phase separation as an organizing principle in bacteria. *Biophys. J.* **120**, 1123–1138 (2021).
5. M. C. Cohan, R. V. Pappu, Making the case for disordered proteins and biomolecular condensates in bacteria. *Trends Biochem. Sci.* **45**, 668–680 (2020).
6. C. Greening, T. Lithgow, Formation and function of bacterial organelles. *Nat. Rev. Microbiol.* **18**, 677–689 (2020).
7. R. J. Emenecker, A. S. Holehouse, L. C. Strader, Biological phase separation and biomolecular condensates in plants. *Annu. Rev. Plant Biol.* **72**, 17–46 (2021).
8. J. Kim, H. Lee, H. G. Lee, P. J. Seo, Get closer and make hotspots: Liquid-liquid phase separation in plants. *EMBO Rep.* **22**, e51656 (2021).
9. D. L. J. Lafontaine, J. A. Riback, R. Bascetin, C. P. Brangwynne, The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **22**, 165–182 (2021).
10. J. Fei *et al.*, Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.* **130**, 4180–4192 (2017).
11. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
12. G. L. Dignon, R. B. Best, J. Mittal, Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* **71**, 53–75 (2020).
13. M. Hardenberg, A. Horvath, V. Ambrus, M. Fuxreiter, M. Vendruscolo, Widespread occurrence of the droplet state of proteins in the human proteome. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33254–33262 (2020).
14. D. A. Adekunle, A. Hubstenberger, The multiscale and multiphase organization of the transcriptome. *Emerg. Top. Life Sci.* **4**, 265–280 (2020).
15. A. S. Lyon, W. B. Peeples, M. K. Rosen, A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **22**, 215–235 (2021).
16. X. Jin *et al.*, Membraneless organelles formed by liquid-liquid phase separation increase bacterial fitness. *Sci. Adv.* **7**, eabh2929 (2021).
17. S. Alberti, D. Dormann, Liquid–liquid phase separation in disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).
18. K. L. Saar *et al.*, Learning the molecular grammar of protein condensates from sequence determinants and embeddings. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019053118 (2021).
19. J. M. Choi, A. S. Holehouse, R. V. Pappu, Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
20. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, e30294 (2017).
21. Y. H. Lin, J. D. Forman-Kay, H. S. Chan, Theories for sequence-dependent phase behaviors of biomolecular condensates. *Biochemistry* **57**, 2499–2508 (2018).
22. B. S. Schuster *et al.*, Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11421–11431 (2020).
23. A. Bremer *et al.*, Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* **14**, 196–207 (2022).
24. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).
25. J. Berry, C. P. Brangwynne, M. Haataja, Physical principles of intracellular organization via active and passive phase transitions. *Rep. Prog. Phys.* **81**, 046601 (2018).
26. C. A. Weber, D. Zwicker, F. Jülicher, C. F. Lee, Physics of active emulsions. *Rep. Prog. Phys.* **82**, 064601 (2019).
27. J. A. Riback *et al.*, Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209–214 (2020).
28. S. Mao, D. Kuldinow, M. P. Haataja, A. Košmrlj, Phase behavior and morphology of multicomponent liquid mixtures. *Soft Matter* **15**, 1297–1311 (2019).
29. S. Mao, M. S. Chakraverti-Wuerthwein, H. Gaudio, A. Košmrlj, Designing the morphology of separated phases in multicomponent liquid mixtures. *Phys. Rev. Lett.* **125**, 218003 (2020).
30. A. K. L. Leung, J. S. Andersen, M. Mann, A. I. Lamond, Bioinformatic analysis of the nucleolus. *Biochem. J.* **376**, 553–569 (2003).
31. D. L. Updike, S. Strome, A genomewide RNAi screen for genes that affect the stability, distribution and function of P granules in Caenorhabditis elegans. *Genetics* **183**, 1397–1419 (2009).
32. S. L. Currie, M. K. Rosen, Using quantitative reconstitution to investigate multicomponent condensates. *RNA* **28**, 27–35 (2022).
33. S. F. Banani *et al.*, Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
34. A. Musacchio, On the role of phase separation in the biogenesis of membraneless compartments. *EMBO J.* **41**, e109952 (2022).
35. S. Zhou, Y. M. Xie, Numerical simulation of three-dimensional multicomponent Cahn–Hilliard systems. *Int. J. Mech. Sci.* **198**, 106349 (2021).
36. K. Shrinivas, M. P. Brenner, Phase separation in fluids with many interacting components. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2108551118 (2021).
37. A. Murugan, Z. Zeravcic, M. P. Brenner, S. Leibler, Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 54–59 (2015).
38. P. Sartori, S. Leibler, Lessons from equilibrium statistical physics regarding the assembly of protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 114–120 (2020).
39. W. M. Jacobs, Self-assembly of biomolecular condensates with shared components. *Phys. Rev. Lett.* **126**, 258101 (2021).
40. R. P. Sear, J. A. Cuesta, Instabilities in complex mixtures with a large number of components. *Phys. Rev. Lett.* **91**, 245701 (2003).
41. W. M. Jacobs, D. Frenkel, Predicting phase behavior in multicomponent mixtures. *J. Chem. Phys.* **139**, 024108 (2013).

42. W. M. Jacobs, D. Frenkel, Phase transitions in biological systems with many components. *Biophys. J.* **112**, 683–691 (2017).
43. I. R. Graf, B. B. Machta, *Thermodynamic stability and critical points in multicomponent mixtures with structured interactions*. https://arxiv.org/abs/2110.11332 (2021). Accessed 22 October 2021.
44. J. W. Gibbs, On the equilibrium of heterogeneous substances. *Trans. Conn. Acad. Arts Sci.* **3**, 1–329 (1876).
45. P. I. Flory, Thermodynamics of high polymer solutions. *J. Chem. Phys.* **10**, 51–61 (1942).
46. J. W. Cahn, J. E. Hilliard, Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.* **28**, 258–267 (1958).
47. X. Xu, C. L. Ting, I. Kusaka, Z. G. Wang, Nucleation in polymers and soft matter. *Annu. Rev. Phys. Chem.* **65**, 449–475 (2014).
48. S. F. Shimobayashi, P. Ronceray, D. W. Sanders, M. P. Haataja, C. P. Brangwynne, Nucleation landscape of biomolecular condensates. *Nature* **599**, 503–506 (2021).
49. D. Moses *et al.*, Hidden structure in disordered proteins is adaptive to intracellular changes. bioRxiv [Preprint] (2021). https://www.biorxiv.org/content/10.1101/2021.11.24.469609v1. Accessed 26 November 2021.
50. O. Adame-Arana, C. A. Weber, V. Zaburdaev, J. Prost, F. Jülicher, Liquid phase separation controlled by pH. *Biophys. J.* **119**, 1590–1605 (2020).
51. M. Hondele, S. Heinrich, P. De Los Rios, K. Weis, Membraneless organelles: Phasing out of equilibrium. *Emerg. Top. Life Sci.* **4**, 331–342 (2020).
52. J. Söding, D. Zwicker, S. Sohrabi-Jahromi, M. Boehning, J. Kirschbaum, Mechanisms of active regulation of biomolecular condensates. *Trends Cell Biol.* **30**, 4–14 (2020).
53. J. Kirschbaum, D. Zwicker, Controlling biomolecular condensates via chemical reactions. *J. R. Soc. Interface* **18**, 20210255 (2021).
54. A. Klosin *et al.*, Phase separation provides a mechanism to reduce noise in cells. *Science* **367**, 464–468 (2020).
55. D. Deviri, S. A. Safran, Physical theory of biological noise buffering by multicomponent phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2100099118 (2021).
56. D. Bracha, M. T. Walls, C. P. Brangwynne, Probing and engineering liquid-phase organelles. *Nat. Biotechnol.* **37**, 1435–1445 (2019).
57. L. Laan, J. H. Koschwanez, A. W. Murray, Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *eLife* **4**, e09638 (2015).
58. E. T. Diepeveen, T. Gehrmann, V. Pourquié, T. Abeel, L. Laan, Patterns of conservation and diversification in the fungal polarization network. *Genome Biol. Evol.* **10**, 1765–1782 (2018).
59. F. Brauns *et al.*, Adaptability and evolution of the cell polarization machinery in budding yeast. bioRxiv [Preprint] (2020). https://www.biorxiv.org/content/10.1101/2020.09.09.290510v1. Accessed 6 December 2021.
60. D. Zwicker, *Project Source Code*, https://github.com/zwicker-group/paper-multicomponent-evolution (2022). Accessed 30 January 2022.
61. B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Routledge, 2018).