

# Theory and simulation of multiphase coexistence in biomolecular mixtures

William M. Jacobs\*

*Department of Chemistry, Princeton University, Princeton, NJ 08544, USA*

(Dated: May 8, 2023)

Biomolecular condensates constitute a newly recognized form of spatial organization in living cells. Although many condensates are believed to form as a result of phase separation, the physicochemical properties that determine the phase behavior of heterogeneous biomolecular mixtures are only beginning to be explored. Theory and simulation provide invaluable tools for probing the relationship between molecular determinants, such as protein and RNA sequences, and the emergence of phase-separated condensates in such complex environments. This review covers recent advances in the prediction and computational design of biomolecular mixtures that phase-separate into many coexisting phases. First, we review efforts to understand the phase behavior of mixtures with hundreds or thousands of species using theoretical models and statistical approaches. We then describe progress in developing analytical theories and coarse-grained simulation models to predict multiphase condensates with the molecular detail required to make contact with biophysical experiments. We conclude by summarizing the challenges ahead for modeling the inhomogeneous spatial organization of biomolecular mixtures in living cells.

## I. INTRODUCTION

The discovery that intracellular “organelles” can exist without membranes has revolutionized molecular and cellular biology [1, 2]. Many such intracellular structures, now collectively referred to as “biomolecular condensates,” have been proposed to form via phase separation [2–5]. Physically, this means that a surface tension holds the phase-separated condensate together, while individual biomolecules—including proteins, RNAs, and other small molecules—exchange between the condensate and the surrounding fluid in dynamic equilibrium. Phase-separated condensates represent a unique form of biological organization compared to traditional membrane-bound organelles, since the absence of a membrane allows for rapid assembly and disassembly in response to stimuli [3].

Over the past 15 years, an increasingly large number of biomolecular condensates have been identified [6]. Because of the wide range of biological phenomena in which condensates play a role, including both fundamental biological processes [7–14] and a variety of pathological conditions [15, 16], it is important to understand the biophysical mechanisms that control which biomolecules partition into specific condensates. Theoretical advances are needed to guide experiments probing the relationship between the properties of individual biomolecules and emergent condensate structures in complex environments. In particular, the physicochemical determinants of condensate composition and stability in heterogeneous intracellular environments—where thousands of biomolecular species are present—are only beginning to be explored. This review summarizes theoretical and simulation efforts in this direction using approaches based on equilibrium thermodynamics.

## A. Linking physicochemical properties and condensate thermodynamics

How do biomolecular determinants such as amino acid or nucleotide primary sequence, secondary/tertiary structure, and chemical modifications control the compositions and spatial organization of phase-separated intracellular condensates (Fig. 1)? This question has been addressed primarily within the context of equilibrium thermodynamics, in which the phase behavior of a macromolecular mixture is governed by free energies at thermal equilibrium. Within this framework, the partitioning of biomolecules into phase-separated condensates is determined by equilibrium chemical potentials, while condensate (dis)assembly dynamics are governed by free-energy gradients close to equilibrium and/or transitions between metastable states. Predictions based on this near-equilibrium assumption generally hold up well when tested against *in vitro* experiments [3, 4, 17]. Thus, while living systems may be more accurately characterized as nonequilibrium steady states under some conditions [18], we will restrict our attention to near-equilibrium approaches for predicting biomolecular phase separation in this review. We will also use the common terminology *liquid–liquid phase separation (LLPS)* [2, 19–22] to describe reversible thermodynamic phase transitions between (potentially complex) fluid phases with different macromolecular concentrations, as our discussion will focus on static properties such as condensate composition and spatial organization. Nonetheless, we note that condensed phases in biology often exhibit viscoelastic dynamical properties and may irreversibly age into solid phases due to the complexity of the interactions among biological macromolecules [6, 22–27].

Concepts from polymer physics have helped shape the prevailing view that transient associations among biomolecules give rise to the overall net attractive interactions required to bring about LLPS [28]. These interactions are commonly referred to as “multivalent,”

---

\* wjacobs@princeton.edu

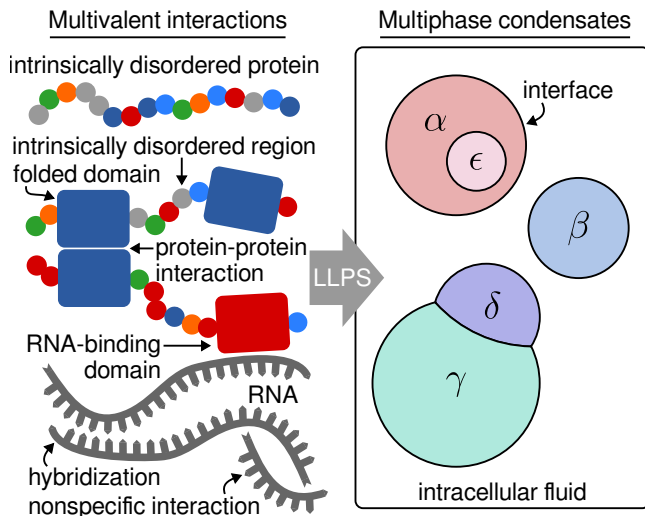


FIG. 1. Multivalent interactions among a wide variety of biological macromolecules, including intrinsically disordered proteins (with amino acids represented by colored circles), multidomain proteins, and nucleic acids, contribute to the thermodynamic driving forces responsible for liquid–liquid phase separation. Phase-separated condensates, including higher-order structures composed of multiple immiscible phases, resemble “membraneless organelles” whose interfaces are stabilized by surface tensions. The molecular compositions within each phase ( $\alpha$ – $\epsilon$ ) are distinct as a result of specific interactions among the constituent biomolecules.

since biomolecules can associate through multiple interaction sites via a variety of forms of noncovalent bonding. Particular attention has been given to conformationally heterogeneous proteins, including intrinsically disordered proteins (IDPs) and multidomain proteins containing intrinsically disordered regions (IDRs) [28]. In the context of IDPs, multivalency refers to the ability of an unfolded protein to engage in many residue–residue contacts with nearby proteins in a condensed phase. Folded domains within multidomain proteins can also contribute to the multivalency required to drive LLPS, either through protein–protein interactions (PPIs) [29] or, in the case of RNA binding domains (RBDs), through interactions with RNA [30]. Finally, nucleic acid mixtures can phase separate under certain conditions due to intermolecular base-pairing [31–33] and nonspecific association [34]. Importantly, the strengths of the net interactions among biopolymers in liquid-like condensates are typically comparable to the thermal energy, since the protein and nucleic acid constituents of biomolecular condensates can often remain fluid on biologically relevant timescales.

### B. Emergence of multiphase coexistence in complex biomolecular mixtures

Biological LLPS results in an enormous diversity of condensates in living cells. Each of these condensates is

associated with a specific chemical composition [35] and may be enriched in many distinct biomolecules relative to the surrounding intracellular fluid [36]. The biological functions of condensates derive directly from this compositional specificity, since the biochemical reactions that take place within the spatial confines of a condensate are dependent on the molecular concentrations that define the local environment. Theoretical descriptions of *in vivo* condensate assembly must therefore account for complex intracellular mixtures comprising thousands of protein and RNA species, which can all potentially interact with one another.

At the simplest level, it is important to distinguish between homotypic and heterotypic interactions between species of the same or different types, respectively. In multicomponent mixtures with strong heterotypic interactions, the tendency of any particular species to partition into a condensate depends on the concentrations of all its potential interaction partners [37]. A consequence is that the equilibrium compositions of coexisting phases may depend on the concentrations of all the components in the mixture, even when there are only two phases in coexistence. This feature can be used to detect the influence of multiple components on phase separation and to infer the relative strengths of homotypic and heterotypic interactions by measuring the volume fractions of coexisting phases at different overall mixture concentrations [33, 38].

Multiple immiscible condensates are commonly found to coexist within a single intracellular compartment [6]. Moreover, depending on the properties of the interfaces between pairs of condensates and between condensates and the surrounding fluid, immiscible condensates can self-organize into spatially organized structures [39]. Well characterized examples include the nucleolus [11, 40] and stress-granule/P-body condensates [29, 41, 42]. It has also become clear that subtle changes in protein and RNA concentrations can perturb the interfacial properties and thus dramatically alter the architecture of multiphase condensates [29, 43]. Nonetheless, predicting multiphase coexistence in the context of heterogeneous intracellular fluids remains a formidable challenge.

### C. Aims and scope of this review

Developing theoretical and computational models of multiphase, multicomponent biomolecular mixtures is essential for understanding the relationship between molecular determinants and biological self-organization via LLPS. The purpose of this article is to highlight a number of advances in this direction. Many recent reviews focusing on theory and simulation, including Refs. [4], [44], and [45], have described coarse-grained modeling approaches for IDPs, multidomain proteins, and nucleic acids. These approaches have primarily been applied to study the properties of single molecules and to mimic *in vitro* experiments on condensate formation.

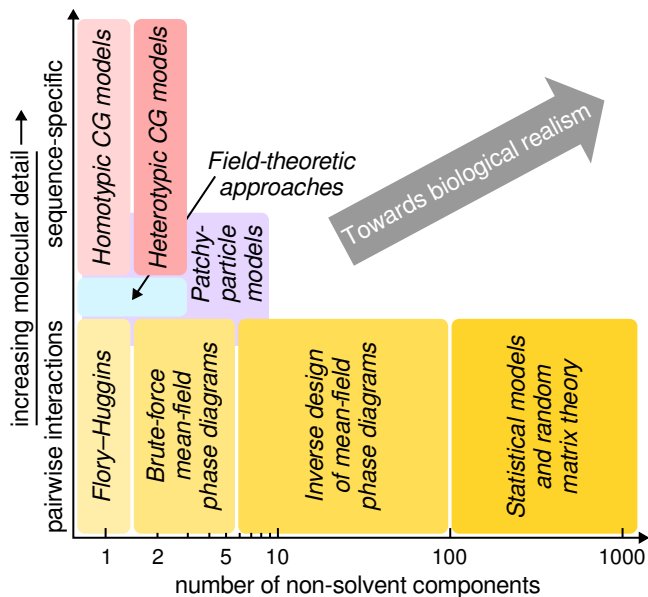


FIG. 2. Computational and theoretical complexity increases with both the level of molecular detail and the number of distinct components in a mixture. Simulation approaches to biomolecular LLPS range from pairwise-interaction mean-field models to sequence-specific coarse-grained (CG) models. However, mixtures with more than three non-solvent components have so far been studied almost exclusively using pairwise mean-field models.

By contrast, we focus here on theoretical challenges that arise when considering multiphase coexistence, especially in mixtures with thousands of components. Studies along these lines have provided complementary insights that are needed to understand biomolecular condensates in an intracellular context (Fig. 2). For broader context, we encourage the reader to consult other recent works, including reviews that emphasize the biological functionality and regulation of condensates [15, 46], the interplay between physical gelation and phase separation of multivalent macromolecules [47], and the conformational dynamics of macromolecules within condensates [48].

In this review, we begin in Sec. II by covering the thermodynamic principles of phase separation in multicomponent fluids. We highlight recently devised numerical methods for calculating multiphase coexistence in both mean-field and classical molecular simulation models. We then discuss theoretical results obtained from mean-field multicomponent mixture models in Sec. III. These studies have provided important insights into phase-behavior scaling relations, although they lack molecular detail and, as such, require assumptions on the statistical properties of intermolecular interactions in complex fluids. In Sec. IV, we examine efforts to describe multicomponent condensates with both analytical and computational models that capture the molecular sequence dependence or the structure of a PPI network. The implications of these studies for the mean-field multicomponent mixture models introduced in Sec. III, and potential extensions

thereof, are discussed. Finally, in Sec. V, we identify key challenges that must be overcome in order to describe inhomogeneous spatial organization in living cells with molecular realism.

## II. THERMODYNAMIC PRINCIPLES OF MULTICOMPONENT LLPS

Phase coexistence describes an equilibrium state in which a material or fluid exists in multiple phases with distinct physicochemical properties, such as oil droplets suspended in aqueous solution. Thermodynamic equilibrium between coexisting phases is established when the temperature, (osmotic) pressure, and chemical potentials of all molecular species are constant throughout the system. Considering a biomolecular solution at constant volume and temperature, the thermodynamic state of the system can be described by the Helmholtz free-energy density,  $f$ . This free energy is a function of the concentrations,  $\{\rho_i\}$ , of all the molecular components in the mixture. (Latin indices will be used throughout to indicate molecular components, while Greek indices will be used to indicate phases. Analogous arguments apply to the Gibbs free-energy density in the case of fluids at constant pressure.) Phase separation can occur when the free-energy density is a nonconvex function of the molecular concentrations (Fig. 3). In such a case, the free energy can be minimized by forming two or more distinct phases—for example, a condensed droplet and the surrounding cytoplasm—each with different concentrations. A mixture phase separates when the overall concentrations of the solution lie within the *coexistence region*, which is bounded by the concentrations of the coexisting phases. Droplets that emerge as a result of this spontaneous process are stabilized by positive surface tensions at the interfaces that form between the coexisting phases. Whenever  $f$  is nonconvex, there is also a *spinodal region* within which the free-energy surface has negative curvature.

In a heterogeneous system comprising many different types of biomolecules, the free-energy surface is a high-dimensional object. Nonetheless, coexistence and spinodal regions can still be determined by examining the convexity and local curvature of the free-energy surface. More precisely, the Hessian matrix  $\partial^2 f / \partial \rho_i \partial \rho_j$  is not positive definite within the spinodal region, implying that a homogeneous mixture within this region is unstable with respect to concentration fluctuations in one or more directions of concentration space. These directions are described by the eigenvectors that correspond to the negative eigenvalues of  $\partial^2 f / \partial \rho_i \partial \rho_j$ . The region of a high-dimensional concentration space in which concentration fluctuations are locally unstable is bounded by a spinodal locus, where the determinant  $|\partial^2 f / \partial \rho_i \partial \rho_j| = 0$ .

The molecular concentrations of coexisting bulk phases can be determined by considering the equal pressure and chemical potential conditions. In multicomponent flu-

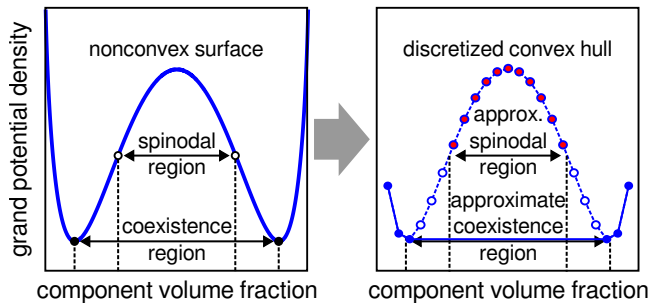


FIG. 3. *Left*: Nonconvex free-energy surfaces lead to phase separation at thermodynamic equilibrium. The inflection points and global minima of the grand potential density,  $\Omega \equiv f - \sum_{i=1}^N \rho_i \mu_i$ , determine the spinodal points and coexistence points, respectively. *Right*: Approximate phase diagrams can be obtained by computing the convex hull (solid line) of a discretized free-energy surface; points on the hull (filled circles) are in one-phase regions, while points not on the hull (empty and red-filled circles) are within a coexistence region. The approximate spinodal region can be determined by identifying points where the Hessian is not positive definite (red-filled circles). Approximate coexistence points can then be refined via nonlinear minimization (see text). This scheme generalizes to higher-dimensional concentration spaces.

ids, these conditions can be satisfied by performing a “common tangent plane construction,” in which a hyperplane is tangent to the free-energy surface at each point in concentration space that corresponds to a coexisting stable phase. A homogeneous mixture with an overall, or “parent”, concentration vector inside the convex hull of the coexisting-phase concentrations can lower its Helmholtz free energy by phase-separating. This convex hull therefore defines the coexistence region, which necessarily encompasses the spinodal region, in a multicomponent fluid. Because the tangent at any point on the free-energy surface is equal to the chemical potential vector,  $\{\mu_i\} = \partial f / \partial \rho_i$ , the common tangent plane construction ensures equal chemical potentials for each species across all phases that are in coexistence. Furthermore, the common tangent plane construction implies that the coexisting phases are all global minima of the grand potential density,  $\Omega(\{\rho_i\}; \{\mu_i\}) \equiv f(\{\rho_i\}) - \sum_i \rho_i \mu_i$ . This fact ensures equal pressures among all bulk phases.

In general, the coexistence concentrations in a multicomponent fluid are not specified uniquely without also prescribing the parent concentrations,  $\{\rho_i\}^{(\text{parent})}$ . The connection between the parent and coexisting-phase concentrations is provided by the conservation law

$$\rho_i^{(\text{parent})} = \sum_{\alpha=0}^K x^{(\alpha)} \rho_i^{(\alpha)}(\{\mu_j\}) \quad \forall i, \quad (1)$$

where  $\alpha$  indexes the phases in a phase-separated state with  $K + 1$  phases, the concentrations  $\{\rho_i^{(\alpha)}\}$  indicate coexisting phases with coexistence chemical potentials  $\{\mu_j\}$ , the volume fractions of the bulk phases are given by  $\{x^{(\alpha)}\}$ , and  $\sum_{\alpha=0}^K x^{(\alpha)} = 1$ . (This indexing convention is

chosen for later convenience, since we are often interested in phase equilibria involving a solvent-majority phase,  $\alpha = 0$ .) Eq. (1) simplifies to the well-known lever rule for binary mixtures (e.g., fluids comprising one macromolecular component plus a solvent).

The spinodal locus coincides with the boundary of the coexistence region at a critical point, where the concentrations of two coexisting phases merge into a single stable phase. Unlike binary mixtures, there is typically no unique critical point in a multicomponent fluid. Instead, multicomponent critical points lie on a temperature-and-concentration-dependent manifold with dimension one less than the number of non-solvent components. Higher-order critical points, where more than two phases simultaneously merge into a single stable phase, are also possible in multicomponent fluids [49].

Multicomponent phase equilibria can equivalently be determined from the excess chemical potential,  $\mu_{\text{ex},i}$ , of each molecular species  $i$ . This quantity represents the contribution to the chemical potential that captures all interactions—both enthalpic and entropic—among the molecules, and is thus a function of all the component concentrations [50]. The excess chemical potential is directly related to the partition coefficient, PC, defined as the ratio of a molecule’s concentration inside (in) and outside (out) of a phase-separated droplet:

$$\text{PC}_i \equiv \frac{\rho_i^{(\text{in})}}{\rho_i^{(\text{out})}} = \exp\left(\beta \mu_{\text{ex},i}^{(\text{out})} - \beta \mu_{\text{ex},i}^{(\text{in})}\right), \quad (2)$$

where  $\beta \equiv 1/k_B T$ ,  $k_B$  is the Boltzmann constant, and  $T$  is the absolute temperature. Partition coefficients are experimentally accessible and biologically relevant quantities, since they quantify the tendency of specific biomolecules to partition spontaneously into phase-separated condensates.

### A. Mean-field models with pairwise interactions

The simplest theoretical descriptions of LLPS are based on mean-field models, which introduce effective parameters to describe how molecules interact with one another. A mean-field model prescribes an approximate free-energy surface in terms of the effective interaction parameters and the component concentrations. The most widely used mean-field models, both in the condensate literature and more generally in biophysics and materials science, make the assumption that the excess chemical potential of species  $i$  can be written in the form

$$\mu_{\text{ex},i}(\{\rho_j\}) = \mu_v(\{\rho_j\}) + \beta^{-1} \sum_{j=1}^N B_{ij} \rho_j, \quad (3)$$

where  $\mu_v$  is a monotonically increasing function that depends only on the concentrations and the excluded volume associated with each molecular species. The second term embodies the assumption of “pairwise interactions”

among the  $N$  non-solvent components, where  $\{B_{ij}\}$  is an  $N \times N$  symmetric matrix of interaction parameters. This assumption underlies the regular solution model of phase-separating mixtures [51], the Flory–Huggins model of homopolymer phase separation [19], and the van der Waals model of non-ideal fluids [52].

The Flory–Huggins model [19] is commonly used to fit experimental data on biomolecular LLPS [53]. Assuming an incompressible fluid with  $N$  non-solvent species, the Flory–Huggins free-energy density is

$$\beta f v_0 = \sum_{i=1}^N \frac{\phi_i}{L_i} \log \phi_i + \phi_0 \log \phi_0 + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \epsilon_{ij} \phi_i \phi_j, \quad (4)$$

where the volume fraction occupied by species  $i$  is  $\phi_i = L_i v_0 \rho_i$ , the degree of polymerization of species  $i$  is  $L_i$ , the size of a monomer is represented by  $v_0$ , and the solvent-occupied volume fraction,  $\phi_0$ , is determined by the incompressibility constraint,  $\sum_{i=0}^N \phi_i = 1$ . We note that, within the context of this model, the “solvent” may itself represent a mixture including non-interacting macromolecules. The interaction parameters  $\{\epsilon_{ij}\}$  are dimensionless. Negative interaction parameters imply that molecules attract one another, while positive interaction parameters imply repulsion. Homotypic and heterotypic interactions are encoded in the on- and off-diagonal elements of  $\{\epsilon_{ij}\}$ , respectively. Eq. (4) is consistent with Eq. (3), since the interaction parameters only enter the free-energy density in a quadratic form. The contribution to the free-energy density from the pairwise interactions can also be written in terms of Flory  $\chi$  parameters,  $\chi_{ij} = \epsilon_{ij} - (\epsilon_{ii} + \epsilon_{jj})/2$ , by extending the sums in the final term of Eq. (4) to include the solvent (component 0) and replacing  $\epsilon_{ij}$  with  $\chi_{ij}$ . This change of variables introduces terms that are linear in  $\{\phi_i\}$  into the free-energy density, which have no effect on the phase behavior. With this alternate notation, the on-diagonal elements  $\{\chi_{ii}\}$  are zero by definition, and the homotypic interactions are encoded by the interactions with the solvent,  $\{\chi_{i0}\}$ .

Two non-solvent components are sufficient to reveal generic effects of homotypic versus heterotypic interactions. In such a mixture, two distinct types of phase transitions can occur: A “condensation” transition can occur if attractive heterotypic interactions are comparable to or stronger than any attractive homotypic interactions, while a “demixing” transition can occur if the heterotypic interactions are significantly less attractive than one or both of the homotypic interactions [54]. Both behaviors have been observed in numerical investigations of two-component-plus-solvent mean-field (e.g., [55]) and molecular simulation models (e.g., [56]). Condensation transitions are analogous to LLPS in simple one-component-plus-solvent fluids, implying that the phase diagram can be fully described by projecting the concentrations onto the parent composition vector [54]. By contrast, mixtures with dissimilar homotypic and heterotypic interaction strengths have more complex phase diagrams. For example, the implications of this complexity for concentration

buffering have recently been explored in Ref. [57] using a two-component-plus-solvent Flory–Huggins model. Concentration buffering was shown to be effective when the tie lines connecting the coexisting condensed and dilute phases are parallel to the concentration “noise distribution.” This observation follows from the generalized lever rule, Eq. (1), with  $K = 1$ , which implies that fluctuations of the parent concentrations in the direction  $\bar{\rho}^{(1)} - \bar{\rho}^{(0)}$  only modify the volume fraction of the condensed phase,  $x^{(1)}$ , leaving the “buffered” concentrations of both non-solvent species in the dilute phase unchanged.

## B. Constructing phase diagrams of multicomponent mean-field models

Moving beyond two-solute scenarios, the construction of high-dimensional phase diagrams becomes considerably more challenging (Fig. 3). An elegant approach for solving this problem in mixtures with up to approximately five non-solvent components was provided in Ref. [58]. This method exploits the fact that the common tangent plane construction is equivalent to convexification of a non-convex free-energy surface. In this method, the free energy of a mean-field model is first evaluated at every point of an  $N$ -dimensional grid over the physical domain  $\phi_i \geq 0 \forall i$  and  $\sum_{i=1}^N \phi_i \leq 1$ . The volume fractions at each grid point and the corresponding free-energy value constitute a single point within an  $(N + 1)$ -dimensional space. The convex hull of all the points within this  $(N + 1)$ -dimensional space can then be determined using standard algorithms [59]. Importantly, grid points that lie within coexistence regions are *not* part of the convex hull. Furthermore, the facets of the convex hull can be analyzed to determine the number of coexisting phases in a coexistence region. This algorithm can be used as “black-box” method for identifying coexistence regions, up to the resolution specified by the concentration-space grid, for any mean-field model.

In order to perform coexistence calculations to greater precision, it is necessary to identify the coexistence chemical potential vector that results in a grand potential with multiple global minima. An efficient approach described in Refs. [29] and [60] involves an iterative two-step algorithm. First, assuming a fixed chemical potential vector, the local minima of the grand potential are identified using initial guesses of each of the coexisting-phase concentrations. Then, the chemical potential vector is adjusted to bring the variance among the values of the grand potential at these local minima to zero. This second step establishes the coexisting phases as global minima of the grand potential. It is advantageous to use estimates of the coexisting-phase concentrations obtained from the convex-hull method as initial guesses when performing these nonlinear minimizations. A similar approach, in which the initial guesses for the coexisting-phase concentrations are obtained from a grid-based search for the spinodal region, was proposed in Ref. [61].

An alternative strategy for calculating phase coexistence has been provided in Ref. [62]. This method uses a nonphysical dynamical scheme, inspired by swapping molecules between metastable phases, in order to eliminate differences between the chemical potentials and the pressures of the phases. Starting from an initial guess of the component volume fractions in each of the  $K + 1$  coexisting phases, the dynamical scheme evolves the volume fractions in each phase  $\alpha$  according to

$$\frac{\partial \phi_i^{(\alpha)}}{\partial t} = \phi_i^{(\alpha)} \beta \sum_{\gamma=0}^K \left[ \phi_i^{(\gamma)} (\mu_i^{(\gamma)} - \mu_i^{(\alpha)}) + (P^{(\gamma)} - P^{(\alpha)}) \right], \quad (5)$$

where  $\{\mu_i^{(\alpha)}\}$  and  $P^{(\alpha)}$  are the component chemical potentials and the pressure, respectively, evaluated in the  $\alpha$  phase with the instantaneous volume fractions  $\{\phi_i\}^{(\alpha)}$ , and  $t$  is the fictitious time associated with these dynamics. At steady state, when  $\partial \phi_i^{(\alpha)} / \partial t = 0$ , Eq. (5) ensures that the phases meet the thermodynamic criteria for coexistence. Crucially, the results of this numerical approach, like the nonlinear minimization scheme described above, depend sensitively on the initial guesses for the coexisting-phase concentrations. In particular, if a candidate phase is not represented in the  $K + 1$  initial concentration vectors, then it is unlikely to be captured in the final set of coexisting phases.

### C. Multicomponent phase coexistence in molecular simulation models via free-energy calculations

Efficient approaches for calculating coexistence among an arbitrary number of fluid phases have also been devised for molecular simulation models. Such models specify a potential energy function that depends on the coordinates of all particles in the simulation volume. As such, Monte Carlo or molecular dynamics (MD) simulation methods are required to sample the configurational phase space. A wide variety of methods are available for computing coexistence between pairs of phases [63]. Within the condensate literature, direct coexistence simulations utilizing a “slab geometry” [44] have become popular due to the ease with which this approach can be implemented. However, in order to compute phase coexistence among a larger number of phases, it is advantageous to work in the grand-canonical ensemble. Grand-canonical phase-coexistence calculations are also ideal for minimizing finite size effects [64].

A robust approach for carrying out multiphase coexistence calculations utilizes a generalization of the multicanonical sampling method [65]. Influenced by earlier simulations of polydisperse fluids [66], Ref. [67] introduced a method to sample an isolated pair of phases in a grand-canonical simulation with multiple free-energy basins (Fig. 4). First, an order parameter  $\Delta\rho_{\alpha\beta} \equiv (\vec{\rho} - \vec{\rho}^{(\alpha)}) \cdot \hat{\nu}_{\alpha\beta}$ , where  $\hat{\nu}_{\alpha\beta} \equiv (\vec{\rho}^{(\beta)} - \vec{\rho}^{(\alpha)}) / |\vec{\rho}^{(\beta)} - \vec{\rho}^{(\alpha)}|$ , is defined to measure the distance along a linear path

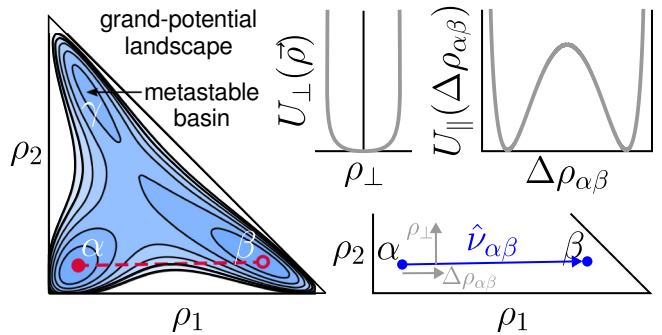


FIG. 4. Multiphase coexistence points can be determined from molecular simulations by sampling the grand-potential landscape. In order to sample two specific phases  $\alpha$  and  $\beta$ , biasing potentials parallel,  $U_{\parallel}$ , and perpendicular,  $U_{\perp}$ , to  $\hat{\nu}_{\alpha\beta}$  are introduced. Reweighting techniques can then be used to tune the component chemical potentials in order to establish equal grand potentials among all coexisting phases (see text).

between the  $\alpha$  and  $\beta$  phases, with concentration vectors  $\vec{\rho}^{(\alpha)}$  and  $\vec{\rho}^{(\beta)}$ , respectively. A biasing potential is then added to constrain fluctuations in orthogonal directions of concentration space,

$$U_{\perp}(\vec{\rho}) \equiv k_{\perp} \left| (\vec{\rho} - \vec{\rho}^{(\alpha)}) - [(\vec{\rho} - \vec{\rho}^{(\alpha)}) \cdot \hat{\nu}_{\alpha\beta}] \hat{\nu}_{\alpha\beta} \right|^{p_{\perp}}, \quad (6)$$

where  $k_{\perp} > 0$  and  $p_{\perp} > 0$  are user-defined constants. An additional biasing potential in the direction of concentration space parallel to  $\hat{\nu}_{\alpha\beta}$ ,  $U_{\parallel}(\Delta\rho_{\alpha\beta})$ , can then be calculated using grand-canonical Wang–Landau simulations [68],

$$\beta U_{\parallel}(\Delta\rho') = \log \int d\mathbf{x} \mathbf{1}_{\Delta\rho_{\alpha\beta}[\vec{\rho}(\mathbf{x}), \Delta\rho']} e^{-\beta\mathcal{H}(\mathbf{x}) - \beta U_{\perp}[\vec{\rho}(\mathbf{x})]}, \quad (7)$$

where  $\mathbf{x}$  represents a particle configuration,  $\mathcal{H}$  is the Hamiltonian of the unbiased model, and  $\mathbf{1}$  is the indicator function. The biasing potential  $U_{\parallel}$  is optimal for “flattening” the free-energy barrier between the  $\alpha$  and  $\beta$ -phase regions of phase space [68]. Finally, performing a multicanonical simulation under the combined potential  $\mathcal{H} + U_{\perp} + U_{\parallel}$  allows the simulation to transit reversibly between the  $\alpha$  and  $\beta$  phases.

Refs. [69] and [60] have demonstrated how this method can be applied to calculate multiphase coexistence points for multicomponent lattice models. Samples obtained from multicanonical simulations between different pairs of phases can be combined via reweighting methods such as MBAR [70] as long as one of the phases is sampled in every simulation. Grand potential differences between all pairs of phases can then be determined, and the chemical potentials can be adjusted in order to find the coexistence point at which all phases have identical pressures at equilibrium. This approach has been successfully applied to compute coexistence points involving more than five phases. Nonetheless, this method also requires prior knowledge of the approximate concentrations of all phases in order to construct the required biasing potentials and sample all the coexisting phases.

### III. PREDICTING AND DESIGNING PHASE BEHAVIOR IN MULTICOMPONENT FLUIDS

We now turn to theoretical studies of mixtures governed by pairwise interactions. We first discuss efforts to predict phase behavior in mixtures with hundreds or thousands of components based on the statistical properties of the pairwise interactions. We then describe recently devised methods to design or “evolve” pairwise interactions in order to stabilize a target phase diagram.

#### A. Multicomponent mixtures with random pairwise interactions

Pairwise interaction models, due to their simplicity, are a natural place to begin exploring how the presence of many distinct molecular components influence the phase behavior of a mixture. However, theoretical progress cannot be made without specifying the form of the interaction matrix, and limited systematic experimental data exist for parameterizing heterotypic interactions. To deal with this lack of information, Ref. [71] proposed that the pairwise interactions can be modeled using a random matrix. Specifically, Ref. [71] considered symmetric random matrices in which the elements are chosen independently from a Gaussian distribution with a prescribed mean and standard deviation. An ensemble of “random mixtures” is thus associated with a particular Gaussian distribution and the number of distinct components  $N$ , such that each mixture in the ensemble is defined by a particular realization of the  $N \times N$  interaction matrix.

Ref. [71] assumed for simplicity that the mixture free-energy density can be described by Eq. (3) with  $\mu_v = 0$ . The resulting free-energy density,  $f$ , is applicable to solutions in which all components are present at low concentrations, and the  $\{B_{ij}\}$  elements in Eq. (3) are referred to as second-virial coefficients [52]. By restricting the study to mixtures with equimolar parent concentrations,  $\rho_i^{(\text{parent})} = \bar{\rho}^{(\text{parent})} \forall i$ , it was shown that the spinodal locus can be predicted directly from the second-virial matrix. The central idea is that unstable concentration fluctuations can be determined from a linear stability analysis of the mean-field free-energy landscape (Fig. 5). With the equimolar parent-concentration assumption, the eigenvalue spectrum of the Hessian matrix,  $\partial^2 f / \partial \rho_i \partial \rho_j$ , is equal to the spectrum of  $\{B_{ij}\}$  plus a constant  $1/\bar{\rho}^{(\text{parent})}$ . Instabilities therefore occur when the minimum eigenvalue of  $\{B_{ij}\}$  is less than  $-1/\bar{\rho}^{(\text{parent})}$ . Applying results from random matrix theory, it was shown that the existence and nature of the dominant instability, which coincides with the minimum eigenvalue of the Hessian matrix, can be determined from the mean,  $b$ , and standard deviation,  $\sigma$ , of the Gaussian distribution of matrix elements in the limit of large  $N$ . Two distinct cases were observed. If the standard deviation among the matrix elements is sufficiently small, such that  $N^{1/2}b/\sigma \lesssim -1$ , then the dominant instabil-

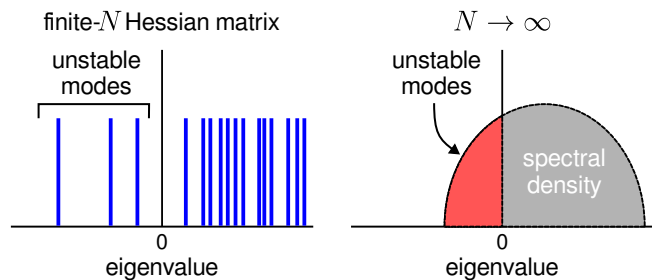


FIG. 5. The spinodal locus, where the mixture becomes unstable with respect to concentration fluctuations, can be predicted using a linear stability analysis. *Left*: Computing the eigenspectrum of the Hessian matrix,  $\partial^2 f / \partial \rho_i \partial \rho_j$ , at the parent concentrations reveals the number of unstable modes, each of which is associated with an orthogonal direction in concentration space. *Right*: Analytical predictions in the large- $N$  limit provide insight into the relationship between the structure and statistical properties of an interaction matrix and the phase behavior of the associated biomolecular mixture.

ity involves concentration fluctuations that are parallel to the equimolar parent concentration vector. This type of instability is consistent with a condensation transition driven by similar homotypic and heterotypic interaction strengths. By contrast, if the standard deviation among the matrix elements is sufficiently large, such that  $N^{1/2}b/\sigma \gtrsim -1$ , then the dominant instability is orthogonal to the parent concentration vector, and individual components demix into phases with differing compositions. Importantly, these behaviors are self-averaging, meaning that the tendency of any particular random-mixture realization to undergo a condensation or demixing transition converges in probability as  $N \rightarrow \infty$ .

Ref. [72] extended these results to mixtures with non-equimolar parent compositions. This work considered a regular-solution free-energy density, in which  $\mu_v = -\log \rho_0$  in Eq. (3). This additional contribution to the free energy accounts for the entropy of the solvent, providing a better physical model of solutions at non-dilute concentrations. Modifying the free-energy density in this way does not qualitatively alter the conclusions of Ref. [71] regarding condensation and demixing in equimolar mixtures. However, consideration of non-equimolar parent compositions reveals a third type of spinodal instability: Demixing transitions can now be classified as either “random,” in which all the components of the eigenvector associated with the instability are of similar order, or “localized,” in which the demixing transition is dominated by only a few species. Random mixtures with a large interaction-parameter variance and equimolar parent compositions tend to undergo random demixing. By contrast, mixtures in which one component has a much higher parent concentration than all the others can undergo a composition-driven transition, in which demixing is localized to the dominant species. The authors emphasized that the direction of composition-driven instabilities cannot be predicted simply by considering the rel-

ative parent concentrations of the components; instead, the interplay between entropic effects and random pairwise interactions tends to amplify the contribution of the dominant component to the unstable concentration fluctuations. In other words, the nature of the instabilities at the spinodal locus of a random mixture depends on both the interaction matrix and the parent concentrations.

Simulation support for the qualitative predictions of Ref. [71] was provided in Refs. [67] and [54]. In these studies, the free-energy calculation strategy described in Sec. II C was applied to compute coexistence between an equimolar dilute phase and a condensed phase in random mixtures with up to 64 non-solvent components. Simulations were conducted using a multicomponent lattice model, with the nearest-neighbor interactions between particles on the lattice specified by a random interaction matrix generated according to the Gaussian prescription of Ref. [71]. Coexistence calculations were then performed to investigate the nature of the phase transition that occurs at the lowest total parent concentration, meaning that the simulated coexistence point represents the lowest-concentration intersection of the equimolar parent concentration vector with any coexistence region. The average phase behavior of the random-mixture ensemble was analyzed by repeating these calculations for many independent realizations of random mixtures with the same interaction mean and variance.

Although the lattice-based coexistence calculations of Refs. [67] and [54] are not directly comparable to theoretical predictions regarding instabilities at the spinodal locus, analogous condensation and demixing transitions were observed in this molecular simulation model. First, the phase behavior at each simulated coexistence point was classified as condensation or demixing according to the angle,  $\theta$ , between the equimolar parent concentration vector and the unit vector connecting the coexisting phases,  $\hat{\nu}_{\alpha\beta}$ . This angle was found to be self-averaging with respect to the number of components,  $N$ , as suggested by random matrix theory [67]. Second, Ref. [54] observed that the distribution of  $\theta$  is bimodal, signifying a sharp transition between these two qualitatively distinct types of phase transitions as the mean and/or variance of the random-interaction distribution was changed. Third, increasing the number of components was found to shift the phase behavior at the simulated coexistence points towards condensation transitions, in line with the predictions of Ref. [71]. This finding implies that the mixing entropy of multicomponent fluids acts to suppress demixing instabilities. However, by contrast with Ref. [71], simulation results indicated that the extreme values of the interaction matrix are more predictive of the simulated coexistence concentrations than the eigen-spectrum of the mean-field Hessian matrix. This observation was exploited to propose a scaling relation for the transition between condensation and demixing behaviors at the phase boundary,  $(\log N)^{1/2} \sim \sigma$ , that differs from the random-matrix-theory prediction for the condensate-demixing crossover at the spinodal locus,  $N^{1/2} \sim \sigma/b$ .

This idea has since been followed up in Ref. [73], which suggested that the coexistence points can be strongly influenced by the tails of the distribution from which the elements of the random interaction matrix are chosen.

Phase separation in mean-field models of mixtures with many components has also been analyzed using phase-field simulations [74]. Deterministic phase-field simulations evolve the spatially varying component volume fractions,  $\{\phi_i(\vec{r})\}$ , on a three-dimensional grid in accordance with linear irreversible thermodynamics [75]. As such, phase-field simulations reach a steady state when the free energy of the simulated volume reaches a local minimum; this steady state may be spatially inhomogeneous if phase separation occurs. Ref. [74] considered a regular-solution free-energy density consistent with Eq. (3), with  $\mu_v = -\log \phi_0(\vec{r}) - \kappa \nabla^2 \phi_i(\vec{r})$ . The second term in  $\mu_v$ , which penalizes the formation of interfaces between phases in a component-independent manner, arises from square-gradient contributions to a Cahn-Hilliard free-energy functional with  $\kappa > 0$  [76]. Simulations then implemented ‘‘Model B dynamics’’ [77], where  $\partial \phi_i / \partial t = \nabla \cdot (M \phi_i \nabla \mu_i)$ , with a component-independent mobility coefficient  $M > 0$ . Upon reaching steady state, compositionally distinct phases were identified by performing a principal component analysis of the spatially varying component concentrations.

Since phase separation in a deterministic phase-field model proceeds via spinodal decomposition, Ref. [74] was able to provide a direct test of the analytical predictions of Ref. [71]. Both condensation and demixing were observed in simulations initialized with equimolar parent concentrations. Consistent with a linear stability analysis at these initial conditions (Fig. 5), Ref. [74] found that the number of phases identified at steady state correlates with the number of negative eigenvalues of the Hessian matrix. Furthermore, the number of steady-state phases could be estimated from the limiting ( $N \rightarrow \infty$ ) spectral density predicted by random matrix theory. This trend was shown to hold for a variety of random-mixture ensembles in which the standard deviation of the independently sampled interaction-matrix elements was either held constant or scaled proportionally to  $N^{1/2}$ . Nonetheless, some caution is warranted in interpreting these results, since the steady-state found via spinodal decomposition may reflect a metastable configuration that does not represent all the equilibrium phases. We shall return to this important consideration below in Sec. III C.

## B. Multicomponent mixtures with structured pairwise interactions

Although random-mixture models are useful for investigating generic features of high-dimensional phase diagrams, they may not reflect the structure of pairwise interactions among real biomolecules. In particular, the assumption that the elements of a  $\{B_{ij}\}$  matrix are independently and identically distributed im-



plies that  $\mathcal{O}(N^2)$  pairwise coefficients characterize the mixture, even though there are only  $N$  chemically distinct biomolecules. Physical interactions arising from the physicochemical features of the biomolecules are instead likely to introduce correlations into the  $\{B_{ij}\}$  matrix.

To address this critical issue, “structured” pairwise interaction models have been introduced and studied using linear stability analysis. Ref. [78] took the approach of grouping components into distinct families, whereby all members within a particular family have similar physicochemical properties. The authors proposed that this relationship could be described via by an interaction matrix of the form  $B = D + C * Z$ , where  $*$  indicates element-wise multiplication.  $D$  and  $C$  are block matrices specifying the mean and standard deviation of the interactions between families, respectively, while  $Z$  is a Gaussian random matrix with zero mean and unit variance. This model reduces to the random-mixture model of Ref. [71] when there is only one family, in which case all interactions have the same mean and variance. Intuitively, a single family of components can demix from a mixture with equimolar parent concentrations if the intra-family interactions are sufficiently more attractive than inter-family interactions. Such “family demixing” tends to dominate over random demixing when the noise amplitude, governed by  $C$ , is small.

Ref. [79] explored an alternative approach in which structured interaction matrices are assumed to have a low matrix rank. This assumption implies that the interaction matrix can be written in the form  $B_{ij} = \sum_{l=1}^r c^{(l)} s_i^{(l)} s_j^{(l)}$ , where the index  $l$  is bounded by the matrix rank,  $r$ . This low-rank decomposition was inspired by a toy model in which each molecular species can be described by  $r$  “molecular features,” which interact according to diagonalized coupling coefficients  $\{c^{(l)}\}$ . The matrix  $\{s_i^{(l)}\}$  specifies the value of each molecular feature for each component  $i$ . In fact, any  $N \times N$  interaction matrix can be written in this form via eigen-decomposition, assuming that  $N - r$  of its eigenvalues are negligible. If all the nonzero eigenvalues of  $\{B_{ij}\}$  are negative, representing net attractive interactions among the molecular features, then the linear-stability condition for the spinodal locus can be recast in terms of a feature covariance matrix. Specifically, this rank- $r$  matrix measures the covariance among the values of the molecular features, weighted by the concentrations of the components expressing these features, in a homogeneous mixture with fixed parent concentrations. The directions of the unstable concentration fluctuations can then be determined from the first principal component of the concentration-weighted molecular-feature distribution. This result bears resemblance to related studies of polydisperse fluids, in which phase transitions have been predicted using so-called “moment free energies” [80, 81]. When  $\{B_{ij}\}$  has both positive and negative eigenvalues, covariance matrices for the net-attractive and net-repulsive molecular features must be considered separately. The extent to which the net-repulsive fea-

tures modify the phase behavior depends on whether their concentration-weighted distribution correlates with that of the net-attractive feature distribution. The authors also showed that this analysis can be extended to predict ordinary and higher-order critical points, whose occurrence depends on higher-order cumulants of the concentration-weighted feature distribution.

An important insight gained from this theory [79] is that the phase behavior of a mixture can be predicted by analyzing properties of the  $r$ -dimensional feature space, which may be much simpler than the  $N$ -dimensional concentration space if  $r \ll N$ . Since intermolecular interactions among conformationally disordered biomolecules are widely believed to arise from a limited number of chemical interactions, such as electrostatic interactions among charged amino acids and hydrophobic forces involving amino acids with aromatic side chains, it is plausible that this is indeed the case. The relationship between this ansatz and findings from sequence-dependent theories will be discussed in Sec. IV. The work of Ref. [79] has also suggested a useful method for coarse-graining a multicomponent fluid into an equivalent binary mixture with the same spinodal and critical points by preserving the second and third cumulants along the first principal component of the concentration-weighted feature distribution. However, it is unclear whether the coexistence manifolds of multicomponent mixtures with low-rank interaction matrices can be simplified in the same way.

### C. Iterative design of multicomponent phase behavior

Taking the next step towards biologically realistic mixtures requires consideration of specific interactions that have emerged due to evolutionary processes. Recent efforts [60, 62, 69] to explore the thermodynamic consequences of evolved interaction specificity have shown that multicomponent mixtures can be designed with the goal of stabilizing a prescribed number of condensed phases. The logic behind this approach is that immense size of the space of possible biomolecular interactions limits the probability that a random-mixture model will produce a phase diagram comparable to the observed complexity of intracellular phase-separated condensates. Indeed, even in the simplest pairwise-interaction models, the “design space” has a dimension of  $N(N + 1)/2$  when all interactions are independently controllable. By contrast, treating multicomponent LLPS as an optimization problem in which the interactions can be systematically tuned has the potential to discover regions of this design space that are relevant to multiphasic condensates.

Ref. [62] demonstrated that the number of coexisting phases in a mean-field pairwise-interaction model can be designed by iterative application of a genetic algorithm. This design process necessitates finding all coexisting phases given a candidate interaction matrix at each iteration. The genetic algorithm is then applied to

evolve a population of interaction matrices in order to identify matrices that result in a target “phase count” of condensed phases. It turns out that this goal is surprisingly easy to achieve owing to the size of the design space when all pairwise interactions are independently tunable. An intuitive strategy of designing block-diagonal matrices, along the lines of Ref. [78], reliably results in phase counts equal to the number of blocks of strongly attractive interactions. However, the genetic algorithm finds solutions to this design problem that are less obviously structured. The authors further showed that designed mixtures with low phase counts tend to be stable with respect to small random perturbations in the interaction energies and that the genetic algorithm can rapidly alter the phase count of a designed mixture, finding new solutions within a few tens or hundreds of iterations.

This iterative design approach comes with a number of caveats, however. First, optimizing for a target phase count does not guarantee that different solutions identified by the genetic algorithm correspond to condensates with similar molecular compositions. Second, although the phase count of a candidate interaction matrix should depend on the parent concentrations according to Eq. (1), Ref. [62] employed a strategy of sampling coexistence points at random parent concentrations. This approach suggests an implicit design goal of maximizing the volume of the  $(K + 1)$ -phase coexistence region within the  $N$ -dimensional concentration space. Third, the reliability and performance of the iterative design algorithm are sensitive to the computational cost and accuracy, respectively, of the intermediate phase-coexistence calculations, which must be repeated for each candidate interaction matrix. This is in fact a very general problem: Regardless of the mixture model, phase-coexistence calculations first require a search for candidate phases, whether by exhaustive grid-based sampling (e.g., [58]; see Sec. II B), randomized initial conditions (e.g., [62]; see Sec. II B), Monte Carlo sampling (e.g., [54]; see Sec. II C), or physical dynamics (e.g., [74]; see Sec. III A). The computational cost of this search problem scales exponentially with the dimension of the concentration space.

#### D. Inverse design of multicomponent phase behavior

Many of the drawbacks of iterative design approaches can be overcome by directly solving the *inverse problem*—designing interactions to yield target phase behavior. Inverse design entails working out constraints on the solution space of biomolecular interactions that correspond to desired collective properties, such as the compositions of condensed phases (Fig. 6). Suitable interactions can be identified in this way without explicitly performing phase-coexistence calculations. As a result, the computational requirements may scale more favorably with the number of components, in particular because the initial search for candidate phases can be avoided.

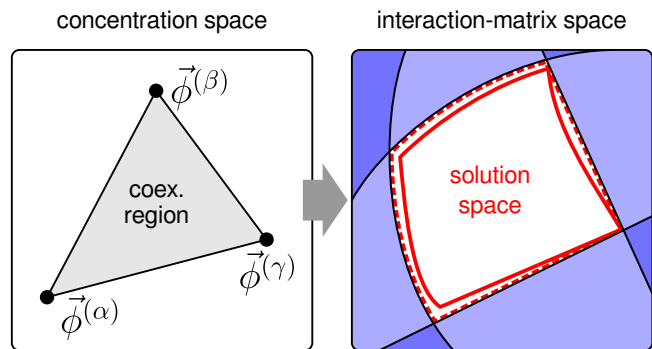


FIG. 6. In the inverse design approach, restrictions on the solution space of pairwise interaction matrices are determined directly from the concentrations of the target phases and the thermodynamic criteria for phase coexistence. *Left*: The target phase diagram consisting of condensed phases  $\{\vec{\phi}^{(\alpha)}\}$ . Any mixture with parent concentrations inside the convex hull of the target phases will phase-separate at equilibrium to establish coexisting phases with the prescribed concentrations. *Right*: Convex programming can be applied to compute the subspace containing interaction matrices that are consistent with the target phase behavior. The convex volume (red dashed line) bounded by the convex-optimization constraints (black lines) closely approximates the solution space to the inverse problem (red solid line). Because many interaction matrices may yield the same phase behavior, regularization is needed to select a particular matrix from the solution space.

An inverse design strategy for mixtures with pairwise interactions was first introduced in Ref. [69]. Because Eq. (3) is linear with respect to  $\{B_{ij}\}$ , the inverse problem can be solved approximately using a convex relaxation. It is therefore possible to prove, within the convex relaxation, whether a pairwise interaction matrix exists for a prescribed set of immiscible phases, and if so, to calculate a suitable interaction matrix with efficient convex programming algorithms [82]. Ref. [69] showed that the thermodynamic requirements for establishing metastable phases with prescribed compositions yield a convex relaxation known as a semidefinite program (SDP). The SDP constraints comprise both affine and eigenvalue inequalities, since the Hessian matrix must be positive definite in each target phase. Solutions to this SDP were shown to result in metastable phases with the desired compositions in mixtures with up to 200 distinct components, both in the context of a Flory–Huggins mean-field model and in Monte Carlo simulations of an associated multicomponent lattice model.

Exploiting the ability to prove feasibility of the SDP, Ref. [69] then studied the probability of finding a feasible solution for an inverse problem with randomly assigned target-phase compositions. This probability was found to drop sharply beyond a certain number of target phases, revealing a thresholding transition reminiscent of the storage capacity in the Hopfield model of neural networks [83] and “multifarious” self-assembly of finite-sized structures [84]. The critical number of condensed phases

associated with this thresholding transition could be predicted using graph-theoretic arguments and depends on both the number of components in the mixture and the fraction of components whose concentrations are enriched in each target phase relative to the surrounding fluid.

A similar convex optimization approach was then applied to design mixtures with prescribed equilibrium phases [60]. A two-step procedure for designing pairwise interaction matrices was proposed. First, a convex relaxation was used to specify an SDP for both the interaction matrix and the approximate coexistence chemical potential vector. Then, the chemical potentials were adjusted to ensure coexistence among the target phases using the nonlinear algorithm described in Sec. II B. A regularization heuristic was also introduced to pick out a unique interaction matrix from within the solution space, eliminating competing condensed phases that were not specified in the phase-diagram design problem. Applying this approach to the Flory–Huggins model, Eq. (4), Ref. [60] provided numerical evidence that while the feasibility of the SDP is independent of the degree of polymerization, the convex relaxation becomes a better approximation of the phase-diagram design problem as the degree of polymerization increases (Fig. 6). Interestingly, coexistence regions with more condensed phases than distinct mixture components can be designed in this way. Furthermore, this inverse design approach is easily extended to include additional optimization goals or constraints on the interactions; for example, it is possible to compute the minimum number of matrix-elements that must be changed in order to switch from one phase diagram to another using this method. Ref. [60] also demonstrated that by mapping interaction matrices to molecular pair potentials, interactions designed using mean-field models can be used to establish coexistence among phases with prescribed compositions in molecular simulation models.

In another application of inverse design, Ref. [85] devised an algorithm to engineer pairwise interactions that produce phase-separated condensates with target morphologies, such as those observed in the nucleolus [40]. At equilibrium, surface tensions control the tendency of macroscopic droplets to exist in nonwetting, partial wetting, or complete wetting configurations (separated, fused, and enveloped droplets, respectively, in Fig. 1). Furthermore, within the Cahn–Hilliard framework [76], the surface tensions between phases of mutually immiscible components are directly related to the pairwise interactions. Ref. [85] showed that predicting multiphase morphologies in multicomponent fluids corresponds to a graph decomposition problem, in which vertices indicate phases and edges indicate shared interfaces between phases. Designing interaction matrices for multicomponent mixtures that phase separate into droplets with prescribed (non)wetting architectures can therefore be achieved by encoding the desired morphology in a graph, enumerating affine inequality constraints on the interactions via graph decomposition, and solving the resulting

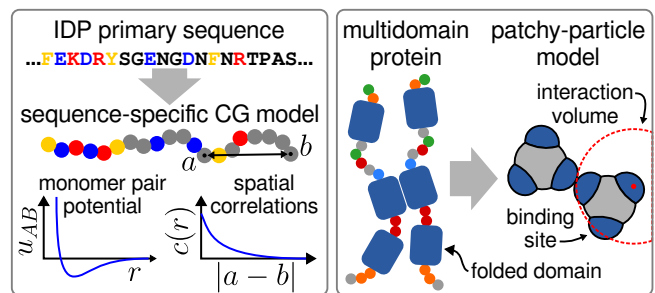


FIG. 7. *Left*: Sequence-dependent CG models represent IDPs as chains of simplified amino acids. Typically, the nonbonded interactions between amino acids of types  $A$  and  $B$  are modeled using a pair potential,  $u_{AB}(r)$ . Analytical theories of sequence-dependent heteropolymer interactions also require a model of the spatial correlations between monomers that are spaced a distance  $|a - b|$  apart in the primary sequence. *Right*: Patchy-particle models of multidomain proteins implement a higher level of coarse-graining by treating the PPI or RNA-binding interfaces on folded domains as specific binding sites on simplified particles; binding sites engage in at most one interaction at a time. Analytical theories associate an interaction volume with each pair of distinct binding-site types.

linear program. Phase-field simulations were then used to demonstrate the efficacy of this design algorithm.

#### IV. SEQUENCE-DEPENDENT THEORIES AND COARSE-GRAINED MOLECULAR MODELS

In parallel with efforts to understand the phase behavior of simplified mixtures with many components, theoretical models have been developed to describe LLPS at a greater level of chemical detail in solutions with a small number of distinct biomolecular species (Fig. 2). In the condensate literature, such models can be broadly classified as sequence-specific coarse-grained (CG) IDP models, which represent nonbonded interactions between amino acids [86–91] or chemical functional groups [92] using pair potentials, and “patchy-particle” [93–95] or “patchy-polymer” [96, 97] CG models, which encode specific interactions between discrete binding sites on each molecule (Fig. 7). We first describe key insights into multicomponent phase behavior from theoretical analyses of these types of models before reviewing recent multicomponent molecular simulation studies.

##### A. Multicomponent field-theoretic approaches

Field-theoretic approaches have been used to predict the sequence-dependent phase diagrams of heteropolymers, with a particular emphasis on polyampholytes. By accounting for chain connectivity, and thus the primary sequence of the heteropolymer, these approaches improve upon mean-field treatments that consider all monomer–

monomer interactions in a polymer solution independently [45]. Field-theoretic approaches incorporate sequence information by modeling the correlations between monomers within a single chain, which decay with increasing separation between monomers along the primary sequence (Fig. 7).

Refs. [98] treated the spatial correlations between monomers with the random-phase approximation (RPA) by assuming that the polymer configurations obey the Gaussian statistics of ideal chains. This assumption means that monomers on different chains are not spatially correlated and that the heteropolymer sequences affect the potential energy, but not the polymer conformations, of a mixture at finite concentration. Despite this simplification, RPA predictions correlate well with experimental measurements of the phase behavior [98] and single-chain properties [99] of charge-neutral polyampholytes. Of particular importance, the RPA theory rationalizes the observed increase in LLPS propensity of charge-neutral sequences with “blocky” as opposed to homogeneous charge patterns [100]. Blocky charge patterns also correlate with smaller radii of gyration of chains in the dilute phase, in line with prior studies using the “sequence charge decoration” order parameter [101, 102] and related blockiness metrics [103, 104]. The RPA theory was extended to charged polyelectrolytes in Ref. [105].

Of relevance to multicomponent mixtures, Ref. [106] applied RPA to mixtures of two distinct charge-neutral polyampholytes. Because RPA ignores spatial correlations between monomers on different chains, the electrostatic contribution to the RPA free energy can be factored into terms arising from each chain individually. The RPA free energy can therefore be mapped at low concentrations to a pairwise-interaction model in which the effective heterotypic interaction,  $B_{12}$ , is the geometric mean of the two homotypic interactions,  $B_{11}$  and  $B_{22}$ . The homotypic interaction coefficients can be calculated by applying RPA to each heteropolymer sequence individually. In light of the discussion in Sec. IIIB, these results indicate that RPA predicts a rank-1 pairwise interaction matrix for charge-neutral polyampholyte mixtures, since  $B_{ij} = B_{ii}^{1/2} B_{jj}^{1/2} \forall i, j$ . This observation further suggests that spatial correlations between different chains are needed to predict higher-rank interaction matrices for charge-neutral polyampholytes.

RPA has also been applied to mixtures of polyelectrolyte mixtures. In solutions with two positively charged and one negatively charged polymer, Ref. [107] predicted that multiphase coacervates can form due to the repulsive heterotypic interactions between two positively charged sequences with differing linear charge densities. Ref. [108] then predicted that differences in the charge patterning between two positively charged sequences with identical linear charge densities is sufficient to drive the formation of two immiscible condensed phases.

An analogous field-theoretic treatment of heteropolymers interacting via short-ranged hydrophobic forces revealed that the leading order contribution to the interac-

tion free energy is given by the sum of the interactions between all pairs of monomers in the mixture [109]. Although this model was not explicitly applied to multicomponent solutions, it suggests that, to leading order, the pairwise interaction matrix for heteropolymers interacting via short-ranged interactions is independent of their primary sequences. In other words, only the frequency of each monomer type in a heteropolymer sequence is relevant at this level of theory [110], and the rank of the interaction matrix cannot exceed the rank of the monomer–monomer interaction matrix, which may itself be rank-deficient [109, 111].

## B. Multicomponent associating fluid models

Concepts from associating fluid theory [112, 113] have been adopted to describe the interactions between binding sites on biomolecules that can only engage in one physical bond at a time. While the methods of Refs. [112] and [113] were originally developed to describe site-specific associative interactions between small molecules, this physical picture extends naturally to multidomain proteins or protein complexes whose constituent domains contain interfaces that interact specifically with other proteins or RNA sequences [93, 114]. The number of such binding sites therefore establishes the coarse-grained “valence” of the multidomain protein or complex (Fig. 7).

Associating fluid theory treats the attractive interactions between pairs of binding sites as perturbations to the free energy of a reference model, which represents the molecular mixture in the absence of binding sites. For example, the Flory–Huggins homopolymer model can serve as a reference model for a mixture of multidomain proteins, with the degree of polymerization  $L_i$  taken to be equal to the number of domains in each protein species  $i$  [93]. The concentration-dependent site–site binding probabilities are then determined from the chemical equilibrium equations

$$X_{iA} + X_{iA} \sum_{j=1}^N \rho_j \sum_{B=1}^{m_j} X_{jB} \Delta_{iA,jB} = 1 \quad \forall i, A, \quad (8)$$

where  $X_{iA}$  represents the probability that the binding site of type  $A$  on a molecule of type  $i$  is not engaged in any associative interaction, and  $m_i$  is the valence of molecule type  $i$ . The matrix  $\{\Delta_{iA,jB}\}$  represents the interaction volumes (i.e., the reciprocals of the dissociation constants) for the associative interactions between binding sites  $A$  and  $B$ , which can in principle depend on spatial correlations in the reference model. Finally, the contribution to the free-energy density due to associative interactions is [113]

$$\beta f_{\text{assoc}} = \sum_{i=1}^N \left[ \rho_i \sum_{A=1}^{m_i} \left( \log X_{iA} - \frac{X_{iA}}{2} \right) + \frac{m_i}{2} \right]. \quad (9)$$

Ref. [115] showed that Eq. (8) has a unique solution and that Eq. (9) leads to a particularly simple

expression for the associative contribution to the excess chemical potential when  $\Delta_{iA,jB}$  is concentration-independent,  $\beta\mu_{\text{assoc},i} = \sum_{A=1}^{m_i} \log X_{iA}$ . Furthermore, in the limit of weak associative interactions, Eq. (9) reduces to a simple pairwise form, such that  $\beta\mu_{\text{assoc},i} \simeq -\sum_{j=1}^N \rho_j \sum_{A=1}^{m_i} \sum_{B=1}^{m_j} \Delta_{iA,jB}$ . With regard to the discussion in Sec. III B, the maximum rank of the pairwise interaction matrix is therefore given by the rank of  $\{\Delta_{iA,jB}\}$  in this limit.

The associating fluid framework has been widely applied to model biomolecular LLPS involving folded domains that interact via specific binding sites. A notable application of the associating fluid framework to multiphase condensates was provided in Ref. [29], which used a simplified representation of an experimentally determined PPI network to predict the compositions and morphologies of coexisting stress granule and P-body condensates. Agreement between theory and experiment regarding the effects of concentration changes and binding-site modifications provided strong evidence that the phase behavior of these condensates is indeed governed by specific PPIs and interactions between RBDs and mRNA.

When the binding sites are assumed to represent individual amino acids of IDPs or short sequence motifs of IDPs and/or RNAs, associating fluid theory is commonly referred to as the “stickers-and-spacers” model of heteropolymer association [4, 116–118]. In this case, a Flory–Huggins homopolymer model with a degree of polymerization much greater than the binding-site valence (i.e., the number of “stickers”) is typically taken as the reference model. Stickers-and-spacers applications of associating fluid theory have been successfully used to rationalize experimental observations of IDP-driven phase separation, including both thermodynamic and dynamical properties, in many contexts [4, 119, 120]. The assignment of the “stickers” to specific amino acids or short sequence motifs has varied depending on context across different studies, however, suggesting that additional contextual information may be required to predict the phase behavior of multicomponent IDP and RNA mixtures from their sequences. For further discussion of applications of associating fluid theory to biopolymers, we direct the reader to recent reviews on this subject, including Refs. [4] and [47].

### C. Insights from coarse-grained molecular simulations of multiphase condensates

#### 1. Polymer simulations with pair potentials

Molecular simulations have provided insights into the accuracy of analytical theories for describing sequence-dependent multicomponent phase behavior. In order to test the RPA predictions of Ref. [106] (see Sec. IV A), Ref. [121] used a combination of field-theoretic and CG MD simulations to study the phase behavior of polyampholyte mixtures. These simulations demonstrated that

pairs of charge-neutral sequences only exhibit demixing when the chains have sufficiently different (i.e., blocky versus uniform) charge distributions. These results are in line with the predictions of the RPA theory. Nonetheless, the authors found that excluded volume interactions—which are present in the MD simulations but are not included in the RPA calculations—are essential for observing demixing in MD simulations. The qualitative agreement with the theoretical predictions was therefore ascribed to the assumption of incompressibility in the RPA calculations. Nonetheless, this observation points to the need for more accurate theoretical treatments that account for excluded volume and interchain correlations.

Moving to systems with a third non-solvent component, Ref. [43] performed simulations of a three-component system comprising a prion-like polypeptide (PLP), an arginine-rich polypeptide (RRP), and RNA. In this system, competition between PLP and RNA for binding to RRP results in the demixing of PLP+RRP condensates into immiscible PLP and RNA+RRP phases when RNA is added. This experimental observation, which bears qualitative resemblance to the competing heterotypic model of Ref. [107] (see Sec. IV A), was reproduced using MD simulations of a CG IDP/RNA model. These simulations also rationalized the experimental observation that the RNA parent concentration controls the morphology of the coexisting condensates.

In an attempt to uncover general sequence determinants of multiphase mixtures, Ref. [122] proposed a computational approach to design IDP sequences that result in multilayered condensates. To this end, the authors used a genetic algorithm to optimize pairs of sequences that form immiscible phases and a stable shared interface, starting from naturally occurring IDP sequences. The authors found that the net homotypic and heterotypic interactions must differ between the optimized IDPs, as expected. In many cases, these net interactions were found to depend primarily on the monomer frequencies, such that the immiscibility of the two phases was not affected by randomizing the sequences of the designed IDPs. However, when the genetic algorithm was initialized using a particular naturally occurring IDP sequence in one of the coexisting phases, the patterning of the amino-acid residues in the optimized partner sequence was found to be crucial for achieving immiscibility. The reasons for this dependence on sequence patterning in some, but not all, optimization scenarios are poorly understood. Nonetheless, sequences generated via this approach could provide challenging test cases for the further development of analytical sequence-dependent theories.

#### 2. Patchy-polymer simulations

“Patchy-polymer” models, which encode one-to-one interactions between binding sites on specific monomers, are appropriate CG models for testing the predictions of associating fluid theory. To explore the design rules

underlying multiphasic systems with this class of models, Ref. [123] introduced a lattice-based CG model of poly-PRM and poly-SH3 multidomain proteins. Proline-rich modules (PRMs) are short IDR sequence motifs that engage in specific interactions with folded SH3 domains [124], and as such form one-to-one binding interactions. Meanwhile, the linkers between motifs in the poly-PRM molecules and between the folded domains in the poly-SH3 molecules, respectively, were modeled either implicitly, representing ideal chains with Gaussian conformational statistics, or explicitly, using a variable number of lattice-site-occupying monomers. Simulations were conducted using two types of poly-SH3 molecules, which competed for binding to the PRMs. The authors found that differences in the linker properties, which tune the effective pairwise interactions between the molecules in the absence of the associative PRM/SH3 interactions, strongly affect the ability of the mixture to form immiscible condensed phases. This observation is consistent with the finding of Ref. [121] that excluded volume interactions are necessary for demixing. By contrast, the interaction volume associated with the attractive PRM/SH3 interactions was found to play a less important role in determining the degree of immiscibility, in line with the predictions of associating fluid theory in the strong-binding limit ( $\rho\Delta \gg 1$ ) of Eq. (8). Ref. [123] also showed that the interfaces of immiscible condensates are similarly affected by the linker properties, since molecules containing linkers with greater excluded volumes are preferentially driven towards interfaces with the dilute phase.

### 3. Patchy-particle models

“Patchy-particle” models allow for simulations with a larger number of distinct molecular species, along with a greater diversity of associative interactions, due to their simplicity. In complex mixtures with a variety of different associative interactions, it is useful to describe the collection of all possible one-to-one binding interactions by introducing an “interaction network” [29]. Ref. [125] explored this network concept using MD simulations of a 6-component mixture comprising 2, 3, and 4-valent patchy particles. The authors considered a nearly fully connected network with almost all equivalent interaction strengths, leading to the formation of a single condensed phase in mixtures with equimolar parent concentrations. Unsurprisingly, the density of associative bonding interactions in the condensed phase was found to correlate with the condensate stability, as measured by the critical temperature. Simulations further revealed that high-valence molecules, which phase separate with high critical temperatures in single-component solutions, tend to increase the critical temperature of multicomponent condensates when added to mixtures of components with lower average valence. Of direct experimental relevance, positive correlations were observed between the critical point of a molecular species in a single-component solu-

tion, its binding-site valence, and its partition coefficient with respect to a multicomponent condensate in a mixture with equimolar parent concentrations.

These observations can be understood qualitatively within the framework of associating fluid theory. Making the simplification that all binding sites interact with one another via the same interaction volume, such that  $\Delta_{iA,jB} = \bar{\Delta} \forall i, A, j, B$ , the solution to Eq. (8) simplifies to  $X_{iA} = \bar{X} \forall i, A$ . The associative contribution to the excess chemical potential (see Sec. IV B) is thus  $\beta\mu_{\text{assoc},i} \approx m_i \log \bar{X}$  in the condensed phase and negligible in the dilute phase, implying that the partition coefficient, Eq. (2), is related to the binding-site valence by  $\text{PC}_i \propto \exp(m_i)$ . An approximate relationship between the stability of the condensed phase and the average valence of the mixture follows by a similar argument. The relatively small variations in interaction strengths in the simulated interaction network [125] can be considered as perturbations on these predictions. However, variations in the geometric arrangements of the binding sites, and their relatively minor effects on the partition coefficients, are not captured at this level of theory.

The patchy-particle model of Ref. [125] was then extended to examine multiphase mixtures in Ref. [126]. The authors modified the interaction network by eliminating heterotypic associative interactions between select molecular species in order to construct immiscible condensates and multilayered structures. Analogously to the results of Ref. [123], simulations demonstrated that strong homotypic associative interactions lead to the formation of multiple immiscible condensates, while the introduction of strong heterotypic associative interactions tends to stabilize a single condensed phase. However, mixtures with competing heterotypic interactions between weakly and strongly associating species showed evidence of multiphase condensate formation.

## V. OUTLOOK AND CHALLENGES

We have reviewed recent progress in the development of statistical and sequence-specific theories of multicomponent fluids and multiphase condensate formation. Further advances in this area have the potential to reveal quantitative relationships between the molecular determinants of biomolecules, whether naturally occurring or engineered, and phase-separated self-organization in heterogeneous mixtures. In particular, inverse-design strategies offer a promising approach for rationally and systematically identifying the physicochemical properties of biomolecular mixtures responsible for the assembly of complex—and biologically functional—condensates. These theoretical and computational efforts will help to provide a roadmap for future experiments on heterogeneous biomolecular mixtures.

Nonetheless, many significant theoretical challenges remain to be explored, particularly with regard to the as-

sumption of thermal equilibrium. Future directions for theoretical and simulation advances in this field include:

1. *Structuring and parameterizing multicomponent mixture models.* Further development of statistical mixture models (see Sec. III) will require incorporating information from sequence-dependent theories and simulations. In this way, it will be possible to investigate the thermodynamic consequences of physically motivated and biomolecularly relevant correlations among interaction parameters in multicomponent fluids, as well as to move beyond pairwise mixture models. Physicochemically motivated constraints should also be incorporated into inverse design approaches.
2. *Extending sequence-dependent coarse-grained simulations and theories to multicomponent mixtures.* Complementary insights can be gained by increasing the number of components in biomolecular mixtures treated using analytical theories or studied via coarse-grained molecular simulation (see Sec. IV). Simulations of recently developed CG IDP models [86–92] have demonstrated impressive agreement with experiments on both single-chain and individual condensed-phase properties, suggesting that multicomponent simulations using these models may also be capable of predicting multiphase coexistence [122] with similar accuracy. Further improvement in the chemical accuracy of multicomponent simulations is likely to be achieved through multiscale approaches that incorporate all-atom simulations of ribonucleic condensates [127–129]. In future simulation studies, it will also be important to consider the role of competition between sequence-dependent clustering, aggregation, and LLPS behaviors, as observed in simple models of single-component heteropolymer solutions [130–132], in multicomponent mixtures.
3. *Accounting for nonequilibrium effects due to kinetic barriers.* Within the near-equilibrium framework, kinetic effects can lead to differences between the phase behavior that is observed in simulations and experiments and what is predicted at global thermodynamic equilibrium. For example, nucleation pathways [133] and slow rates of transitions be-

tween metastable states [134–136] can affect the molecular compositions and multiphasic organization of phase-separated condensates on biologically relevant timescales. The consequences of these nonequilibrium effects in systems with many components require further exploration.

4. *Exploring differences in phase behavior at nonequilibrium steady states.* Phase separation can also occur in fluids at nonequilibrium steady states (NESSs), which can arise due to chemostatted chemical reactions [18]. Differences between thermal equilibrium and a NESS can manifest, for example, in the nucleation behavior [137, 138] as well as the growth and coarsening dynamics [139–143] of phase-separated droplets. The implications of chemically driven NESSs for multiphase self-organization are largely unexplored.
5. *Developing theoretical tools for emerging experimental applications.* A variety of experimental platforms for manipulating biomolecular LLPS have recently been developed using “designer” peptides [144–147], nucleic acids [32, 33, 148], and non-biological polymers [149]. Chemically specific computational tools are needed to guide the rational design of multicomponent, multiphasic mixtures using these experimental platforms. With a better understanding of condensate compositional control in heterogeneous environments, combined theoretical and experimental engineering approaches have the potential to bring about practical techniques for manipulating complex biological processes *in vivo* [150].

In summary, LLPS can give rise to highly nontrivial spatial organization in multicomponent biomolecular fluids. Nevertheless, considerable gaps persist in our understanding of the relationship between molecular-level properties and emergent phase behavior in heterogeneous mixtures. Addressing this multifaceted question therefore represents an important way in which chemical theory and simulation can contribute to research at the forefront of molecular and cell biology, while helping to elucidate the origins of self-organization in living systems.

This work is supported by the National Science Foundation (DMR-2143670).

- 
- [1] Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoesge, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **2009**, *324*, 1729–1732.
  - [2] Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, eaaf4382.
  - [3] Berry, J.; Brangwynne, C. P.; Haataja, M. Physical principles of intracellular organization via active and passive phase transitions. *Rep. Prog. Phys.* **2018**, *81*, 046601.
  - [4] Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical principles underlying the complex biology of intracellular phase transitions. *Ann. Rev. Biophys.* **2020**, *49*, 107–133.
  - [5] Villegas, J. A.; Heidenreich, M.; Levy, E. D. Molecular and environmental determinants of biomolecular con-

- densate formation. *Nat. Chem. Biol.* **2022**, 1–11.
- [6] Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Bio.* **2017**, *18*, 285–298.
  - [7] Boija, A.; Klein, I. A.; Sabari, B. R.; Dall’Agnese, A.; Coffey, E. L.; Zamudio, A. V.; Li, C. H.; Shrinivas, K.; Manteiga, J. C.; Hannett, N. M., et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **2018**, *175*, 1842–1855.
  - [8] Sabari, B. R.; Dall’Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C., et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **2018**, *361*, eaar3958.
  - [9] Wang, A.; Conicella, A. E.; Schmidt, H. B.; Martin, E. W.; Rhoads, S. N.; Reeb, A. N.; Nourse, A.; Ramirez Montero, D.; Ryan, V. H.; Rohatgi, R., et al. A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *EMBO J.* **2018**, *37*, e97452.
  - [10] Kim, T. H.; Tsang, B.; Vernon, R. M.; Sonenberg, N.; Kay, L. E.; Forman-Kay, J. D. Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates regulation of translation and deadenylation. *Science* **2019**, *365*, 825–829.
  - [11] Lafontaine, D. L.; Riback, J. A.; Bascetin, R.; Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 165–182.
  - [12] Klosin, A.; Oltsch, F.; Harmon, T.; Honigmann, A.; Jülicher, F.; Hyman, A. A.; Zechner, C. Phase separation provides a mechanism to reduce noise in cells. *Science* **2020**, *367*, 464–468.
  - [13] Su, X.; Ditlev, J. A.; Hui, E.; Xing, W.; Banjade, S.; Okrut, J.; King, D. S.; Taunton, J.; Rosen, M. K.; Vale, R. D. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **2016**, *352*, 595–599.
  - [14] Tiwary, A. K.; Zheng, Y. Protein phase separation in mitosis. *Curr. Opin. Cell Biol.* **2019**, *60*, 92–98.
  - [15] Alberti, S.; Hyman, A. A. Biomolecular condensates at the nexus of cellular stress, protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 196–213.
  - [16] Jack, A.; Ferro, L. S.; Trnka, M. J.; Wehri, E.; Nadgir, A.; Nguyenla, X.; Fox, D.; Costa, K.; Stanley, S.; Schaletzky, J., et al. SARS-CoV-2 nucleocapsid protein forms condensates with viral genomic RNA. *PLoS Biol.* **2021**, *19*, e3001425.
  - [17] Wei, M.-T.; Chang, Y.-C.; Shimobayashi, S. F.; Shin, Y.; Strom, A. R.; Brangwynne, C. P. Nucleated transcriptional condensates amplify gene expression. *Nature cell biology* **2020**, *22*, 1187–1196.
  - [18] Söding, J.; Zwicker, D.; Sohrabi-Jahromi, S.; Boehning, M.; Kirschbaum, J. Mechanisms for active regulation of biomolecular condensates. *Trends Cell Biol.* **2020**, *30*, 4–14.
  - [19] Colby, R. H.; Rubinstein, M. *Polymer Physics*; Oxford University Press, 2003.
  - [20] McMaster, L. P. *Aspects of liquid–liquid phase transition phenomena in multicomponent polymeric systems*; ACS Publications, 1975.
  - [21] Taratuta, V. G.; Holschbach, A.; Thurston, G. M.; Blankschtein, D.; Benedek, G. B. Liquid-liquid phase separation of aqueous lysozyme solutions: Effects of pH and salt identity. *J. Phys. Chem.* **1990**, *94*, 2140–2144.
  - [22] Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid–liquid phase separation in biology. *Ann. Rev. Cell Dev. Biol.* **2014**, *30*, 39–58.
  - [23] Jawerth, L.; Fischer-Friedrich, E.; Saha, S.; Wang, J.; Franzmann, T.; Zhang, X.; Sachweh, J.; Ruer, M.; Ijavi, M.; Saha, S., et al. Protein condensates as aging Maxwell fluids. *Science* **2020**, *370*, 1317–1323.
  - [24] Mittag, T.; Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* **2022**,
  - [25] Alshareedah, I.; Singh, A.; Quinn, A.; Banerjee, P. R. Determinants of Viscoelasticity and Flow Activation Energy in Biomolecular Condensates. *bioRxiv preprint* **2022**, 2022–12.
  - [26] Tejedor, A. R.; Sanchez-Burgos, I.; Estevez-Espinosa, M.; Garaizar, A.; Collepardo-Guevara, R.; Ramirez, J.; Espinosa, J. R. Protein structural transitions critically transform the network connectivity and viscoelasticity of RNA-binding protein condensates but RNA can prevent it. *Nat. Comm.* **2022**, *13*, 5717.
  - [27] Alshareedah, I.; Borchers, W. M.; Cohen, S. R.; Farag, M.; Singh, A.; Bremer, A.; Pappu, R. V.; Mittag, T.; Banerjee, P. R. Sequence-encoded grammars determine material properties and physical aging of protein condensates. *bioRxiv preprint* **2023**, 2023–04.
  - [28] Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **2015**, *11*, 899–904.
  - [29] Sanders, D. W.; Kedersha, N.; Lee, D. S.; Strom, A. R.; Drake, V.; Riback, J. A.; Bracha, D.; Eeftens, J. M.; Iwanicki, A.; Wang, A., et al. Competing protein-RNA interaction networks control multiphase intracellular organization. *Cell* **2020**, *181*, 306–324.
  - [30] Lin, Y.; Protter, D. S.; Rosen, M. K.; Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **2015**, *60*, 208–219.
  - [31] King, J. T.; Shakya, A. Phase separation of DNA: From past to present. *Biophys. J.* **2021**, *120*, 1139–1149.
  - [32] Rovigatti, L.; Bomboi, F.; Sciortino, F. Accurate phase diagram of tetravalent DNA nanostars. *J. Chem. Phys.* **2014**, *140*, 154903.
  - [33] Hegde, O.; Li, T.; Sharma, A.; Borja, M.; Jacobs, W. M.; Rogers, W. B. Competition between self-assembly and phase separation induces re-entrant condensation of DNA liquids. *arXiv preprint* **2023**, arXiv:2301.06134.
  - [34] Wadsworth, G. M.; Zahurancik, W. J.; Zeng, X.; Pullara, P.; Lai, L. B.; Sidharthan, V.; Pappu, R. V.; Gopalan, V.; Banerjee, P. R. RNAs undergo phase transitions with lower critical solution temperatures. *bioRxiv preprint* **2022**, 10.1101/2022.10.17.512593.
  - [35] Ditlev, J. A.; Case, L. B.; Rosen, M. K. Who’s in and who’s out—compositional control of biomolecular condensates. *J. Mol. Biol.* **2018**, *430*, 4666–4684.
  - [36] Xing, W.; Muhlrad, D.; Parker, R.; Rosen, M. K. A quantitative inventory of yeast P body proteins reveals principles of composition and specificity. *Elife* **2020**, *9*, e56525.
  - [37] Riback, J. A.; Zhu, L.; Ferrolino, M. C.; Tolbert, M.; Mitrea, D. M.; Sanders, D. W.; Wei, M.-T.; Kriwacki, R. W.; Brangwynne, C. P. Composition-



- dependent thermodynamics of intracellular phase separation. *Nature* **2020**, *581*, 209–214.
- [38] Qian, D.; Welsh, T. J.; Erkamp, N. A.; Qamar, S.; Nixon-Abell, J.; Krainer, G.; George-Hyslop, P. S.; Michaels, T. C.; Knowles, T. P. Tie-Line Analysis Reveals Interactions Driving Heteromolecular Condensate Formation. *Phys. Rev. X* **2022**, *12*, 041038.
- [39] Fare, C. M.; Villani, A.; Drake, L. E.; Shorter, J. Higher-order organization of biomolecular condensates. *Open Biol.* **2021**, *11*, 210137.
- [40] Feric, M.; Vaidya, N.; Harmon, T. S.; Mitrea, D. M.; Zhu, L.; Richardson, T. M.; Kriwacki, R. W.; Pappu, R. V.; Brangwynne, C. P. Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **2016**, *165*, 1686–1697.
- [41] Guillén-Boixet, J.; Kopach, A.; Holehouse, A. S.; Wittmann, S.; Jahnel, M.; Schlüßler, R.; Kim, K.; Trussina, I. R.; Wang, J.; Mateju, D., et al. RNA-induced conformational switching and clustering of G3BP drive stress granule assembly by condensation. *Cell* **2020**, *181*, 346–361.
- [42] Yang, P.; Mathieu, C.; Kolaitis, R.-M.; Zhang, P.; Messing, J.; Yurtsever, U.; Yang, Z.; Wu, J.; Li, Y.; Pan, Q., et al. G3BP1 is a tunable switch that triggers phase separation to assemble stress granules. *Cell* **2020**, *181*, 325–345.
- [43] Kaur, T.; Raju, M.; Alshareedah, I.; Davis, R. B.; Potoyan, D. A.; Banerjee, P. R. Sequence-encoded and composition-dependent protein-RNA interactions control multiphasic condensate morphologies. *Nat. Comm.* **2021**, *12*, 872.
- [44] Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Ann. Rev. Phys. Chem.* **2020**, *71*, 53.
- [45] Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Theories for sequence-dependent phase behaviors of biomolecular condensates. *Biochemistry* **2018**, *57*, 2499–2508.
- [46] Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 215–235.
- [47] Pappu, R. V.; Cohen, S. R.; Dar, F.; Farag, M.; Kar, M. Phase Transitions of Associative Biomacromolecules. *Chem. Rev.* **2023**.
- [48] Abyzov, A.; Blackledge, M.; Zweckstetter, M. Conformational dynamics of intrinsically disordered proteins regulate biomolecular condensate chemistry. *Chem. Rev.* **2022**, *122*, 6719–6748.
- [49] Griffiths, R. B.; Wheeler, J. C. Critical points in multi-component systems. *Phys. Rev. A* **1970**, *2*, 1047.
- [50] Chandler, D. *Introduction to modern statistical mechanics*; Oxford University Press, 1987.
- [51] Porter, D. A.; Easterling, K. E. *Phase transformations in metals and alloys*; CRC press, 2009.
- [52] Hansen, J.-P.; McDonald, I. R. *Theory of simple liquids: With applications to soft matter*; Academic Press, 2013.
- [53] Qian, D.; Michaels, T. C.; Knowles, T. P. Analytical Solution to the Flory–Huggins Model. *J. Phys. Chem. Lett.* **2022**, *13*, 7853–7860.
- [54] Jacobs, W. M.; Frenkel, D. Phase transitions in biological systems with many components. *Biophys. J.* **2017**, *112*, 683–691.
- [55] Shek, A. C.; Kusumaatmaja, H. Spontaneous phase separation of ternary fluid mixtures. *Soft Matt.* **2022**, *18*, 5807–5814.
- [56] Mazararakos, K.; Zhou, H.-X. Multiphase organization is a second phase transition within multi-component biomolecular condensates. *J. Chem. Phys.* **2022**, *156*, 191104.
- [57] Deviri, D.; Safran, S. A. Physical theory of biological noise buffering by multicomponent phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2100099118.
- [58] Mao, S.; Kuldinow, D.; Haataja, M. P.; Košmrlj, A. Phase behavior and morphology of multicomponent liquid mixtures. *Soft Matter* **2019**, *15*, 1297–1311.
- [59] Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)* **1996**, *22*, 469–483.
- [60] Chen, F.; Jacobs, W. M. Programmable phase behavior in fluids with designable interactions. *arXiv preprint* **2023**, arXiv:2301.12958.
- [61] Lin, Y.-H.; Wessén, J.; Pal, T.; Das, S.; Chan, H. S. *Phase-Separated Biomolecular Condensates: Methods and Protocols*; Springer, 2022; pp 51–94.
- [62] Zwicker, D.; Laan, L. Evolved interactions stabilize many coexisting phases in multicomponent liquids. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2201250119.
- [63] Frenkel, D.; Smit, B. *Understanding molecular simulation: From algorithms to applications*; Elsevier, 2001.
- [64] Wilding, N. B. Critical-point and coexistence-curve properties of the Lennard-Jones fluid: A finite-size scaling study. *Phys. Rev. E* **1995**, *52*, 602–611.
- [65] Berg, B. A.; Neuhaus, T. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.* **1992**, *68*, 9.
- [66] Buzzacchi, M.; Sollich, P.; Wilding, N. B.; Müller, M. Simulation estimates of cloud points of polydisperse fluids. *Phys. Rev. E* **2006**, *73*, 046110.
- [67] Jacobs, W. M.; Frenkel, D. Predicting phase behavior in multicomponent mixtures. *J. Chem. Phys.* **2013**, *139*, 024108.
- [68] Wang, F.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, *86*, 2050.
- [69] Jacobs, W. M. Self-assembly of biomolecular condensates with shared components. *Phys. Rev. Lett.* **2021**, *126*, 258101.
- [70] Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- [71] Sear, R. P.; Cuesta, J. A. Instabilities in complex mixtures with a large number of components. *Phys. Rev. Lett.* **2003**, *91*, 245701.
- [72] Thewes, F. C.; Krüger, M.; Sollich, P. Composition Dependent Instabilities in Mixtures With Many Components. *arXiv preprint* **2022**, arXiv:2211.04170.
- [73] Girard, M. On kinetics and extreme values in systems with random interactions. *Phys. Biol.* **2022**, *20*, 016006.
- [74] Shrinivas, K.; Brenner, M. P. Phase separation in fluids with many interacting components. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2108551118.
- [75] Onsager, L. Reciprocal relations in irreversible processes. I. *Phys. Rev.* **1931**, *37*, 405.
- [76] Cahn, J. W.; Hilliard, J. E. Free energy of a nonuniform system. I. Interfacial free energy. *J. Chem. Phys.* **1958**, *28*, 258–267.
- [77] Provatas, N.; Elder, K. *Phase-field methods in materials*

- science and engineering*; John Wiley & Sons, 2011.
- [78] Carugno, G.; Neri, I.; Vivo, P. Instabilities of complex fluids with partially structured and partially random interactions. *Phys. Biol.* **2022**, *19*, 056001.
- [79] Graf, I. R.; Machta, B. B. Thermodynamic stability and critical points in multicomponent mixtures with structured interactions. *Phys. Rev. Research* **2022**, *4*, 033144.
- [80] Sollich, P. Predicting phase equilibria in polydisperse systems. *J. Phys. Condens. Mat.* **2001**, *14*, R79.
- [81] Sollich, P.; Warren, P. B.; Cates, M. E. Moment free energies for polydisperse systems. *Adv. Chem. Phys.* **2001**, *116*, 265–336.
- [82] Boyd, S.; Boyd, S. P.; Vandenberghe, L. *Convex optimization*; Cambridge University Press, 2004.
- [83] Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 2554–2558.
- [84] Murugan, A.; Zeravcic, Z.; Brenner, M. P.; Leibler, S. Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 54–59.
- [85] Mao, S.; Chakraverti-Wuerthwein, M. S.; Gaudio, H.; Košmrlj, A. Designing the morphology of separated phases in multicomponent liquid mixtures. *Phys. Rev. Lett.* **2020**, *125*, 218003.
- [86] Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comp. Biol.* **2018**, *14*, e1005941.
- [87] Dignon, G. L.; Zheng, W.; Best, R. B.; Kim, Y. C.; Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 9929–9934.
- [88] Latham, A. P.; Zhang, B. Maximum entropy optimized force field for intrinsically disordered proteins. *J. Chem. Theory Comput.* **2019**, *16*, 773–781.
- [89] Joseph, J. A.; Reinhardt, A.; Aguirre, A.; Chew, P. Y.; Russell, K. O.; Espinosa, J. R.; Garaizar, A.; Collepardo-Guevara, R. Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy. *Nat. Comput. Sci.* **2021**, *1*, 732–743.
- [90] Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, e2111696118.
- [91] Tesei, G.; Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open. Res. Europe* **2022**, *2*, 94.
- [92] Benayad, Z.; von Bülow, S.; Stelzl, L. S.; Hummer, G. Simulation of FUS protein condensates with an adapted coarse-grained model. *J. Chem. Theory Comput.* **2020**, *17*, 525–537.
- [93] Jacobs, W. M.; Oxtoby, D. W.; Frenkel, D. Phase separation in solutions with specific and nonspecific interactions. *The Journal of Chemical Physics* **2014**, *140*, 05B626.1.
- [94] Espinosa, J. R.; Garaizar, A.; Vega, C.; Frenkel, D.; Collepardo-Guevara, R. Breakdown of the law of rectilinear diameter and related surprises in the liquid-vapor coexistence in systems of patchy particles. *J. Chem. Phys.* **2019**, *150*, 224510.
- [95] Joseph, J. A.; Espinosa, J. R.; Sanchez-Burgos, I.; Garaizar, A.; Frenkel, D.; Collepardo-Guevara, R. Thermodynamics and kinetics of phase separation of protein-RNA mixtures by a minimal model. *Biophys. J.* **2021**, *120*, 1219–1230.
- [96] Zhang, Y.; Xu, B.; Weiner, B. G.; Meir, Y.; Wingreen, N. S. Decoding the physical principles of two-component biomolecular phase separation. *Elife* **2021**, *10*, e62403.
- [97] Murata, Y.; Niina, T.; Takada, S. The stoichiometric interaction model for mesoscopic MD simulations of liquid-liquid phase separation. *Biophys. J.* **2022**, *121*, 4382–4393.
- [98] Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Sequence-specific polyampholyte phase separation in membraneless organelles. *Phys. Rev. Lett.* **2016**, *117*, 178101.
- [99] Lin, Y.-H.; Chan, H. S. Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophys. J.* **2017**, *112*, 2043–2046.
- [100] McCarty, J.; Delaney, K. T.; Danielsen, S. P.; Fredrickson, G. H.; Shea, J.-E. Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins. *J. Phys. Chem. Lett.* **2019**, *10*, 1644–1652.
- [101] Sawle, L.; Ghosh, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* **2015**, *143*, 08B615.1.
- [102] Firman, T.; Ghosh, K. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *J. Chem. Phys.* **2018**, *148*, 123305.
- [103] Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 8183–8188.
- [104] Das, R. K.; Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 13392–13397.
- [105] Lin, Y.-H.; Brady, J. P.; Chan, H. S.; Ghosh, K. A unified analytical theory of heteropolymers for sequence-specific phase behaviors of polyelectrolytes and polyampholytes. *J. Chem. Phys.* **2020**, *152*, 045102.
- [106] Lin, Y.-H.; Brady, J. P.; Forman-Kay, J. D.; Chan, H. S. Charge pattern matching as a ‘fuzzy’ mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J. Phys.* **2017**, *19*, 115003.
- [107] Chen, X.; Chen, E.-Q.; Shi, A.-C.; Yang, S. Multiphase coacervates driven by electrostatic correlations. *ACS Macro Lett.* **2021**, *10*, 1041–1047.
- [108] Chen, X.; Chen, E.-Q.; Yang, S. Multiphase Coacervation of Polyelectrolytes Driven by Asymmetry of Charged Sequence. *Macromolecules* **2023**, *56*, 3–14.
- [109] Wessén, J.; Das, S.; Pal, T.; Chan, H. S. Analytical Formulation and Field-Theoretic Simulation of Sequence-Specific Phase Separation of Protein-Like Heteropolymers with Short-and Long-Spatial-Range Interactions. *J. Phys. Chem. B* **2022**, *126*, 9222–9245.
- [110] De Sancho, D. Phase separation in amino acid mixtures is governed by composition. *Biophys. J.* **2022**, *121*, 4119–4127.
- [111] Li, H.; Tang, C.; Wingreen, N. S. Nature of driving force for protein folding: A result from analyzing the

- statistical potential. *Phys. Rev. Lett.* **1997**, *79*, 765.
- [112] Wertheim, M. S. Fluids with highly directional attractive forces. I. Statistical thermodynamics. *J. Stat. Phys.* **1984**, *35*, 19–34.
- [113] Chapman, W. G.; Gubbins, K. E.; Jackson, G.; Radosz, M. SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilib.* **1989**, *52*, 31–38.
- [114] Jacobs, W. M.; Knowles, T. P.; Frenkel, D. Oligomers of heat-shock proteins: Structures that don't imply function. *PLoS Comp. Biol.* **2016**, *12*, e1004756.
- [115] Michelsen, M. L.; Hendriks, E. M. Physical properties from association models. *Fluid Phase Equilib.* **2001**, *180*, 165–174.
- [116] Semenov, A. N.; Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. *Macromolecules* **1998**, *31*, 1373–1385.
- [117] Tanaka, F. Theoretical study of molecular association and thermoreversible gelation in polymers. *Polym. J.* **2002**, *34*, 479–509.
- [118] Michels, J. J.; Brzezinski, M.; Scheidt, T.; Lemke, E. A.; Parekh, S. H. Role of solvent compatibility in the phase behavior of binary solutions of weakly associating multivalent polymers. *Biomacromolecules* **2021**, *23*, 349–364.
- [119] Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D., et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **2018**, *174*, 688–699.
- [120] Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694–699.
- [121] Pal, T.; Wessén, J.; Das, S.; Chan, H. S. Subcompartmentalization of polyampholyte species in organelle-like condensates is promoted by charge-pattern mismatch and strong excluded-volume interaction. *Phys. Rev. E* **2021**, *103*, 042406.
- [122] Chew, P. Y.; Joseph, J. A.; Collepardo-Guevara, R.; Reinhardt, A. Thermodynamic origins of two-component multiphase condensates of proteins. *Chem. Sci.* **2023**, *14*, 1820–1836.
- [123] Harmon, T. S.; Holehouse, A. S.; Pappu, R. V. Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins. *New J. Phys.* **2018**, *20*, 045002.
- [124] Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F., et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483*, 336–340.
- [125] Espinosa, J. R.; Joseph, J. A.; Sanchez-Burgos, I.; Garaizar, A.; Frenkel, D.; Collepardo-Guevara, R. Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 13238–13247.
- [126] Sanchez-Burgos, I.; Espinosa, J. R.; Joseph, J. A.; Collepardo-Guevara, R. Valency and binding affinity variations can regulate the multilayered organization of protein condensates with many components. *Biomolecules* **2021**, *11*, 278.
- [127] Zheng, W.; Dignon, G. L.; Jovic, N.; Xu, X.; Regy, R. M.; Fawzi, N. L.; Kim, Y. C.; Best, R. B.; Mittal, J. Molecular details of protein condensates probed by microsecond long atomistic simulations. *J. Phys. Chem. B* **2020**, *124*, 11671–11679.
- [128] Gruijs da Silva, L. A.; Simonetti, F.; Hutten, S.; Riemenschneider, H.; Sternburg, E. L.; Pietrek, L. M.; Gebel, J.; Dötsch, V.; Edbauer, D.; Hummer, G., et al. Disease-linked TDP-43 hyperphosphorylation suppresses TDP-43 condensation and aggregation. *EMBO J.* **2022**, *41*, e108443.
- [129] Galvanetto, N.; Ivanović, M. T.; Chowdhury, A.; Sottini, A.; Nüesch, M.; Nettels, D.; Best, R.; Schuler, B. Ultrafast molecular dynamics observed within a dense protein condensate. *bioRxiv preprint* **2022**, 2022–12.
- [130] Statt, A.; Casademunt, H.; Brangwynne, C. P.; Panagiotopoulos, A. Z. Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *J. Chem. Phys.* **2020**, *152*, 075101.
- [131] Rana, U.; Brangwynne, C. P.; Panagiotopoulos, A. Z. Phase separation vs aggregation behavior for model disordered proteins. *J. Chem. Phys.* **2021**, *155*, 125101.
- [132] Rekhi, S.; Sundaravadivelu Devarajan, D.; Howard, M. P.; Kim, Y. C.; Nikoubashman, A.; Mittal, J. Role of Strong Localized vs. Weak Distributed Interactions in Disordered Protein Phase Separation. *bioRxiv preprint* **2023**, 10.1101/2023.01.27.525976.
- [133] Shimobayashi, S. F.; Ronceray, P.; Sanders, D. W.; Haataja, M. P.; Brangwynne, C. P. Nucleation landscape of biomolecular condensates. *Nature* **2021**, *599*, 503–506.
- [134] Garaizar, A.; Espinosa, J. R.; Joseph, J. A.; Krainer, G.; Shen, Y.; Knowles, T. P.; Collepardo-Guevara, R. Aging can transform single-component protein condensates into multiphase architectures. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, e2119800119.
- [135] Erkamp, N. A.; Sneideris, T.; Ausserwöger, H.; Qian, D.; Qamar, S.; Nixon-Abell, J.; George-Hyslop, P. S.; Schmit, J. D.; Weitz, D. A.; Knowles, T. P. Multiphase condensates from a kinetically arrested phase transition. *bioRxiv preprint* **2022**, 10.1101/2022.09.19.508534.
- [136] Lin, A. Z.; Ruff, K. M.; Jalihal, A. P.; Dar, F.; King, M. R.; Lalmansingh, J. M.; Posey, A. E.; Seim, I.; Gladfelter, A. S.; Pappu, R. V. Dynamical control enables the formation of demixed biomolecular condensates. *bioRxiv preprint* **2023**, 10.1101/2023.01.04.522702.
- [137] Cho, Y.; Jacobs, W. M. Tuning nucleation kinetics via nonequilibrium chemical reactions. *Phys. Rev. Lett.* **2023**, *130*, 128203.
- [138] Ziethen, N.; Kirschbaum, J.; Zwicker, D. Nucleation of chemically active droplets. *arXiv preprint* **2022**, arXiv:2212.12224.
- [139] Zwicker, D.; Hyman, A. A.; Jülicher, F. Suppression of Ostwald ripening in active emulsions. *Phys. Rev. E* **2015**, *92*, 012317.
- [140] Zwicker, D.; Seyboldt, R.; Weber, C. A.; Hyman, A. A.; Jülicher, F. Growth and division of active droplets provides a model for protocells. *Nat. Phys.* **2017**, *13*, 408–413.
- [141] Wurtz, J. D.; Lee, C. F. Chemical-reaction-controlled phase separated drops: Formation, size selection, and

- coarsening. *Phys. Rev. Lett.* **2018**, *120*, 078102.
- [142] Weber, C. A.; Zwicker, D.; Jülicher, F.; Lee, C. F. Physics of active emulsions. *Rep. Prog. Phys.* **2019**, *82*, 064601.
- [143] Kirschbaum, J.; Zwicker, D. Controlling biomolecular condensates via chemical reactions. *J. R. Soc. Interface* **2021**, *18*, 20210255.
- [144] Simon, J. R.; Carroll, N. J.; Rubinstein, M.; Chilkoti, A.; López, G. P. Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nat. Chem.* **2017**, *9*, 509–515.
- [145] Fisher, R. S.; Elbaum-Garfinkle, S. Tunable multiphase dynamics of arginine and lysine liquid condensates. *Nat. Comm.* **2020**, *11*, 4628.
- [146] Heidenreich, M.; Georgeson, J. M.; Locatelli, E.; Rovigatti, L.; Nandi, S. K.; Steinberg, A.; Nadav, Y.; Shimon, E.; Safran, S. A.; Doye, J. P., et al. Designer protein assemblies with tunable phase diagrams in living cells. *Nat. Chem. Biol.* **2020**, *16*, 939–945.
- [147] Baruch Leshem, A.; Sloan-Dennison, S.; Massarano, T.; Ben-David, S.; Graham, D.; Faulds, K.; Gottlieb, H. E.; Chill, J. H.; Lampel, A. Biomolecular condensates formed by designer minimalistic peptides. *Nat. Comm.* **2023**, *14*, 1–11.
- [148] Di Leo, S.; Marni, S.; Plata, C. A.; Fraccia, T. P.; Smith, G. P.; Maritan, A.; Suweis, S.; Bellini, T. Pairing statistics and melting of random DNA oligomers: Finding your partner in superdiverse environments. *PLoS Comp. Biol.* **2022**, *18*, e1010051.
- [149] Lu, T.; Spruijt, E. Multiphase complex coacervate droplets. *J. Am. Chem. Soc.* **2020**, *142*, 2905–2914.
- [150] Lyons, H.; Veetil, R. T.; Pradhan, P.; Fornero, C.; De La Cruz, N.; Ito, K.; Eppert, M.; Roeder, R. G.; Sabari, B. R. Functional partitioning of transcriptional regulators by patterned charge blocks. *Cell* **2023**, *186*, 327–345.