Decoding the physical principles of two-component biomolecular phase separation

Yaojun Zhang

Center for the Physics of Biological Function, Princeton University

Bin Xu and Benjamin G. Weiner

Department of Physics, Princeton University

Yigal Meir

Department of Physics, Ben Gurion University of the Negev

Department of Physics, Princeton University and

Department of Molecular Biology, Princeton University

Ned S. Wingreen*

Department of Molecular Biology, Princeton University and

Lewis-Sigler Institute for Integrative Genomics, Princeton University

(Dated: August 24, 2020)

Cells possess a multiplicity of non-membrane bound compartments, which form via liquid-liquid phase separation. These condensates assemble and dissolve as needed to enable central cellular functions. One important class of condensates is those composed of two associating polymer species that form one-to-one specific bonds. What are the physical principles that underlie phase separation in such systems? To address this question, we employed coarse-grained molecular dynamics simulations to examine how the phase boundaries depend on polymer valence, stoichiometry, and binding strength. We discovered a striking phenomenon – for sufficiently strong binding, phase separation is suppressed at rational polymer stoichiometries, which we termed the magic-ratio effect. We further developed an analytical dimer-gel theory that confirmed the magic-ratio effect and disentangled the individual roles of polymer properties in shaping the phase diagram. Our work provides new insights into the factors controlling the phase diagrams of biomolecular condensates, with implications for natural and synthetic systems.

* Author to whom correspondence should be addressed. Email: wingreen@princeton.edu.

I. INTRODUCTION

Eukaryotic cells are host to a multiplicity of non-membrane bound compartments. Recent studies have shown that these compartments form via liquid-liquid phase separation [1–3]. The phase-separated condensates enable many central cellular functions – from ribosome assembly, to RNA regulation and storage, to signaling and metabolism [2, 4]. Unlike conventional liquid-liquid phase separation, e.g. water-oil demixing, the underlying interactions that drive biomolecular phase separation typically involve strong one-to-one saturable interactions, often among multiple components. As a result, the phase diagrams of biomolecular condensates are complex and are sensitive to a variety of physical properties of the biomolecules, included number of binding sites, binding strengths, and additional non-specific interactions. Importantly, these physical parameters can be subject to biological regulation, and can thus directly impact the organization and function of the condensates. It is therefore crucial to understand how the physical properties of the components shape the phase diagram of biomolecular condensates.

Biomolecular condensates typically contain tens to hundreds of types of molecules. Yet, when characterized in detail, only a small number of components are responsible for condensate formation [4]. One class of such condensates are those formed by the association of two essential components. In the simplest case, each component consists of repeated domains that bind in a one-to-one fashion with the domains of the other component. Such two-component condensates have been observed in several natural and engineered contexts, including the algal pyrenoid, P-bodies, and promyelocytic leukemia (PML) nuclear bodies [5–7].

Previous simulations [5, 8] of average cluster size in such two-component systems revealed a striking phenomenon – for sufficiently strong binding, the formation of large clusters is suppressed when the valence of one species equals or is an integral multiple of the valence of the other species, favoring the formation of small stable oligomers instead of a condensate. The phenomenon reminiscent of the exact filling of atomic shells leading to the unreactive noble gases was termed the “magic-number” effect. However, as cluster size is subject both to finite-size effects and to ambiguity between a thermodynamic phase transition and a sol-gel percolation transition [9], it provides at best a qualitative measure of phase separation. Moreover, the previous studies focused on equal monomer stoichiometry, whereas biomolecular condensates cover a broad range of stoichiometries both *in vitro* [6, 7] and *in vivo* [10].

Here, we directly delineate the full phase diagram of such two-component systems. Using coarse-grained molecular dynamics simulations, we explore systematically how phase boundaries

depend on valence, stoichiometry, and binding strength of two associating polymers. Our studies reveal an unanticipated effect – when the numbers of polymers of the two types have a rational stoichiometry (1:1, 1:2, etc.) and their valences are such that small oligomers can form with few unsatisfied bonds, then phase separation is strongly suppressed, which we call the “magic-ratio” effect. To understand the magic-ratio effects better, we develop a two-component sticker theory à la Semenov and Rubinstein [11, 12]. We model the system as dominated by polymer dimers in the dilute phase and by a condensate of independent stickers in the dense phase. The resulting analytical theory captures the magic-ratio effect discovered in simulations, and allows us to disentangle the individual roles of valence, stoichiometry, specific-bond strength, and non-specific attraction in determining the phase boundaries of two-component multivalent systems. Living cells regulate the valence and interactions of biomolecules through chemical modification, or on a slower timescale, tune the stoichiometry via synthesis/degradation or sequestration, and over evolutionary time, adapt the strength of specific and non-specific interactions through mutation of molecular sequences. Understanding the individual roles of these biologically tunable variables thus brings new insights into possible cellular strategies for regulating the formation and dissolution of biomolecular condensates.

II. SIMULATION RESULTS

We perform coarse-grained molecular-dynamics simulations using LAMMPS [13] to determine the phase boundaries of two-component multivalent systems (Fig. 1). Briefly, we model the two polymer species as flexible linear chains of beads connected by harmonic springs (Fig. 1A). Each bead represents one associative domain of the polymer. To ensure associative domains of different polymer types bind in a one-to-one fashion, we impose a finite-ranged attractive interaction between beads of different types. This, however, could lead to more than one-to-one associations. Therefore, to avoid such unwanted associations, we impose strong repulsive interactions between beads of the same type over a large enough range to prevent other beads overlapping with a bound pair, thus preventing multiple-to-one binding (Figs. 1B and S1), see Methods and Supplemental Material for details.

To find the phase boundaries, we simulate hundreds of polymers of types A and B with, respectively, m and n monomers (an $A_m:B_n$ system) in a box with periodic boundary conditions (Figs. 1C and 1D). We initialize the system by constructing a dense slab of polymers in the middle of the box [14]. The system evolves and relaxes according to Langevin dynamics [15]. After the system

FIG. 1. Coarse-grained molecular-dynamics simulations of two-component multivalent associative polymers. (A) The system consists of two types of polymers A (blue) and B (red) of varying lengths and concentrations. Depicted are A and B polymers of length 10, denoted as A₁₀ and B₁₀. Each polymer is modeled as a linear chain of spherical particles connected by harmonic bonds. Monomers of different types interact pairwise through an attractive potential, while repulsion between monomers of the same type prevents them from overlapping and thus ensures one-to-one binding of monomers of different types (see Methods and Supplemental Material for details). (B) Snapshots of dimers formed by A₁₀ and B₁₀ with one-to-one bonds. (C) Snapshot of a simulation with 125 A₁₀ and 125 B₁₀ polymers. The system phase separates into a dense phase (middle region) and a dilute phase (two sides) in a 250nm×50nm×50nm simulation box with periodic boundary conditions. (D) Same as C but with 138 A₁₀ and 112 B₁₀ polymers, yielding an overall monomer concentration ratio (stoichiometry) 1.23. (E) Monomer concentration profiles of A₁₀:B₁₀ systems at various overall stoichiometries (total monomer concentration 6.64mM), each with the center of the dense phase aligned at $x = 0$ and averaged over time and over ten simulation repeats (see Methods and Supplemental Material). All simulations performed in LAMMPS [13].

has achieved equilibrium, two phases coexist: a dilute phase consisting of dimers and other small oligomers, and a dense phase of an interconnected polymer condensate. We measure the corresponding density profile (Fig. 1E) and calculate the dilute- and dense-phase concentrations by averaging the density profile over the regions ($x \leq -100\text{nm}$ or $x \geq 100\text{nm}$) and ($-10\text{nm} \leq x \leq 10\text{nm}$), respectively. See Methods and Supplemental Material for simulation details.

A. Equal Monomer Stoichiometry

It was shown previously that for equal monomer stoichiometry in the strong-binding regime, clustering is substantially suppressed when the number of binding sites on one polymer species is an integer multiple of the number of binding sites on the other, as this condition favors the assembly of small oligomers in which all binding sites are saturated [5, 8]. What does this magic-number effect imply for the actual phase diagram? To address this question, we fix the valence of polymer A at 14 and systematically vary the valence of polymer B from 5 to 16 while keeping the two monomer concentrations the same, i.e., at equal global monomer stoichiometry.

Figures 2A and 2B show simulation results for total monomer concentrations of the dilute and dense phases for $A_{14}:B_5$ to $A_{14}:B_{16}$ systems. In the strong binding regime, for magic-number cases, i.e. when the valence of B is 7 or 14, the dilute-phase concentration shows pronounced peaks (Fig. 2A, black curve). What is the origin of the peak at $A_{14}:B_{14}$? Intuitively, when the dilute phase of the two-component system is dominated by dimers (for systems $A_{14}:B_{12}$ to $A_{14}:B_{16}$, as supported by cluster size analysis in Fig. S2), each of these dimers has high translational entropy, whereas polymers in the dense condensate have low translational entropy. For $A_{14}:B_{14}$, all binding sites can pair up in a dimer just as well as in the condensate, so the energy per polymer is not necessarily lower in the condensate. Why then is the condensate still competitive with the dilute phase? In a dimer, the binding sites of A_{14} must match all the binding sites of B_{14} , leading to a reduced overall conformational entropy. By comparison, the polymers in the condensate are more independent, binding to multiple members of the other species and enjoying a relatively higher overall conformational entropy. Because the translational entropy of each dimer decreases as their concentration goes up, the condensed phase eventually becomes more favorable and so the system phase separates with increasing concentration. Therefore, phase separation in $A_{14}:B_{14}$ is primarily driven by a competition between translational entropy and conformational entropy.

By contrast, for $A_{14}:B_{13}$ and $A_{14}:B_{15}$, one of the monomers in the dimer cannot be paired, and for $A_{14}:B_{12}$ and $A_{14}:B_{16}$, two monomers per dimer cannot be paired. Therefore, forming a condensate not only increases the conformational entropy but more importantly lowers the energy of these systems. This significantly tilts the balance in favor of condensation. As a result, the dilute-phase concentration is sharply peaked at $A_{14}:B_{14}$, falling off rapidly for increasingly unequal polymer lengths. We note that the dense-phase concentration shows no such feature (Fig. 2B), indicating that the peak at $A_{14}:B_{14}$ does not arise from differences in the internal structure of the dense phase.

FIG. 2. Simulations of associative polymers reveal a magic-number effect with respect to relative valence. Monomer concentrations in (A) dilute and (B) dense phases for simulated polymer systems at different binding strengths. U_0 denotes the depth of the potential well, in units of $k_B T$ (see Methods and Supplemental Material for details). The valence of polymer A is 14, and the valence of polymer B ranges from 5 to 16. Global monomer stoichiometry is 1 and total monomer concentration is 6.64mM. Histograms of cluster size in (C) A₁₄:B₁₄ and (D) A₁₄:B₆ systems, for $U_0 = 14$. “Counts” refer to number of clusters. Cluster size is measured in monomers. Red dots indicate the dominant oligomer in the dilute phase.

The dilute phase of two-component systems is not always dominated by dimers (Fig. S2). For example, the dilute phase of the A₁₄:B₇ system is dominated by fully-bonded trimers with 1 A₁₄ and 2 B₇, the dilute phase of A₁₄:B₈ is dominated by trimers with 1 A₁₄ and 2 B₈, which has two unpaired monomers per trimer, and the dilute phase of A₁₄:B₆ is dominated by oligomers with 3 A₁₄ and 7 B₆, which although fully-bonded is not small (Fig. 2D). Consistent with the above logic, we find another peak in the dilute-phase concentration at A₁₄:B₇ (Fig. 2A). More generally, in contrast to the magic-number systems, the dilute phases in other cases are dominated by oligomers which are not capable of being fully bonded (high energy) and/or not small (low translational entropy) (Fig. S2). The dilute-phase concentration is therefore lower in these non-magic-number cases.

How do the phase boundaries depend on the strength of binding? Figure 2A shows that, for non-magic-number systems, the dilute-phase concentration decreases monotonically with increasing binding strength, whereas for magic-number systems the dependence can be non-monotonic. This

difference is attributed to the distinct underlying driving forces for phase separation. For non-magic-number systems, as clustering allows a larger fraction of binding sites to be paired, the stronger the binding, the more the energy is lowered by condensate formation. Therefore, the dilute-phase concentration drops as binding strength increases (or as temperature decreases). Such energy-dependence is expected for conventional phase-separation models, such as Flory-Huggins [16, 17].

Interestingly, for the magic-number system $A_{14}:B_{14}$, the dilute-phase concentration first decreases with increasing binding strength in the weak binding regime, similar to non-magic-number systems. However, as the binding energy is increased further, most of binding sites pair up in both dilute and dense phases. Phase separation is then primarily driven by a competition between conformational and translational entropy. The pairing up of binding sites reduces the conformational entropy of both the dense and dilute phases. By contrast, the translational entropy of the dilute-phase components is almost unaffected. Consequently, the dilute phase becomes more competitive relative to the condensate, so the dilute phase boundary shifts to higher concentration.

By comparison, the dense-phase concentration increases monotonically with increasing binding strength for all systems (Fig. 2B). This follows because the stronger the binding, the more monomers are paired, which tightens the condensate structure. We note that at substantially higher binding energies than studied here, essentially all the binding sites are satisfied in both magic-number and non-magic-number systems, and the phase boundaries become independent of binding energy.

B. Unequal Monomer Stoichiometry

How do the phase boundaries depend on overall monomer stoichiometry? Figures 3A and 3B show total monomer concentrations of the dilute and dense phases for magic-number systems $A_8:B_8$ to $A_{14}:B_{14}$ at different global monomer stoichiometries. For each system, the dilute-phase concentration peaks at equal monomer ratio, falls off initially as the ratio deviates from 1, and then curves back up. What is the origin of the peak at equal stoichiometry? Recall that, in the strong binding regime, phase separation of magic-number systems is primarily driven by a competition between translational entropy and conformational entropy. Now consider starting with a system at equal monomer concentration, and adding more of one polymer species to the system. At the beginning, the added polymers readily enter the dense phase, which relaxes the conformational constraint that every monomer in the condensate has to pair with a partner. This

FIG. 3. Simulations of associative polymers reveal a magic-ratio effect with respect to polymer stoichiometry. Monomer concentrations in (A) dilute and (B) dense phases for equal polymer length systems (i.e., $A_n:B_n$) at different global stoichiometries. Monomer concentrations in (C) dilute and (D) dense phases for systems where polymer B is 1 monomer shorter than polymer A (i.e., $A_n:B_{n-1}$) at different stoichiometries; black dots indicate cases where the number of polymers of each type is the same. Interaction strength $U_0 = 14$ and total monomer concentration is 6.64mM.

increase of the conformational entropy of the condensate makes it more competitive, so the dilute-phase concentration decreases. However, as the ratio between the two polymers is increased further, it becomes possible to form a spectrum of dilute-phase oligomers which typically contain one extra polymer of the majority type (Table S1). These new oligomers have more relaxed structures than fully-bonded dimers, which raises the conformational entropy of the dilute phase. Therefore, the dilute phase is favored over the condensate and its concentration curves back up.

Figure 3A also reveals that the dilute-phase concentration decreases with increasing polymer valence. This follows in part because translational entropy in the dilute phase is per dimer center of mass, whereas conformational entropy in both phases scales with the number of monomers. The entropic gain of joining the dense phase is therefore more on a per monomer basis for longer polymers, so the dilute-phase concentration decreases with increasing valence. As a less apparent yet important point, Fig. 3A also shows that increasing polymer valence enhances both the width and relative height of the peak in the dilute-phase concentration.

Figures 3C and 3D show total monomer concentrations of the dilute and dense phases for un-

equal valence polymers $A_8:B_7$ to $A_{14}:B_{13}$ at different global monomer stoichiometries. The dilute phase boundary shows a symmetric minimum around equal stoichiometry for $A_8:B_7$, yet surprisingly, the phase boundary becomes asymmetric and then peaks at equal *polymer* stoichiometry with increasing polymer length (Fig. 3C). What is the origin of these peaks? Taking the $A_{14}:B_{13}$ system as an example, its dilute phase is dominated by dimers with an unpaired A monomer. This strongly disfavors the dilute phase in the strong binding regime at equal monomer stoichiometry. However, as the overall A:B monomer stoichiometry increases, the excess As cannot be paired anyway. In particular, at equal polymer stoichiometry (denoted as black dots in Fig. 3C), forming dimers is no longer energetically costly. The result is strong suppression of phase separation at equal polymer stoichiometry. We term this the "magic-ratio" effect, as it occurs for rational ratios of associative polymers with strong one-to-one bonds. The effect is enhanced for longer polymers with few unsatisfied bonds, hence the clear peak in the dilute phase concentration for the $A_{14}:B_{13}$ system. Taking this case as an example, we can simply understand why the peak occurs at equal polymers stoichiometry. To the left of the $A_{14}:B_{13}$ peak at equal polymer stoichiometry, the dilute-phase concentration is low as dimers are energetically disfavored because more bonds can be satisfied in the condensate. By contrast, to the right of the peak, the dilute-phase concentration is low for a different reason – because the condensate is entropically favored, similar to the peak with respect to stoichiometry for magic-number systems. Eventually, the dilute-phase concentration curves back up due to formation of higher oligomers in the dilute phase, as discussed for magic-number systems.

We note that for all these systems the dense-phase concentration shows no such striking features. Rather, the concentration decreases monotonically as the global monomer stoichiometry departs from 1 and as the valence of polymers decreases (Figs. 3B and 3D).

Above, we considered the role of both relative valence and relative stoichiometry. By plotting phase boundaries as joint functions of valence and stoichiometry, we obtain a unified picture: Figures 4A and 4B show the dilute- and dense-phase concentrations for systems $A_{14}:B_{12-16}$ at global monomer stoichiometries 14:12-16. Notably, the dilute-phase concentration is peaked along the diagonal (Fig. 4A), i.e. at equal *polymer* stoichiometry, which we termed the "magic-ratio effect". Intuitively, all cases along the diagonal favor 1:1 polymer dimers: the dimers enjoy high translation entropy and there is no energy penalty involved in their formation. Thus, a dilute phase of dimers is strongly favored at equal polymer stoichiometry.

FIG. 4. Simulations of associative polymers reveal a magic-ratio effect with respect to polymer stoichiometry. Monomer concentrations in (A) dilute and (B) dense phases for systems $A_{14}:B_{12-16}$ at global stoichiometries 14:12-16. Interaction strength $U_0 = 14$ and total monomer concentration is 6.64mM.

III. DIMER-GEL THEORY RESULTS

While our simulations have revealed that a magic-ratio effect influences the boundaries of phase separation for associating polymers, we desire a deeper understanding of the interplay of factors such as overall valence, stoichiometry, and interaction strength. To this end, we develop a mean field theory of two-component associative polymers à la Semenov and Rubinstein [11, 18].

Specifically, we consider a system of A and B polymers as in our simulations. Each polymer is a linear chain of L_1 or L_2 monomers of type A or type B, respectively. Without loss of generality, we take $L_1 \geq L_2$. Monomers (stickers) of different types associate in a one-to-one fashion. Our simulations suggest that for polymers of similar valence close to equal polymer stoichiometry the dilute phase is dominated by dimers and the dense phase is a condensate. Therefore, we assume that polymers can associate either as dimers or, alternatively, as a condensate in which pairs of monomers bind independently. This assumption of independence is a mean field approximation, as it neglects correlations between monomers in the same chain, and thus only applies when the polymers strongly overlap, i.e. at densities above the semidilute regime [19].

The partition function of such a system can be divided into three parts: $Z = Z_{\text{ni}}Z_sZ_{\text{ns}}$, where Z_{ni} , the partition function of a solution of non-interacting polymers, captures the translational and conformational entropy of the two polymer species, Z_s captures specific interactions between associating monomers, and Z_{ns} captures all non-specific interactions.

The corresponding free-energy density for the mixed non-interacting polymers is [11]

$$\frac{F_{\text{ni}}}{k_B T} = \frac{c_1}{L_1} \ln \frac{c_1}{e L_1} + \frac{c_2}{L_2} \ln \frac{c_2}{e L_2}, \quad (1)$$

where c_1 and c_2 are the concentrations of A and B polymers measured in terms of monomers. Note

that the terms for the conformational entropy of non-interacting polymers are omitted in Eq. (1), as they are linear in c_1 and c_2 and thus do not influence the phase boundaries.

To include specific interactions, we first consider the partition function $Z_s(N_{d1}, N_{d2}, N_b)$ for states with exactly N_{d1} and N_{d2} total numbers of monomers of A and B types in dimers (i.e. number of dimers equals $N_{d1}/L_1 = N_{d2}/L_2$) and N_b additional monomer pairs,

$$Z_s(N_{d1}, N_{d2}, N_b) = P(N_{d1}, N_{d2}, N_b) W(N_{d1}, N_{d2}, N_b) \exp(N_{d1}\epsilon_d/L_1 + N_b\epsilon_b). \quad (2)$$

In Eq. (2), P is the number of different ways that polymers and monomers can be chosen to pair up to form dimers and independent bonds,

$$P(N_{d1}, N_{d2}, N_b) = \binom{N_1/L_1}{N_{d1}/L_1} \binom{N_2/L_2}{N_{d2}/L_2} (N_{d1}/L_1)! \binom{N_1 - N_{d1}}{N_b} \binom{N_2 - N_{d2}}{N_b} N_b!, \quad (3)$$

where N_1 and N_2 are the total numbers of monomers of A and B types. (Note that in Eq. (3) if $L_1 > L_2$, the excess monomers of type A in dimers do not form additional bonds.) In Eq. (2), W is the probability that all chosen polymers and monomers are, respectively, close enough to their specified partners in the non-interacting state to form dimers and independent bonds,

$$W(N_d, N_b) = \left(\frac{v_d}{V}\right)^{\frac{N_{d1}}{L_1}} \left(\frac{v_b}{V}\right)^{N_b}, \quad (4)$$

where v_d and v_b are effective interaction volumes and V is the system volume. The last term in Eq. (2) is the Boltzmann factor for specific interactions, where ϵ_d and ϵ_b are the effective binding energies of dimers and monomer pairs, in units of $k_B T$.

The part of the free-energy density due to specific interactions is

$$\frac{F_s}{k_B T} = -\frac{1}{V} \ln Z_s. \quad (5)$$

Using Stirling's approximation $\ln N! = N \ln N - N$, we obtain

$$\begin{aligned} \frac{F_s}{k_B T} = & -\frac{c_1}{L_1} \ln c_1 + (1 - L_1) \frac{c_1 - c_{d1}}{L_1} \ln(c_1 - c_{d1}) + (c_1 - c_{d1} - c_b) \ln(c_1 - c_{d1} - c_b) \\ & -\frac{c_2}{L_2} \ln c_2 + (1 - L_2) \frac{c_2 - c_{d2}}{L_2} \ln(c_2 - c_{d2}) + (c_2 - c_{d2} - c_b) \ln(c_2 - c_{d2} - c_b) \\ & + \frac{c_{d1}}{L_1} \ln(ec_{d2} L_1 K_d) + c_b \ln(ec_b K_b), \end{aligned} \quad (6)$$

where $K_d \equiv e^{-\epsilon_d}/v_d$ and $K_b \equiv e^{-\epsilon_b}/v_b$ are, respectively, the dissociation constants of a dimer and of a pair of monomers. c_{d1} and c_{d2} are the concentrations of monomers of A and B types in dimers (so $c_{d1}/L_1 = c_{d2}/L_2$), and c_b is the concentration of independent bonds.

In the thermodynamic limit, F_s will be minimized with respect to c_{d1} , c_{d2} and c_b , which implies

$$K_d c_{d2} L_1 (c_1 - c_{d1})^{L_1-1} (c_2 - c_{d2})^{L_2-1} = (c_1 - c_{d1} - c_b)^{L_1} (c_2 - c_{d2} - c_b)^{L_2}, \quad (7)$$

$$K_b c_b = (c_1 - c_{d1} - c_b) (c_2 - c_{d2} - c_b). \quad (8)$$

Note that if c_b in Eq. (7) and c_{d1} and c_{d2} in Eq. (8) are set to zero, these equations reduce to

$$K_d \rho_d = (\rho_1 - \rho_d) (\rho_2 - \rho_d), \quad (9)$$

$$K_b c_b = (c_1 - c_b) (c_2 - c_b), \quad (10)$$

where ρ_1 , ρ_2 , and ρ_d are the total concentrations of A and B polymers and dimers (measured in polymeric units), i.e., $\rho_1 = c_1/L_1$, $\rho_2 = c_2/L_2$, and $\rho_d = c_{d1}/L_1 = c_{d2}/L_2$. Eqs. (9) and (10) are consistent with the definitions of the dissociation constants of a dimer and of an independent bond, respectively.

The free-energy density due to non-specific interactions can in general be written as a power expansion in the concentrations [11, 19],

$$\frac{F_{ns}}{k_B T} = \frac{1}{2} \sum_{ij} v_{ij} c_i c_j + \frac{1}{6} \sum_{ijk} w_{ijk} c_i c_j c_k, \quad (11)$$

where the sum is over all the species in the system, including free polymers/monomers, dimers and independent bonds, and v_{ij} and w_{ijk} are two- and three-body interaction parameters. For our simulation system, we derive a specific form of F_{ns} by taking into account that (1) we are interested in the strong-binding regime where the magic-ratio effect is observed, (2) there is no non-specific interaction between free polymers of different types in our simulation, and (3) non-specific interactions are only important at high concentrations. The result is

$$\frac{F_{ns}}{k_B T} = \frac{v_b}{2} \max(c_1, c_2)^2 + \frac{w_b}{6} \max(c_1, c_2)^3, \quad (12)$$

where v_b and w_b are the two- and three-body interaction parameters for a solution of independent bonds. See Appendix ?? for details of the derivation.

Finally, substituting the conditions Eqs. (7) and (8) into Eq. (6), we obtain the total free-energy density $F = F_{ni} + F_s + F_{ns}$,

$$\begin{aligned} \frac{F}{k_B T} = & \frac{c_1}{L_1} \ln \frac{c_1 - c_{d1}}{e L_1} + c_1 \ln \frac{c_1 - c_{d1} - c_b}{c_1 - c_{d1}} + \frac{c_{d1}}{L_1} + \\ & \frac{c_2}{L_2} \ln \frac{c_2 - c_{d2}}{e L_2} + c_2 \ln \frac{c_2 - c_{d2} - c_b}{c_2 - c_{d2}} + c_b + \frac{v_b}{2} \max(c_1, c_2)^2 + \frac{w_b}{6} \max(c_1, c_2)^3, \end{aligned} \quad (13)$$

where c_{d1} , c_{d2} , and c_b are the solutions of Eqs. (7) and (8). Equations (7), (8), and (13) form a complete set which predicts the free-energy density of the two-component associative polymer system at given total monomer concentrations, c_1 and c_2 , of the two species.

Intuitively, in the strong-binding regime, i.e. when $c_1, c_2 \gg K_d, K_b$, polymers either associate as dimers or as independent bonds depending on their relative free energies. In the limit that dimers are preferred ($\rho_d = \min(\rho_1, \rho_2)$ and $c_b = 0$), the contribution from specific interactions is

$$\frac{F_s^{\text{dim}}}{k_B T} = \rho_d \ln K_d + (\rho - \rho_d) \ln \frac{\rho - \rho_d}{e} - \rho \ln \frac{\rho}{e}, \quad (14)$$

where $\rho = \max(\rho_1, \rho_2)$ is the concentration of the majority species in polymeric units. The terms on the right of Eq. (14) reflect, respectively, the free-energy density due to dimer formation, translational entropy of leftover polymers, and loss of translational entropy of the majority species (in effect, the formation of each dimer removes the translation entropy of one free polymer). In the opposite limit that independent bonds are preferred ($c_b = \min(c_1, c_2)$ and $\rho_d = 0$),

$$\frac{F_s^{\text{ind}}}{k_B T} = c_b \ln K_b + (c - c_b) \ln \frac{c - c_b}{e} - c \ln \frac{c}{e}, \quad (15)$$

where $c = \max(c_1, c_2)$, and the terms are analogous to those in Eq. (14). Numerical studies show that the full $F_s(c_1, c_2)$ in Eq. (6) is always well approximated by the lower of the two limiting values of F_s (Eqs. (14) and (15)).

In which regions of concentration space are dimers versus independent bonds preferred? For a magic-number system composed of two polymer species of valence L at equal monomer stoichiometry, $F_s^{\text{dim}}/k_B T = \rho \ln(K_d e/\rho)$ and $F_s^{\text{ind}}/k_B T = c \ln(K_b e/c)$. Comparing the two expressions, dimers are favored at low concentrations, whereas a gel connected by independent bonds is favored at high concentrations. The transition occurs when $F_s^{\text{dim}} = F_s^{\text{ind}}$, i.e. at concentration $c_0 = e(K_b^L/(K_d L))^{1/(L-1)}$. Away from equal stoichiometry, the transition occurs at a lower concentration $c_s = c_0(s-1)^{s-1}s^{-s}$, where $s = \max(c_1, c_2)/\min(c_1, c_2) > 1$ (see Supplemental Material for details). As c_s decreases rapidly with increasing s (Fig. 5A inset, white curve), the preference for dimers over a gel exhibits a sharp peak around equal stoichiometry.

To give a concrete example of the above analysis, we extract the values of K_d for dimers from simulations, choose a value of K_b for independent bonds close to the dissociation constant of a pair of monomers (see Supplemental Material for details), and numerically solve Eqs. (7) and (8) for c_{d1} , c_{d2} , and c_b to find the fraction of monomers in dimers and independent bonds for all concentrations (c_1, c_2) . We find that indeed for polymers of equal valence, dimers are favored at low concentrations and independent bonds at high concentrations. The dimer dominated region extends sharply to

higher concentrations in a narrow zone around the diagonal, as quantitatively captured by c_s (Figs. 5A inset and S3A). For polymers of similar but unequal valence, the dimer dominated region extends to higher concentrations along the direction of equal *polymer* stoichiometry (Figs. 5B inset and S3B).

Finally, to extract the phase boundaries, we substitute the values of c_{d1} , c_{d2} , and c_b into Eq. (13) to first obtain the free energy as a function of c_1 and c_2 . The free-energy landscape has two basins, one at small concentrations corresponding to the dilute dimer-dominated phase, and one at high concentrations corresponding to the dense independent-bond-dominated gel-phase (Fig. S4). We locate the phase boundaries by applying convex-hull analysis to this free-energy landscape (see Supplemental Material).

Does the dimer-gel theory capture the magic-ratio effect revealed by our simulations? Figures 5A and 5B show the phase diagrams of $A_8:B_8$ and $A_8:B_7$ systems. In both cases, the phase boundaries on the dilute side extend sharply into the two-phase region along the direction of equal polymer stoichiometry (tie lines along this direction are denoted by black dots). Figures 5C and 5D show the dilute- and dense-phase concentrations for systems $A_8:B_{6-10}$ at global monomer stoichiometries 8:6-10. Notably, the dilute-phase concentrations are substantially shifted up around the diagonal, verifying the magic-ratio effect observed in simulations (Fig. 4A).

One of the major assumptions of the dimer-gel theory is the mean-field approximation, i.e. monomers of different types can associate independently in the dense phase. This approximation captures a key feature of the dense phase, namely that a single polymer binds to multiple partners. Nevertheless, because monomers belonging to the same polymer are tethered together, correlations in binding exist (Fig. S7). Therefore, what should be considered to be “independent” is not individual monomers but rather segments of the binding correlation length (~ 1.8 monomers). The dense phase of a valence 14 system is thus more accurately described by the theory at valence $14/1.8 \approx 8$. We therefore present results for valence 8 systems in Fig. 5. (The theoretical phase diagrams and the dilute- and dense-phase concentrations for valence 14 systems also verify the magic-ratio effect (Fig. S6).)

The dimer-gel theory has only a handful of parameters: the valences L_1 and L_2 of polymers A and B, the dissociation constants K_d and K_b of dimers and independent bonds, and the non-specific interaction parameters v_b and w_b . How are the phase boundaries and the magic-ratio effects determined collectively by these parameters? If valence is increased while keeping all other parameters fixed in the theory, for equal valence polymers we find that the dilute-phase concentration decreases, while the dense-phase concentration increases, and the peak with respect to stoichiometry

FIG. 5. A dimer-gel theory predicts the magic-ratio effect. Phase diagrams of (A) $A_8:B_8$ and (B) $A_8:B_7$ systems: one-phase region white, two-phase region green. The dilute- and dense-phase concentrations are connected by representative tie lines. The tie line along the direction of equal polymer stoichiometry is denoted with a black dot. Insets: fraction of monomers in dimers for (A) $A_8:B_8$ and (B) $A_8:B_7$ systems. White curve in A inset is the transition boundary between dimer- and independent bonds-dominated regions predicted by c_s . Dashed white line in (B) inset denotes equal polymer stoichiometry. Monomer concentrations in (C) dilute and (D) dense phases for systems $A_8:B_{6-10}$ at global monomer stoichiometries 8:6-10. The total monomer concentration is the same as in simulations, 6.64 mM. For details see Supplemental Material. Parameters: $v_b = 9 \times 10^{-2} \text{mM}^{-1}$, $w_b = 7 \times 10^{-3} \text{mM}^{-2}$, $K_b = 3.8 \times 10^{-3} \text{mM}$, and K_d values in Table S3.

is enhanced in terms of the dilute-phase peak-to-valley ratio (Figs. S8A and S8B). If valence is increased for unequal valence polymers, we observe that the shape of the dilute phase boundary transitions from a shoulder to a peak (Figs. S8C and S8D). All these features are consistent with the simulation results in Fig. 3.

For the theory to agree quantitatively with the phase boundaries from simulations, we find that smaller values of non-specific interaction parameters are necessary for higher valence systems (Figs. S8C and S8D). Intuitively, this follows because higher valence polymers have more backbone bonds, which bring bound monomer pairs closer together in the dense phase – effectively reducing the non-specific repulsion between them. Finally, the dimer-gel theory also predicts that the magic-

ratio effect disappears in the weak-binding regime (Fig. S9), consistent with our simulation results (Fig. 2).

IV. DISCUSSION

Intracellular phase separation is driven by multivalent interactions between macromolecules. These interactions are separated into two classes [4, 20]: (1) specific interactions, such as binding between protein domains, are relatively strong and involve specific partners, and (2) non-specific interactions, such as electrostatic and hydrophobic interactions, which are much weaker, more generic, and non-saturable. Multivalent systems with specific interactions allow for “orthogonal” condensates to form: the specific interactions holding together one class of droplets will typically not interfere with those holding together another class. Motivated by the key role of specific interactions in intracellular phase separation, we focused on exploring the effects of specific interactions on the phase boundaries of two-component associative polymers. Specifically, we combined coarse-grained molecular dynamics simulations and analytical theory to examine the individual roles of valence, stoichiometry, and binding strength on the phase boundaries. In particular, we identified a magic-ratio effect: for sufficiently strong binding, phase separation is strongly suppressed at equal *polymer* stoichiometry.

The magic-ratio effect occurs exclusively in the strong-binding regime. Are specific protein-protein, protein-RNA, and RNA-RNA interactions strong enough to lead to the magic-ratio effect? The onset of the effect in our simulations occurs around $U_0 = 9k_B T$ (Fig. 2A), which corresponds to a monomer-monomer dissociation constant $K_d = 0.4\text{mM}$. This value is consistent with the onset K_d of 1-2.5mM estimated from 3D lattice simulations with one polymer and one rigid component [8]. For comparison, the measured K_d for a SUMO protein domain with a SIM peptide is 0.01mM [7] and for an SH3 domain and a PRM peptide is 0.35mM [6]. Thus for systems as strongly interacting as SUMO-SIM or SH3-PRM, the magic-ratio effect in principle allows for novel mechanisms of regulation. For example, chemical modification of the effective valence of one component to change into or out of a magic-ratio condition could shift the phase boundary as a possible means of condensate regulation [5]. Magic-ratio effects could also manifest in other experimental systems, such as non-biological polymers, DNA origami [21], or patchy colloid systems [22]. As an inverse problem, the magic-ratio effect could be exploited to determine the relative valence of associating biomolecules by measuring their phase diagram.

While the magic-ratio effect is robust with respect to the strength of non-specific interactions,

these interactions do strongly influence phase boundaries. For example, the dilute-phase concentrations in our simulations are $\sim 1\text{mM}$, while the reported values for biological systems are typically tens of μM or less. The discrepancy is likely due to different strengths of non-specific attraction. Indeed, increasing the non-specific attraction in simulations from $0.06k_{\text{B}}T$ to $0.12k_{\text{B}}T$ leads to a 50% reduction in the dilute-phase concentration (Fig. S10). Future work will explore the interplay between specific and non-specific interactions, and their roles in determining the physical properties of droplets, such as surface tension, viscosity, etc.

The simulations and theory presented here are aimed at providing conceptual understanding. Quantitative descriptions of related real systems will likely require additional features, such as details of molecular shape and flexibility, linker lengths, as well as range and type of interactions. Nevertheless, we have discovered a universal magic-ratio effect, and provided evidence that it shapes the phase boundaries of multivalent associative polymers. We expect these general findings to be independent of most system details. We hope that our work will stimulate exploration of magic-ratio effects in both natural and synthetic multivalent, multicomponent systems.

V. METHODS

A. Modeling two-component multivalent associative polymers

We perform coarse-grained molecular-dynamics simulations using LAMMPS [13] to simulate two-component multivalent associative polymers. Individual polymers are modeled as linear chains of spherical particles connected by harmonic bonds (Fig. 1A). Harmonic bonds are modeled using a harmonic potential:

$$U_{\text{b}}(r) = k(r - r_{\text{b}})^2, \quad (16)$$

where $r_{\text{b}} = 4.5\text{nm}$ is the mean bond length, $k = 20k_{\text{B}}T/r_{\text{b}}^2$ is the bond stiffness, k_{B} is the Boltzmann constant, and $T = 300\text{K}$ is room temperature.

Monomers of the same type interact through a softened, truncated Lenard-Jones potential:

$$U_{\text{r}}(r) = 4\epsilon\lambda \left[(1 - \lambda)^2 + \left(\frac{r}{\sigma}\right)^6 \right]^{-2} - 4\epsilon\lambda \left[(1 - \lambda)^2 + \left(\frac{r}{\sigma}\right)^6 \right]^{-1} \quad \text{for } r < r_{\text{c}}, \quad (17)$$

where $\epsilon = 0.15k_{\text{B}}T$, $\lambda = 0.68$, $\sigma = 3.5\text{nm}$, and $r_{\text{c}} = 5\text{nm}$. These parameters effectively lead to a monomer of diameter $d \simeq 3\text{nm}$ and a weak attractive tail of depth $0.06k_{\text{B}}T$. The weak attractive tail is employed solely to promote a more compact condensate.

Monomers of different types interact through an attractive potential:

$$U_a(r) = -\frac{1}{2}U_0 \left(1 + \cos \frac{\pi r}{r_0}\right) \quad \text{for } r < r_0, \quad (18)$$

where $U_0 = 14k_B T$ is used in all simulations, except as indicated in Fig. 2. The attraction cut-off distance is $r_0 = 2\text{nm}$. Note that due to the strong repulsion between monomers of the same type, simultaneous binding of two monomers of one type to a monomer of the other type is energetically highly disfavored. This ensures one-to-one binding of monomers of different types (Figs. 1B and S1D).

B. Determining the phase boundaries

Each system consists of n_1 and n_2 polymers of types A and B, respectively. The number of polymers are determined by their valences/lengths (L_1 and L_2) and global A:B monomer stoichiometry (s) through

$$n_1 = \text{round} \left(\frac{Ns}{L_1(1+s)} \right), \quad (19)$$

$$n_2 = \text{round} \left(\frac{N}{L_2(1+s)} \right). \quad (20)$$

The round function is used as we can only simulate an integer number of polymers. For all simulations except those in Fig. 4, we use $N = 2500$ and a box of size $250\text{nm} \times 50\text{nm} \times 50\text{nm}$. Simulations in Fig. 4 are performed with $N = 5000$ and a box of size $315\text{nm} \times 63\text{nm} \times 63\text{nm}$.

Simulations are equilibrated using a Langevin thermostat in the NVT ensemble at $T = 300\text{K}$ with periodic boundary conditions, i.e. the system evolves according to Langevin dynamics [15]

$$m \frac{d^2 \vec{r}_i}{dt^2} = -\gamma \frac{d\vec{r}_i}{dt} - \nabla_{\vec{r}_i} U(\vec{r}_1, \dots, \vec{r}_N) + \vec{f}, \quad (21)$$

where $\vec{r}_i$ is the coordinate of particle i , m is its mass, γ is the friction coefficient, $\vec{f}$ is random thermal noise, and the energy $U(\vec{r}_1, \dots, \vec{r}_N)$ contains all interactions between particles, including harmonic bonds, non-specific, and specific interactions (Eqs. (16)-(18)).

To promote phase equilibrium and ensure that only a single dense condensate is formed, we first initialize the simulation by confining polymers in the region $-50\text{nm} < x < 50\text{nm}$. The attractive interaction between A and B monomers (Eq. (18)) is gradually switched on from $U_0 = 0$ to 14 over 10^8 time steps. This annealing procedure leads to the formation of a dense phase close to its equilibrated concentration. The confinement is then removed, and the system is equilibrated to allow the formation of a dilute phase and relaxation of the dense phase. We record the positions

of all particles for 400 recordings after the system has achieved equilibrium. For each choice of valence and stoichiometry, we perform 10 simulation replicates with different random seeds.

To find the dilute- and dense-phase concentrations, we center the simulation box to the center of mass of the dense phase for each recording. We then compute the monomer concentration histogram along the x axis with a bin size $1/50$ of box length. The resulting concentration profile has high values in the middle corresponding to the dense-phase concentration, and low values on the two sides corresponding to the dilute-phase concentration (Fig. 1E). The dilute- and dense-phase concentrations are calculated by averaging the concentration profile over the regions ($x \leq -100\text{nm}$ or $x \geq 100\text{nm}$) and $(-10\text{nm} \leq x \leq 10\text{nm})$, respectively. Simulations with different total numbers of particles show consistent dilute- and dense-phase concentrations (Table S2), suggesting the effect of finite size is minor.

VI. ACKNOWLEDGMENTS

We thank Guanhua He, Martin Jonikas, and Daniel Lee for insightful suggestions and comments. This work was supported in part by the National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030), and NSF grant 1521553.

-
- [1] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, Germline p granules are liquid droplets that localize by controlled dissolution/condensation, *Science* **324**, 1729 (2009).
 - [2] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry, *Nat. Rev. Mol. Cell Biol.* **18**, 285 (2017).
 - [3] D. Bracha, M. T. Walls, and C. P. Brangwynne, Probing and engineering liquid-phase organelles, *Nat. Biotechnol.* **37**, 1435 (2019).
 - [4] J. A. Ditlev, L. B. Case, and M. K. Rosen, Who’s in and who’s out—compositional control of biomolecular condensates, *J. Mol. Biol.* **430**, 4666 (2018).
 - [5] E. S. F. Rosenzweig, B. Xu, L. K. Cuellar, A. Martinez-Sanchez, M. Schaffer, M. Strauss, H. N. Cartwright, P. Ronceray, J. M. Plitzko, F. Förster, *et al.*, The eukaryotic CO_2 -concentrating organelle is liquid-like and exhibits dynamic reorganization, *Cell* **171**, 148 (2017).
 - [6] P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, *et al.*, Phase transitions in the assembly of multivalent signalling proteins, *Nature* **483**, 336 (2012).

- [7] S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, and M. K. Rosen, Compositional control of phase-separated cellular bodies, *Cell* **166**, 651 (2016).
- [8] B. Xu, G. He, B. G. Weiner, P. Ronceray, Y. Meir, M. C. Jonikas, and N. S. Wingreen, Rigidity enhances a magic-number effect in polymer phase separation, *Nat. Commun.* **11**, 1 (2020).
- [9] T. S. Harmon, A. S. Holehouse, and R. V. Pappu, Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins, *New J. Phys.* **20**, 045002 (2018).
- [10] D. W. Sanders, N. Kedersha, D. S. Lee, A. R. Strom, V. Drake, J. A. Riback, D. Bracha, J. M. Eeftens, A. Iwanicki, A. Wang, *et al.*, Competing protein-rna interaction networks control multiphase intracellular organization, *Cell* **181**, 306 (2020).
- [11] A. N. Semenov and M. Rubinstein, Thermoreversible gelation in solutions of associative polymers. 1. statics, *Macromolecules* **31**, 1373 (1998).
- [12] J.-M. Choi, A. S. Holehouse, and R. V. Pappu, Physical principles underlying the complex biology of intracellular phase transitions, *Annu. Rev. Biophys.* **49** (2020).
- [13] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comput. Phys.* **117**, 1 (1995).
- [14] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model, *PLoS Comput. Biol.* **14**, e1005941 (2018).
- [15] P. Langevin, Sur la théorie du mouvement brownien [on the theory of brownian motion], *C. R. Acad. Sci.* **146**, 530 (1908).
- [16] P. J. Flory, Thermodynamics of high polymer solutions, *J. Chem. Phys.* **10**, 51 (1942).
- [17] M. L. Huggins, Solutions of long chain compounds, *J. Chem. Phys.* **9**, 440 (1941).
- [18] B. Xu, *Protein Phase Separation In and Out of Cells*, Ph.D. thesis, Princeton University (2018).
- [19] P.-G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell university press, 1979).
- [20] C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, and M. K. Rosen, Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein, *Mol. Cell* **63**, 72 (2016).
- [21] Y. Hu and C. M. Niemeyer, From dna nanotechnology to material systems engineering, *Adv. Mater.* **31**, 1806294 (2019).
- [22] E. Bianchi, R. Blaak, and C. N. Likos, Patchy colloids: State of the art and perspectives, *Phys. Chem. Chem. Phys.* **13**, 6397 (2011).