Preserving condensate structure and composition by lowering sequence complexity

Amogh Sood[1] and Bin Zhang[1,*]

[1]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA*

E-mail: binz@mit.edu

Abstract

Biological condensates play a vital role in organizing cellular chemistry. They selectively partition biomolecules, preventing unwanted cross-talk and buffering against chemical noise. Intrinsically disordered proteins (IDPs) serve as primary components of these condensates due to their flexibility and ability to engage in multivalent, non-specific interactions, leading to spontaneous aggregation. Theoretical advancements are critical at connecting IDP sequences with condensate emergent properties to establish the so-called molecular grammar. We proposed an extension to the stickers and spacers model, incorporating non-specific pairwise interactions between spacers alongside specific interactions among stickers. Our investigation revealed that while spacer interactions contribute to phase separation and co-condensation, their non-specific nature leads to disorganized condensates. Specific sticker-sticker interactions drive the formation of condensates with well-defined structures and molecular composition. We discussed how evolutionary pressures might emerge to affect these interactions, leading to the prevalence of low complexity domains in IDP sequences. These domains suppress spurious interactions and facilitate the formation of biologically meaningful condensates.

Significance Statement

Biomolecular condensates serve as pivotal mechanisms in cellular organization, often characterized by an abundance of intrinsically disordered proteins (IDPs) that undergo frequent mutations in their sequences. Despite this, IDP sequences exhibit non-random patterns, yet the precise relationship between these sequences and the emergent properties of condensates remains unclear. To address this gap, we propose a molecular theory that delineates how various sequence features of IDPs contribute to the organization and composition of condensates. This theory not only sheds light on the evolution of IDPs but also elucidates the emergence of non-random sequence patterns as essential elements for the formation of functional condensates. Correspondingly, we posit that the prevalence of low-complexity regions within IDPs is a result of evolutionary selection.

Introduction

Biomolecular condensates serve as important mechanisms for the hierarchical and vectorial organization of chemistry within cells.[1–7] They are non-stoichiometric assemblies of biomolecules that can form via spontaneous or driven processes and exhibit characteristics of phase separation and percolation.[8–10] The chemical environment and physicochemical conditions within condensates are distinct from their surroundings,[11–18] enabling cells to selectively partition biomolecules, prevent unwanted cross-talk and interference between various biochemical pathways, and buffer against chemical noise.

In recent years, these condensates have gained significant prominence and attracted substantial research interest. They have been found to primarily form through the involvement of intrinsically disordered proteins (IDPs).[19–22] These unique proteins lack a fixed three-dimensional structure, enabling them to be highly flexible and engage in multivalent interactions. Multivalent interactions involve the binding of multiple partners simultaneously and are often nonspecific in nature, leading to the spontaneous aggregation and condensation

2

of various cellular components. Many research groups have attempted to establish what is known as the "molecular grammar",[23–34] which connects amino acid sequences with protein phase behaviors and the collective physical properties of condensates.

Establishing a theoretical framework is crucial for advancing our understanding of biological condensates. The "stickers and spacers" model, originally developed within the context of polymer gelation theory,[35–37] has gained popularity for modeling protein condensates.[1,2,25,38–40] In this model, biomolecules within condensates are envisioned as possessing two distinct functional components: "stickers" and "spacers". Stickers represent specific molecular domains or motifs with a high affinity for one another, facilitating interaction and bringing molecules into close proximity, thereby contributing to the condensate's cohesive and ordered structure. Conversely, spacers act as flexible linkers connecting the stickers, enabling the dynamic and transient nature of interactions within the condensate. This model offers a conceptual framework that assists in interpreting experimental observations regarding condensate stability and material properties.[41]

The simplicity of the stickers and spacers model, crucial for its theoretical elegance, encounters limitations when applied to realistic biomolecules. Specifically, when dealing with an IDP, the identification of stickers and spacers is often challenging.[1,3,42,43] Most amino acids undergo some level of interactions, and determining when to disregard these interactions to categorize them as spacers remains ambiguous. We propose that instead of entirely disregarding their interactions, it's essential to consider the interactions between spacer sequences and explore their impact on the physical properties of condensates. This exploration could offer insights into the evolutionary pressure of IDP sequences, arising from the interplay between sticker and spacer sequences. Moreover, such an investigation could enhance our understanding of real-world biological condensates, where spacers might deviate from ideal behavior but still participate in weak interactions.

We expanded the stickers and spacers model to include explicit non-specific pairwise interactions between spacers, alongside the specific interactions among sticker motifs. This

extension allowed us to systematically explore the intricate interplay between specific and non-specific interactions in determining the structural and compositional properties of condensates. Our investigation revealed that spacer interactions contribute to phase separation and the co-condensation of multiple molecules. However, the non-specific nature of these interactions results in disorganized condensates with undefined molecular compositions. In contrast, specific sticker-sticker interactions drive the formation of condensates with robust contacts and precise compositions. Subsequently, we discussed the implications of our theory for the evolution of protein sequences, asserting the existence of evolutionary constraints even on segments of protein sequences that interact non-specifically. These constraints ensure the functionality of condensates. This evolutionary pressure naturally favors the emergence of low complexity domains to suppress spurious interactions, facilitating the formation of biologically meaningful condensates.

Theory

Stickers and random spacers model

We present a generalized version of the stickers and spacers model to investigate the phase behaviors of associative polymers (Fig. 1). These polymers consist of N monomers, with f privileged monomers that exhibit specific attractive interactions, denoted by a strength of $-u_a$, leading to the formation of noncovalent, physical bonds. We refer to these privileged monomers as stickers. For simplicity, we consider stickers composed of a single monomer. However, it's important to note that in biological sequences, stickers may consist of multiple amino acids. The remaining monomers are designated as spacers.

In a departure from the traditional model, we consider the interaction energy between a pair of spacers, or a spacer-sticker pair, to be a random variable, ϵ_i, drawn from a normal distribution, $\mathcal{N}(\bar{\epsilon}, \Delta\epsilon^2)$, with mean $\bar{\epsilon}$ and variance $\Delta\epsilon^2$. The contribution to the total energy
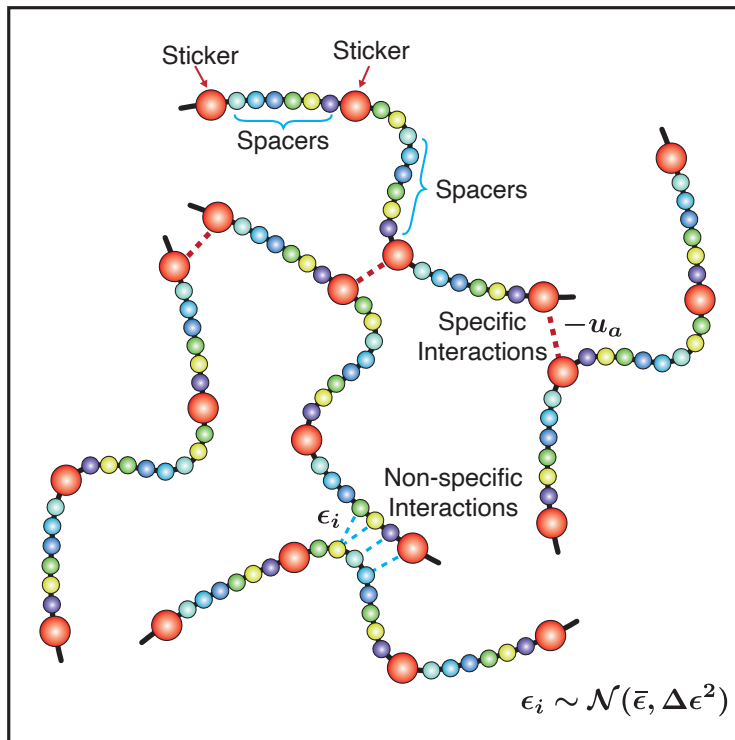
4

Figure 1: A schematic illustration of the stickers and random spacers model. Red spheres indicate stickers that interact specifically, and the strength for sticker-sticker interactions is a well-defined number, $-u_a$. We indicate the random spacers using shades of blue-green. These contribute non-specific interactions, and the pairwise interactions between adjacent pairs of spacers are chosen from a normal distribution, $\mathcal{N}(\bar{\epsilon}, \Delta\epsilon^2)$.

for a given configuration on the lattice from non-specific interactions is

$$-E = \sum_{i \in \text{n.n.}} \epsilon_i. \tag{1}$$

The summation denotes a sum over neighboring spacer-spacer pairs and non-bonded sticker and spacer pairs. Our use of random energy follows the tradition of protein folding theory,[44,45] allowing the derivation of expressions with a mean field theory[46] that are not specific to considerations of any particular sequence.

In the following, we explore two systems to investigate the influence of spacer interactions on the organization and compositions of condensates. Initially, we examine a homotypic system consisting of a single polymeric species encompassing both specifically interacting sticker

moieties and non-specifically interacting spacer residues. In this homotypic system, we formulate a mean-field free energy and investigate percolation coupled with phase separation. Our analysis reveals how the distribution of non-specific spacer interactions impacts the critical temperature, critical concentration, gel point, and degree of conversion. Subsequently, we extend this model to a heterotypic system involving two polymeric species, denoted as A and B. We highlight the necessity of finely tuning non-specific interactions to ensure robust composition in the dense phase for $A - B$ mixtures.

Phase behaviors of a single component system

We begin by examining a single-component system comprising n_p identical polymer chains on a lattice containing n sites. Let $\varphi \equiv n_p N/n$ represent the fraction of sites occupied by the polymer. The formation of $n_p m$ bonds occurs between $2n_p m$ stickers, and we define the degree of conversion as $p = \frac{2m}{f}$. Here, the f sticker segments are assumed to be uniformly distributed, effectively partitioning the chain into $f + 1$ segments. We denote the expected length of each individual segment as l. By symmetry, $(f + 1)l = N - f$. Consequently, $l = \frac{N-f}{f+1} \approx \frac{N}{f}$ for $N \gg f$ and $f \gg 1$.

The partition function for forming $n_p m$ bonds between $2n_p m$ stickers is,

$$\mathcal{Z} = \int \mathrm{d}E P(E)\Omega_{\mathrm{ref}}\Omega_{\mathrm{st}} \exp(\beta u_a n_p m) \exp(-\beta E). \tag{2}$$

Here, the probability distribution of the energy, $P(E)$, is Gaussian, since E is a sum of independent random variables following the Gaussian distribution (Eq. 1). The mean and variance of $P(E)$ are given by

$$-n^{-1}\overline{E}[p, \varphi] = \frac{(z - p/l)\varphi^2 \bar{\epsilon}}{2}, \tag{3a}$$

$$n^{-1}\Delta E^2[p, \varphi] = \frac{(z - p/l)\varphi^2 \Delta\epsilon^2}{2}. \tag{3b}$$

Here, z denotes the coordination number of the lattice. The term p/l appears to account for the fact that once a sticker forms a physical bond with another sticker, it is no longer available to interact with spacers. $\beta = (k_B T)^{-1}$ where k_B is the Boltzmann's constant. For convenience, we set $k_B = 1$ and report all parameters in natural units.

Ω_{ref} is the number of ways of placing the n_p polymers on the and is a standard result in polymer physics.[47] Its expression is given by,

$$\log(\Omega_{\text{ref}}) = -n\left[\frac{\varphi}{N}\log\left(\frac{\varphi}{Ne}\right) + (1-\varphi)\log(1-\varphi)\right]. \tag{4}$$

Ω_{st} corresponds to configurational entropy of forming sticker-sticker bonds. Following Semenov and Rubinstein[35], we derive the expression as

$$\Omega_{\text{st}} = \frac{(n_p f)!}{(n_p f - 2n_p m)!(n_p m)!2^{n_p m}}\left(\frac{z-2}{n}\right)^{n_p m}. \tag{5}$$

In deriving the above expression, we assumed that the chains are strongly overlapping ($\varphi >> \varphi_{\text{overlap}} \sim N^{-1/2}$) and inter-chain bonds between stickers dominate over intra-chain ones. Thus, to arrive at Ω_{st}, we first count the number of ways of choosing $2n_p m$ stickers out of $n_p f$ stickers to form bonds. We then multiply this number by the number of ways of pairing $2n_p m$ stickers together which is given by $(2n_p m - 1)!!$. Finally, we multiply this by the probability that all of the chosen stickers are neighboring each other. We also used the relation $(2k - 1)!! = \frac{2k!}{2^k k!}$ to further simplify the expression finally yielding Eq. 5.

The integration in Eq. 2 is readily performed, and we use $-\beta F = \log(\mathcal{Z})$ to obtain the free energy. The free energy must be minimised with respect to the degree of conversion, p, yielding the condition

$$\frac{p}{(1-p)^2} = \frac{(z-2)\varphi}{l}\exp\left(\beta u_a - \varphi\left(\beta\tilde{\epsilon} + \frac{\beta^2}{2}\Delta\epsilon^2\right)\right). \tag{6}$$

7

Remembering that $p \in [0, 1]$, we derive

$$p = 1 - \frac{\sqrt{\frac{4\varphi\lambda[\varphi]}{l} + 1} - 1}{\frac{2\varphi\lambda[\varphi]}{l}}, \tag{7}$$

where $\lambda[\varphi] \equiv (z - 2) \exp\left(\beta u_a - \varphi\left(\beta\tilde{\epsilon} + \frac{\beta^2 \Delta\epsilon^2}{2}\right)\right)$. We substitute the condition, Eq. 6, into the original expression for βF, to obtain the free energy density $\beta\mathcal{F} \equiv n^{-1}\beta F|_{\partial_p F=0}$ as

$$\beta\mathcal{F} = \frac{\varphi}{N} \log\left(\frac{\varphi}{Ne}\right) + (1 - \varphi)\log(1 - \varphi) + \frac{\varphi}{l}\left[\frac{p}{2} + \log(1 - p)\right] - \frac{z\varphi^2\beta\tilde{\epsilon}}{2} - \frac{z\varphi^2\beta^2\Delta\epsilon^2}{4} \tag{8}$$

Eq. 8 provides the starting point for deriving equilibrium properties of the system. For example, to determine the critical behaviour we compute the chemical potential, $\mu = \partial_\varphi\mathcal{F}$. The critical point is then the intersection of the nullclines $\partial_\varphi\mu = 0$ and $\partial_\varphi^2\mu = 0$. Additionally, we also consider the concentration of chains with all f stickers free, $\mathcal{C}_{\text{free}}$, defined as,

$$\mathcal{C}_{\text{free}} = \frac{\varphi}{N}(1 - p)^f. \tag{9}$$

The gel line is obtained by the condition $\partial_\varphi\mathcal{C}_{\text{free}} = 0$. We also compute φ_{dense}, the concentration in the dense phase, by using $\Pi \approx 0$, where Π is the osmotic pressure given by the relation,

$$\beta\Pi = \beta(\varphi\partial_\varphi\mathcal{F} - \mathcal{F}). \tag{10}$$

The concentration in the dilute phase (φ_{dilute}) can then be obtained by equating the the chemical potentials in the two phases.

Phase behaviors for a two component system

To further understand the influence of spacer interactions on condensate composition, we examine systems composed of two types of chains, denoted as A and B. Here, n_a and n_b

represent the number of chains, while N_a and N_b denote the degree of polymerization for chains A and B, respectively. Chain A comprises f_a stickers, whereas chain B consists of f_b stickers. We permit specific $A - B$ sticker-sticker interactions $(-u_{ab})$ while prohibiting self-interactions (i.e., $u_{aa} = u_{bb} = 0$). This assumption mirrors a common scenario involving two proteins with specific binding.

The partition function for the two component system takes the familiar form,

$$\mathcal{Z} = \int \mathrm{d}E P(E) \Omega_{\mathrm{ref}} \Omega_{\mathrm{st}} \exp(\beta k u_{ab}) \exp(-\beta E). \tag{11}$$

We once again assume the spacer-spacer and spacer-sticker interactions between $A - A$, $B - B$ and $A - B$ chains are each drawn from a normal distribution $\mathcal{N}(\bar{\epsilon}_{xx}, \Delta \epsilon_{xx}^2)$ where, $xx \in \{aa, bb, ab\}$. Correspondingly, $P(E)$ is Gaussian with mean and variance given by

$$-n^{-1}\bar{E}[p,\varphi] = \frac{1}{2}\left[(z - p/l)\,\varphi_a^2 \bar{\epsilon}_{aa} + \left(z - \frac{\varphi_a}{\varphi_b}\frac{p}{l}\right)\varphi_b^2 \bar{\epsilon}_{bb} + \left(2z\varphi_a\varphi_b - \frac{p}{l}\varphi_a(\varphi_a + \varphi_b)\right)\bar{\epsilon}_{ab}\right], \tag{12a}$$

$$n^{-1}\Delta E^2[p,\varphi] = \frac{1}{2}\left[(z - p/l)\varphi_a^2 \Delta\epsilon_{aa}^2 + \left(z - \frac{\varphi_a}{\varphi_b}\frac{p}{l}\right)\varphi_b^2 \Delta\epsilon_{bb}^2 + \Delta\epsilon_{ab}^2\left(2z\varphi_a\varphi_b - \frac{p}{l}\varphi_a(\varphi_a + \varphi_b)\right)\right]. \tag{12b}$$

The mixing entropy is given by,[47]

$$-n^{-1}\log(\Omega_{\mathrm{ref}}) = \frac{\varphi_a}{N_a}\log\left(\frac{\varphi_a}{N_a e}\right) + \frac{\varphi_b}{N_b}\log\left(\frac{\varphi_b}{N_b e}\right) + (1 - \varphi_a - \varphi_b)\log(1 - \varphi_a - \varphi_b). \tag{13}$$

We assume the existence of k pairs of $A - B$ sticker-sticker bonds in the system. Therefore, $\Omega_{\mathrm{st}} = \binom{n_a f_a}{k}\binom{n_b f_b}{k}k!\left(\frac{z-2}{n}\right)^k$. In order to define the degree of conversion we note that the maximum possible number of bonds is $\min\{n_a f_a, n_b f_b\}$, and without loss of generality we can assume it is chain A.[39] Therefore, we define $p_{ab} \equiv k/n_a f_a$. For algebraic convenience, we assume $\frac{N_a}{f_a} \equiv l_a \approx \frac{N_b}{f_b} \equiv l_b = l$. Consequently, the configurational entropy of the sticker-

sticker bonds can be readily computed as:

$$-n^{-1}\log(\Omega_{\text{st}}) = -\frac{\varphi_a}{N_a}p_{ab}f_a \log\left(\frac{(z-2)\varphi_b f_b}{eN_b}\right) + \frac{\varphi_a}{N_a}f_a\left(p_{ab}\log p_{ab} + (1-p_{ab})\log(1-p_{ab})\right)$$

(14)

$$+ \frac{\varphi_b}{N_b}f_b\left(1 - p_{ab}\frac{(\varphi_a/N_a)f_a}{(\varphi_b/N_b)f_b}\right)\log\left(1 - p_{ab}\frac{(\varphi_a/N_a)f_a}{(\varphi_b/N_b)f_b}\right).$$

Here we have used the Stirling's approximation (i.e, for $n >> 1$, $\log(n!) \approx n\log n - n + \mathcal{O}(\log n)$) and the following helpful identities, $\frac{kn_af_a}{n_bf_bn_af_a} = p_{ab}\frac{(\varphi_a/N_a)f_a}{(\varphi_b/N_b)f_b}$, $\varphi_t = N_tn_t/n$, $k/n = p_{ab}n_af_a/n = p_{ab}f_a\varphi_a/N_a$.

We perform the integration in Eq. 11 and use $-\beta F = \log(\mathcal{Z})$ to obtain the free energy. We then minimise it with respect to the degree of conversion, p_{ab}, to yield,

$$\frac{p_{ab}}{(1-p_{ab})\left(1 - p_{ab}\frac{\varphi_a}{\varphi_b}\right)} = \frac{(z-2)\varphi_b}{l}\exp\left(\beta u_{ab} - \frac{\beta}{2}\left[\bar{\epsilon}_{aa}\varphi_a + \bar{\epsilon}_{bb}\varphi_b + \bar{\epsilon}_{ab}(\varphi_a + \varphi_b)\right]\right.$$

(15)

$$\left. - \frac{\beta^2}{4}\left[\Delta\epsilon_{aa}^2\varphi_a + \Delta\epsilon_{bb}^2\varphi_b + \Delta\epsilon_{ab}^2(\varphi_a + \varphi_b)\right]\right).$$

We substitute the above expression into the original expression for βF to obtain the free energy density $\beta\mathcal{F} \equiv n^{-1}\beta F|_{\partial_p F=0}$

$$\beta\mathcal{F} = \frac{\varphi_a}{N_a}\log\left(\frac{\varphi_a}{N_ae}\right) + \frac{\varphi_b}{N_b}\log\left(\frac{\varphi_b}{N_be}\right) + (1 - \varphi_a - \varphi_b)\log(1 - \varphi_a - \varphi_b)$$

(16)

$$+ \frac{\varphi_a}{l}p + \frac{\varphi_a}{l}\log(1-p) + \frac{\varphi_b}{l}\log\left(1 - p\frac{\varphi_a}{\varphi_b}\right)$$

$$- \beta z\varphi_a^2\frac{\bar{\epsilon}_{aa}}{2} - \beta z\varphi_b^2\frac{\bar{\epsilon}_{bb}}{2} - \beta z\varphi_a\varphi_b\bar{\epsilon}_{ab}$$

$$- \frac{\beta^2}{4}z\varphi_a^2\Delta\epsilon_{aa}^2 - \frac{\beta^2}{4}z\varphi_b^2\Delta\epsilon_{bb}^2 - \frac{\beta^2}{2}z\varphi_a\varphi_b\Delta\epsilon_{ab}^2.$$

One can recover the expressions derived for the homotypic case from the above equation by replacing $\varphi_a = \varphi_b \equiv \varphi$, up to a $\varphi \to 2\varphi$ transformation.

For the two component system, the spinodal can be obtained from the nullcline of the

determinant of the hessian matrix of $\mathcal{F}$, namely, $\det(\mathcal{H}_{\mathcal{F}}) = 0$. We determine the binodal by solving the following system of equations

$$\mu_a^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) = \mu_a^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right), \tag{17}$$
$$\mu_b^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) = \mu_b^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right),$$
$$\Pi^{(0)}\left(\varphi_a^{(0)}, \varphi_b^{(0)}\right) = \Pi^{(1)}\left(\varphi_a^{(1)}, \varphi_b^{(1)}\right).$$

where, $\mu_k = \partial_{\varphi_k}\mathcal{F}$ and $\Pi = \sum_k \varphi_k(\partial_{\varphi_k}\mathcal{F}) - \mathcal{F}$. These equations ensure that the chemical potentials and the osmotic pressure are identical between coexisting phases. We solve the above equations by numerically finding roots to the the function $\mathcal{B}(\{\varphi_x^k\}) \equiv (\mu_a^{(0)} - \mu_a^{(1)}, \mu_b^{(0)} - \mu_b^{(1)}, \Pi^{(0)} - \Pi^{(1)}, 0) = 0$ using standard python libraries. More details on the numerical solutions are provided in the Supporting Information.

Results

Non-specific spacer interactions facilitate phase separation

As described in the Theory section, we introduce a novel model designed to investigate the phase behavior of condensate-forming proteins (Fig. 1). Expanding upon the stickers and spacers model,[1,2,35–37] we identify specific chemical groups within proteins that exhibit robust interactions as stickers. However, contrary to prevalent approaches in contemporary literature, we account for heterogeneity in interactions among spacers. For simplicity, we assume that the strength of spacer interactions follows a normal distribution, $\mathcal{N}(\bar{\epsilon}, \Delta\epsilon^2)$, characterized by a mean of $\bar{\epsilon}$ and a variance of $\Delta\epsilon^2$. This variability accommodates the diversity of amino acids, resulting in weak yet significant interactions.[44,48,49] We refer to this model as the "stickers and random spacers" model, or STARS.

We compute the complete phase diagram of the STARS model to investigate the impact of nonspecific spacer interactions on condensate behaviors. For a protein solution with identical

molecules, we derive its free energy density $\mathcal{F}$ (Eq. 8). From this, we calculate the chemical potentials as $\mu = \partial_\varphi \mathcal{F}$, where φ represents the polymer volume fraction. The binodal is obtained by equating the chemical potentials of each component in coexisting phases, while the spinodal is obtained as the nullcline of $\partial_\varphi \mu = 0$. These lines delineate the boundaries among the unstable, meta-stable, and stable regions of the phase diagram. The critical point (T_c, φ_c) denotes the temperature and concentration at which the solution first becomes unstable, aiding our understanding of the macromolecular solubility of the system.

We present the phase diagram for the STARS model in Fig. 2a. The sticker interactions are set as $u_a = 5k_B T$ and $\Delta\epsilon = 2k_B T$ with $\bar{\epsilon} = 0$. Additional model parameters are included in the caption. For comparison, we compute the phase diagram for a stickers and spacers model with similar parameters (Fig. 2b). Similar to the stickers and spacers model,[50,51] the STARS model exhibits both a percolation (gelation) transition and phase separation. At high temperatures with $T > T_c$, the system can undergo a sol-gel transition without phase separation. Decreasing the temperature further leads to phase separation coupled with gelation.

The inclusion of spacer interactions also leads to quantitative alterations in the phase diagram. In Fig. 2c, we observe a sharp rise in the critical temperature T_c with increasing $\Delta\epsilon$, thereby promoting phase separation. The critical concentration, φ_c, follows a similar trend, albeit exhibiting a slight increase at larger values of $\Delta\epsilon$ before quickly plateauing. Notably, these trends remain consistent regardless of other system parameters (Figs. S1 and S2). Spacer interactions, while chosen from a normal distribution with a zero mean, can produce contacts with negative energies that are favored by the Boltzmann factor, as defined in the partition function (Eq. 2). Consequently, these contacts stabilize the condensed phase, thereby facilitating phase separation.
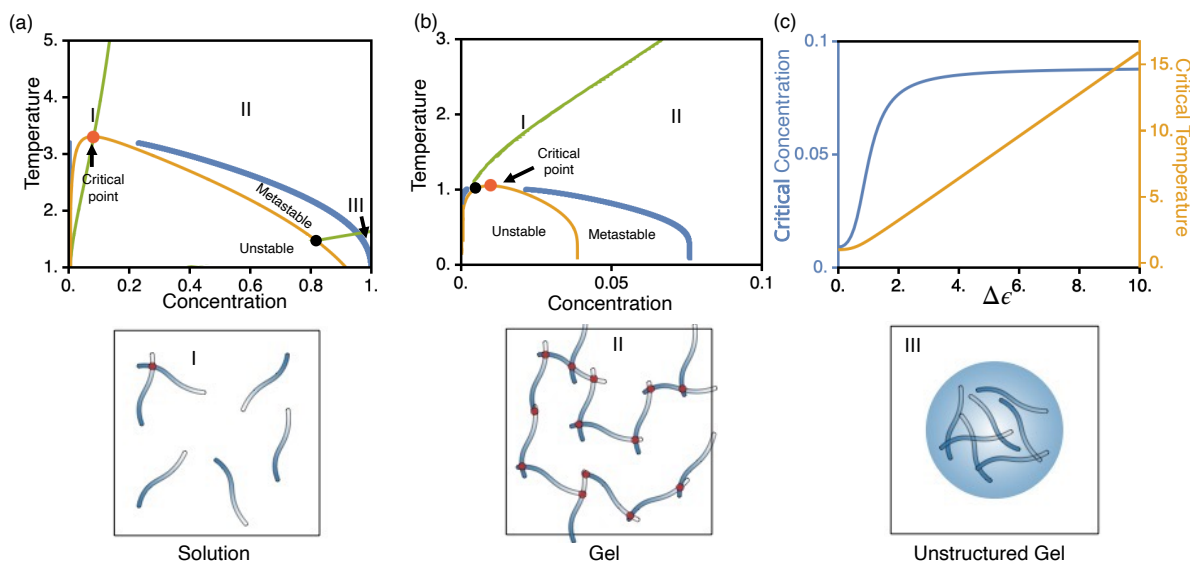
Figure 2: Phase behavior of the stickers and random spacers model. (a, b) Phase diagrams for the STARS model with $\Delta\epsilon = 2$ (a) and the stickers and spacers model with $\Delta\epsilon = 0$ (b). We plot the spinodal (orange) and binodal (blue) curves that demarcate the boundaries between the stable, meta-stable and unstable regions in the phase diagram. The critical point is highlighted in red. For the STARS model, the gel line (green) crosses the binodal twice, partitioning the stable phase into three regions. Illustrative configurations for the three regions corresponding to the solution phase, the gel phase, and the unstructured gel are shown in the bottom. (c) Dependence of the critical point on the strength of non-specific interaction among spacers, $\Delta\epsilon$. We set $u_a = 5, \bar{\epsilon} = 0, l = 10, N = 100$, and $z = 6$ when computing the phase diagrams.

Non-specific spacer interactions modulate condensate organization

Fig. 2c indicates that diversifying IDP sequences might increase the variation in their interaction energies, potentially contributing positively to phase separation. However, it seems counter-intuitive that many IDPs, known to participate in condensate formation, would evolve sequences featuring low complexity regions.[52–55] These regions often contain amino acid repeats with suppressed sequence diversification and interaction patterns, leading to smaller $\Delta\epsilon$ values.

We propose that spacer-spacer interactions could negatively impact condensate structures arising from physical crosslinks between stickers. These crosslinks differentiate biomolecular condensates from simple liquids, generating heterogeneous environments with specific protein-protein interfaces. Robust contacts inside condensates could facilitate fast process-

ing of intermediates, such as in metabolic channeling.[56] Additionally, physical crosslinks influence the viscoelastic and rheological properties crucial for condensate function.[1,2,50] For instance, differences in viscoelasticity between nucleolar core and outer layers strongly affect ribosomal assembly.[57]

To characterize condensate structure, we introduce the degree of conversion, p, measuring the fraction of bonded stickers. Additionally, we study the dependence of p on $\Delta\epsilon$ at a constant temperature $T = 1$. For the parameters outlined in the caption of Fig. 3, the system undergoes phase separation coupled with gelation at this temperature. We denote the polymer volume fraction of the dense phase as φ_{dense} and use it to determine p from Eq. 7. As depicted in Fig. 3a, the degree of conversion initially increases slightly, reaching a maximum before decreasing to zero with increasing $\Delta\epsilon$. This suggests that spacer-spacer interactions not only influence the overall macromolecular solubility but also mediate network properties of the stickers.

Moreover, spacer-spacer interactions can qualitatively change the phase diagram. We study $\mathcal{C}_{\text{free}}$, measuring the concentration of polymer chains with all stickers free (i.e., not bound to other stickers, Eq. 9). As per established theories, for $\Delta\epsilon = 0$, $\mathcal{C}_{\text{free}}$ first increases with concentration (φ) in the pre-gel regime, achieving a maximum at the gel point (Fig. 3b, blue). Subsequently, in the post-gel regime, $\mathcal{C}_{\text{free}}$ monotonically decreases to zero as an increasing number of stickers form crosslinks.[35] Consequently, the maximum of $\mathcal{C}_{\text{free}}$ often defines the gel line ($\partial_\varphi \mathcal{C}_{\text{free}} = 0$), depicted in Fig. 2 as the green curves.

Strikingly, we note that the monotonic decrease in $\mathcal{C}_{\text{free}}$ as a function of increasing φ in the post-gel regime for $\Delta\epsilon \neq 0$ no longer holds true. As shown by the orange line in Fig. 3b, $\mathcal{C}_{\text{free}}$ decreases to a minimum before increasing again at higher concentrations. Once again, these observations remain qualitatively insensitive to the exact choice of system parameters (Figs S1 and S2).

This non-monotonic behavior results in the first derivative of $\mathcal{C}_{\text{free}}$ crossing the binodal twice in Fig. 2a, partitioning the stable regions into three. Regions I and II resemble the
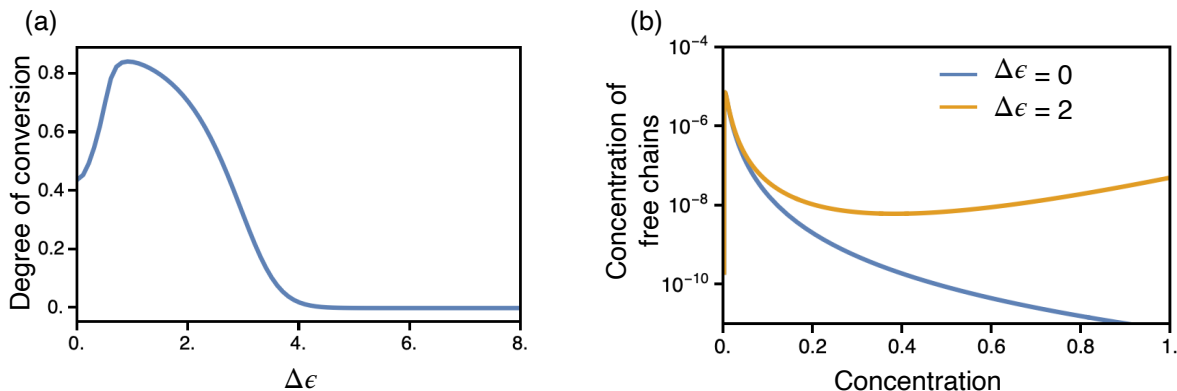
14

Figure 3: Impact of non-specific interactions among spacer interactions on the network properties of condensates. (a) The degree of conversion evaluated at φ_{dense} (the concentration of polymers in the dense phase) shows a moderate increase followed by a subsequent decrease as we widen the spread of the spacer-spacer interaction energy distribution by increasing $\Delta\epsilon$. (b) The concentration of free chains (with all stickers free) increases with concentration in the pre-gel regime and reaches a maximum at the gel-point. It monotonically decreases in the post-gel regime when $\Delta\epsilon = 0$, but exhibits non-monotonic behavior for $\Delta\epsilon \neq 0$. We set $u_a = 5$, $T = 1$, $\bar{\epsilon} = 0$, $l = 10$, $N = 100$, and $z = 6$.

typical solution and gel phase found in the stickers and spacers model (see Fig. 2b). However, in region III, C_{free} increases again due to a drop in sticker-sticker crosslinks. This transition leads to a less structured gel due to the competition between spacer interactions and the formation of sticker-sticker bonds.

Non-specific spacer interactions modulate condensate composition

Until now, our focus has been on single-component systems. However, biomolecular condensates within cells commonly consist of multiple molecules.[8,10,58–61] These condensates possess well-defined compositions, allowing only certain molecules to selectively partition into them.[62] Maintaining such a specific composition is crucial for their function in both physiological contexts and potential therapeutic applications.[13,63–65] To explore further, we investigated whether spacer interactions influence condensate composition in a minimal system comprising two components.

We aimed to model a scenario where A serves as the host, incorporating a polymeric guest B. Consequently, we focused on homotypic $A - A$ interactions ($\bar{\epsilon}_{aa} = 1$, $\Delta\epsilon_{aa} = 0$) to

15

facilitate the demixing of polymer A from B.[16] To simplify, our attention was directed toward sequence diversification in protein B. This diversification results in both self-interactions and cross-interactions with A. Hence, we set both ϵ_{bb} and ϵ_{ab} as random variables with variance $\Delta\epsilon_{bb} \equiv \Delta\epsilon_{ab}$ and zero-mean (for convenience).

For this two-component system, we derived the free energy expression (Eq. 16) to study its phase behaviors. Fig. 4 illustrates the phase behavior at constant temperature ($T = 1$) as a function of composition in the $\varphi_a - \varphi_b$ plane, considering varying strengths of specific (u_{ab}) and nonspecific interactions ($\Delta\epsilon_{bb}, \Delta\epsilon_{ab}$). The orange lines denote the spinodal, while the blue lines represent the binodal. Additionally, the gray lines represent tie lines connecting coexisting phases.

We observed that the inclusion of nonspecific spacer-spacer interactions facilitates the co-condensation of A and B. Specifically, in scenarios where sticker-sticker interactions are sufficiently weak (only $\bar{\epsilon}_{aa} = 1$ without any $A - B$ or $B - B$ spacer interactions), the negative slope of tie lines in Fig. 4a suggests a demixing behavior. Stable phases at the top left and bottom right exhibit enrichment in only one component without a balanced presence of both. However, upon introducing nonspecific spacer interactions in Fig. 4b, tie line slopes become positive, indicating stable phases enriched or depleted in both components simultaneously– signifying co-condensation. A similar behavior was observed by Deviri and Safran[16] using the Flory-Huggins theory. The observed splay in the tie lines in Fig. 4b indicates that slight concentration fluctuations in the dilute phase can lead to significantly distinct compositions in the dense phase.

Furthermore, enhancing interaction strength among stickers (while maintaining $\Delta\epsilon_{ab} = \Delta\epsilon_{bb} = 0$) also facilitates co-condensation. As shown in Fig. 4c, a new stable phase emerges within the diagram's center. Tie lines connect this stable phase with regions depleted in one component at the top left and lower right corners. In this new stable region, $p_{ab} \approx 1$. Additionally, the stable co-condensation phase has a narrower range of possible concentrations for A and B. This narrow concentration range is unlike the wider range supported in Fig. 4b
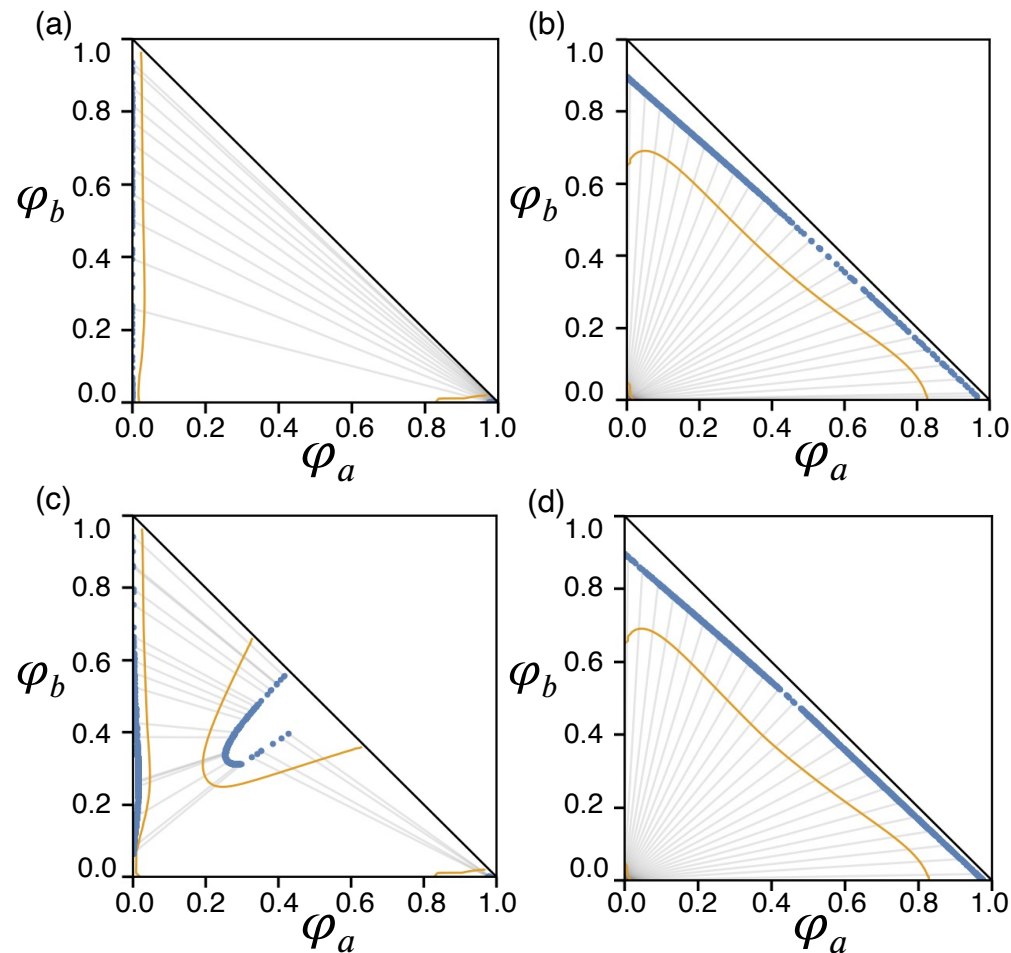
16

Figure 4: Phase diagrams showing the spinodal line (orange, solid) and binodal line (blue, dots) for a two component system with (a) $u_{ab} = 2, \Delta\epsilon_{bb} = \Delta\epsilon_{ab} = 0$, (b) $u_{ab} = 2, \Delta\epsilon_{bb} = \Delta\epsilon_{ab} = 1.1$, (c) $u_{ab} = 5, \Delta\epsilon_{bb} = \Delta\epsilon_{ab} = 0$, and (d) $u_{ab} = 5, \Delta\epsilon_{bb} = \Delta\epsilon_{ab} = 1.1$. The tie-lines (light grey, solid) connects co-existing points on the binodal. We set $N_a = N_b = 10, l = 2, z = 6, \bar{\epsilon}_{aa} = 1, \Delta\epsilon_{aa} = 0$, and $\bar{\epsilon}_{bb} = \bar{\epsilon}_{ab} = 0$ in all systems.

and d, where a broader spectrum of mixing ratios exists in the stable condensed phase. In our current setup, the stable phase in Fig. 4c is centered around $(0.5, 0.5)$ due to symmetric sticker distributions in A and B. Altering the interaction strengths among stickers or the relative abundance of A or B stickers can aid in promoting phases with varied compositions. A detailed survey of titrating $\Delta\epsilon_{bb}, \Delta\epsilon_{ab}$ on the phase diagram in both the weak $(u_{ab} = 2)$ and strong $(u_{ab} = 5)$ sticker-sticker interaction regimes is shown in the Supporting Information (Figs S3 and S4).

Thus, while both sticker-sticker and spacer interactions induce condensation of multiple

components, sticker interactions robustly generate condensates with more defined molecular compositions and structures. This property can be advantageous in biological systems requiring precise rationing of different species to optimize efficiency in specific chemical processes. These observations remain qualitatively insensitive to the exact choice of system parameters (Figs S5 and S6).

Conclusions and Discussion

We have introduced a new theory aimed at establishing connections between the emergent physical properties of condensates and protein sequences. Specifically, by extending from the stickers and spacers model, we investigate how the interplay between specific and non-specific interactions can influence both the structural integrity and compositional specificity of biomolecular condensates. Specific interactions are defined as those exclusively formed between a pair of stickers due to their chemical selectivity, while non-specific interactions can occur among spacers and between spacers and non-bonded stickers. We analytically solved the new stickers and random spacers model to assess its phase behaviors.

Our findings demonstrate that non-specific spacer interactions have the capacity to promote phase separation and the condensation of multi-component systems. However, these non-specific interactions fail to generate condensates with robust organizational structures necessary for fine-tuning material properties and molecular compositions. Conversely, specific interactions show a similar ability to promote phase separation while establishing well-defined interaction networks within the condensate, resulting in precise compositions.

Revisiting the definition of stickers and spacers.

While it's acknowledged that stickers and spacers are context-dependent, it's common to identify potential stickers by pinpointing residues known to have the most pronounced impact on saturation concentrations.[3,66,67] However, our study suggests that relying solely on

18

critical behavior as a criterion for identifying potential stickers might be less robust. This is because spacers, characterized by non-specific interactions, can also influence both the critical and percolation behavior of the system. Alternatively, from a functional standpoint, we propose that stickers can also be identified as chemical groups that facilitate protein-protein interactions through specific binding. According to this definition, natural candidates for stickers are molecular recognition features (MoRFs) or short linear motifs (SLiMs). These consist of short stretches of adjacent amino acids essential for molecular recognition and protein binding.[68–70] Regions undergoing significant mutations leading to higher levels of binding promiscuity, however, should be classified as spacers. Identifying stickers and spacers based on interaction specificity helps in understanding the sequence features of IDPs and their evolutionary conservation, as we discuss below.

Evolutionary pressure and the rise of low complexity.

Our study offers insight into the evolution of IDP sequences. IDPs are recognized for their high evolvability, characterized by rapid mutation rates.[71,72] Unlike globular proteins that require well-defined 3D structures, IDPs theoretically possess a broader sequence space for exploration.[55] However, IDPs do not explore all possible unfoldable sequences. Instead, they adopt simplified sequences enriched with low complexity domains, utilizing only a subset of amino acids. Furthermore, despite their poor conservation in alignments, recent studies have revealed that orthologous IDPs share many conserved molecular features,[73–75] indicating non-randomness and suggesting evolutionary constraints that favor functionally fit sequence patterns.

Yet, the mere formation of biological condensates cannot explain the conserved features in IDP sequences. As discussed in the main text, the diversification of spacer sequences can effectively reduce saturation concentrations and facilitate phase separation. Therefore, to just optimize condensate stability, IDPs would not heavily favor low complexity domains that suppress interactions.

We propose that the formation of condensates with robust structures and compositions serves as the main pressure for IDP evolution. This pressure strengthens interactions among stickers while concurrently suppressing promiscuous interactions among spacers, leading to the prevalence of low complexity domains in IDPs. Lowering sequence complexity limits the range of possible amino acids within the sequence, naturally reducing the complexity and fluctuation of unwanted interactions. Additionally, for sequences of identical amino acid composition, interaction complexity can be further reduced by creating repeating, alternating aromatic residue sequences, such as FG-repeats in nucleoporins or YG-repeats in FUS-LCD.[3,76,77] These alternating sequences are known to exhibit weaker interactions compared to sequences where amino acids of the same type are clustered together.[78]

The suggested evolutionary pressure inherently drives the preservation of stickers to maintain condensate compositional specificity. Molecular recognition features and short linear motifs, acknowledged for their evolutionary conservation, support this notion.[73] Conversely, spacers encounter an evolutionary push to diminish their complexity. This pressure doesn't favor any particular sequence, hence contributing to their high mutation rate. Nevertheless, mutations resulting in excessively strong interactions, even when localized within the low complexity domains (i.e., spacers), can disrupt normal function as well.[79,80]

Acknowledgments

Competing interests: Authors declare that they have no competing interests.

Code and Data Availability

Data presented in this study is available upon reasonable request to the corresponding author.

References

(1) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual Review of Biophysics* **2020**, *49*, 107–133.

(2) Pappu, R. V.; Cohen, S. R.; Dar, F.; Farag, M.; Kar, M. Phase Transitions of Associative Biomacromolecules. *Chemical Reviews* **2023**, *123*, 8945–8987.

(3) Ginell, G. M.; Holehouse, A. S. *Phase-Separated Biomolecular Condensates*; Springer US, 2022; p 95–116.

(4) Latham, A. P.; Zhang, B. Molecular determinants for the layering and coarsening of biological condensates. *Aggregate* **2022**, *3*, e306.

(5) Berry, J.; Brangwynne, C. P.; Haataja, M. Physical principles of intracellular organization via active and passive phase transitions. *Reports on Progress in Physics* **2018**, *81*, 046601.

(6) Uversky, V. N. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18–30.

(7) Woodruff, J. B.; Hyman, A. A.; Boke, E. Organization and Function of Non-dynamic Biomolecular Condensates. *Trends Biochem. Sci.* **2018**, *43*, 81–94.

(8) Sabari, B. R.; Dall'Agnese, A.; Young, R. A. Biomolecular Condensates in the Nucleus. *Trends in Biochemical Sciences* **2020**, *45*, 961–977.

(9) Sanulli, S.; J. Narlikar, G. Liquid-like interactions in heterochromatin: Implications for mechanism and regulation. *Current Opinion in Cell Biology* **2020**, *64*, 90–96, Cell Nucleus.

(10) Banani, S. F.; Lee, H. O.; Hyman, A. A.; Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology* **2017**, *18*, 285–298.

(11) Hnisz, D.; Shrinivas, K.; Young, R. A.; Chakraborty, A. K.; Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **2017**, *169*, 13–23.

(12) 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.

(13) Kilgore, H. R.; Young, R. A. Learning the chemical grammar of biomolecular condensates. *Nature Chemical Biology* **2022**, *18*, 1298–1306.

(14) 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.

(15) Lyon, A. S.; Peeples, W. B.; Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nature Reviews Molecular Cell Biology* **2020**, *22*, 215–235.

(16) Deviri, D.; Safran, S. A. Physical theory of biological noise buffering by multicomponent phase separation. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2100099118.

(17) Ye, S.; Latham, A. P.; Tang, Y.; Hsiung, C.-H.; Chen, J.; Luo, F.; Liu, Y.; Zhang, B.; Zhang, X. Micropolarity governs the structural organization of biomolecular condensates. *Nat. Chem. Biol.* **2023**,

(18) Latham, A. P.; Zhu, L.; Sharon, D. A.; Ye, S.; Willard, A. P.; Zhang, X.; Zhang, B.

Frustrated Microphase Separation Produces Interfacial Environment within Biological Condensates. *bioRxiv* **2023**,

(19) Borcherds, W.; Bremer, A.; Borgia, M. B.; Mittag, T. How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? *Current Opinion in Structural Biology* **2021**, *67*, 41–50.

(20) Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annual Review of Cell and Developmental Biology* **2014**, *30*, 39–58.

(21) Mittag, T.; Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Molecular Cell* **2022**, *82*, 2201–2214.

(22) Dignon, G. L.; Best, R. B.; Mittal, J. Biomolecular Phase Separation: From Molecular Driving Forces to Macroscopic Properties. *Annual Review of Physical Chemistry* **2020**, *71*, 53–75.

(23) Schuster, B. S.; Regy, R. M.; Dolan, E. M.; Kanchi Ranganath, A.; Jovic, N.; Khare, S. D.; Shi, Z.; Mittal, J. Biomolecular Condensates: Sequence Determinants of Phase Separation, Microstructural Organization, Enzymatic Activity, and Material Properties. *J. Phys. Chem. B* **2021**, *125*, 3441–3451.

(24) 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*, 1–31.

(25) Wang, J.; Choi, J. M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 688–699.e16.

(26) Bhandari, K.; Cotten, M. A.; Kim, J.; Rosen, M. K.; Schmit, J. D. Structure-Function Properties in Disordered Condensates. *J. Phys. Chem. B* **2021**, *125*, 467–476.

(27) Pak, C. W.; Kosno, M.; Holehouse, A. S.; Padrick, S. B.; Mittal, A.; Ali, R.; Yunus, A. A.; Liu, D. R.; Pappu, R. V.; Rosen, M. K. Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol. Cell* **2016**, *63*, 72–85.

(28) 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.

(29) Ranganathan, S.; Shakhnovich, E. I. Dynamic metastable long-living droplets formed by sticker-spacer proteins. *Elife* **2020**, *9*, 1–25.

(30) Jacobs, W. M.; Frenkel, D. Phase Transitions in Biological Systems with Many Components. *Biophysical Journal* **2017**, *112*, 683–691.

(31) Shrinivas, K.; Brenner, M. P. Phase separation in fluids with many interacting components. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2108551118.

(32) Sanders, D. W. et al. Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell* **2020**, *181*, 306–324.e28.

(33) 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. Commun.* **2021**, *12*.

(34) Saar, K. L.; Morgunov, A. S.; Qi, R.; Arter, W. E.; Krainer, G.; Lee, A. A.; Knowles, T. P. Learning the molecular grammar of protein condensates from sequence determinants and embeddings. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*.

(35) Semenov, A. N.; Rubinstein, M. Thermoreversible Gelation in Solutions of Associative Polymers. 1. Statics. *Macromolecules* **1998**, *31*, 1373–1385.

(36) Dobrynin, A. V. Phase Diagram of Solutions of Associative Polymers. *Macromolecules* **2004**, *37*, 3881–3893.

(37) Tanaka, F. Theory of Thermoreversible Gelation. *Macromolecules* **1989**, *22*, 1988–1994.

(38) Choi, J.-M.; Hyman, A. A.; Pappu, R. V. Generalized models for bond percolation transitions of associative polymers. *Phys. Rev. E* **2020**, *102*, 042403.

(39) Prusty, D.; Pryamitsyn, V.; Olvera de la Cruz, M. Thermodynamics of Associative Polymer Blends. *Macromolecules* **2018**, *51*, 5918–5932.

(40) Choi, J.-M.; Dar, F.; Pappu, R. V. LASSI: A lattice model for simulating phase transitions of multivalent proteins. *PLOS Computational Biology* **2019**, *15*, e1007028.

(41) Mittag, T.; Pappu, R. V. A conceptual framework for understanding phase separation and addressing open questions and challenges. *Mol. Cell* **2022**, *82*, 2201–2214.

(42) Orlando, G.; Raimondi, D.; Tabaro, F.; Codicè, F.; Moreau, Y.; Vranken, W. F. Computational identification of prion-like RNA-binding proteins that form liquid phase-separated condensates. *Bioinformatics* **2019**, *35*, 4617–4623.

(43) Vernon, R. M.; Chong, P. A.; Tsang, B.; Kim, T. H.; Bah, A.; Farber, P.; Lin, H.; Forman-Kay, J. D. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **2018**, *7*.

(44) Bryngelson, J. D.; Wolynes, P. G. Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences* **1987**, *84*, 7524–7528.

(45) Pande, V.; Grosberg, A.; Tanaka, T. Statistical mechanics of simple models of protein folding and design. *Biophysical Journal* **1997**, *73*, 3192–3210.

(46) Derrida, B. Random-Energy Model: Limit of a Family of Disordered Models. *Physical Review Letters* **1980**, *45*, 79–82.

(47) Doi, M. *Introduction to Polymer Physics*, paperback ed.; Oxford University Press, 1990; p 136.

(48) Latham, A. P.; Zhang, B. Maximum Entropy Optimized Force Field for Intrinsically Disordered Proteins. *J. Chem. Theory Comput.* **2020**, *16*, 773–781.

(49) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Sci.* **2021**, *50*, pro.4094.

(50) Rubinstein, M.; Dobrynin, A. V. Solutions of associative polymers. *Trends in Polymer Science* **1997**, *5*, 181–186.

(51) Tanaka, F. *Polymer Physics*; Cambridge University Press, 2011.

(52) Van Der Lee, R. et al. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **2014**, *114*, 6589–6631.

(53) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **2001**, *42*, 38–48.

(54) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.

(55) Uversky, V. N. Intrinsically disordered proteins and their ”Mysterious” (meta)physics. *Front. Phys.* **2019**, *7*, 8–23.

(56) Pareek, V.; Sha, Z.; He, J.; Wingreen, N. S.; Benkovic, S. J. Metabolic channeling: predictions, deductions, and evidence. *Mol. Cell* **2021**, *81*, 3775–3785.

(57) Mitrea, D. M.; Cika, J. A.; Stanley, C. B.; Nourse, A.; Onuchic, P. L.; Banerjee, P. R.; Phillips, A. H.; Park, C.-G.; Deniz, A. A.; Kriwacki, R. W. Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. *Nature communications* **2018**, *9*, 842.

(58) Latham, A. P.; Zhang, B. On the stability and layered organization of protein-DNA condensates. *Biophys. J.* **2022**, *121*, 1727–1737.

(59) Rostam, N.; Ghosh, S.; Chow, C. F. W.; Hadarovich, A.; Landerer, C.; Ghosh, R.; Moon, H.; Hersemann, L.; Mitrea, D. M.; Klein, I. A.; Hyman, A. A.; Toth-Petroczy, A. CD-CODE: crowdsourcing condensate database and encyclopedia. *Nat. Methods* **2023**,

(60) Leicher, R.; Osunsade, A.; Chua, G. N. L.; Faulkner, S. C.; Latham, A. P.; Watters, J. W.; Nguyen, T.; Beckwitt, E. C.; Christodoulou-Rubalcava, S.; Young, P. G.; Zhang, B.; David, Y.; Liu, S. Single-stranded nucleic acid binding and coacervation by linker histone H1. *Nat. Struct. Mol. Biol.* **2022**, *29*, 463–471.

(61) Lafontaine, D. L. J.; Riback, J. A.; Bascetin, R.; Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 165–182.

(62) Lyons, H.; Veettil, 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.e28.

(63) Thody, S. A.; Clements, H. D.; Baniasadi, H.; Lyon, A. S.; Sigman, M. S.; Rosen, M. K. Small Molecule Properties Define Partitioning into Biomolecular Condensates. *bioRxiv* **2022**,

(64) Kilgore, H. R.; Mikhael, P. G.; Overholt, K. J.; Boija, A.; Hannett, N. M.; Dongen, C. V.; Lee, T. I.; Chang, Y.-T.; Barzilay, R.; Young, R. A. Distinct chemical environments in biomolecular condensates. *Nature Chemical Biology* **2023**,

(65) Klein, I. A.; Boija, A.; Afeyan, L. K.; Hawken, S. W.; Fan, M.; Dall'Agnese, A.; Oksuz, O.; Henninger, J. E.; Shrinivas, K.; Sabari, B. R., et al. Partitioning of cancer therapeutics in nuclear condensates. *Science* **2020**, *368*, 1386–1392.

(66) 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 (80-.).* **2020**, *367*, 694–699.

(67) Bremer, A.; Farag, M.; Borcherds, W. M.; Peran, I.; Martin, E. W.; Pappu, R. V.; Mittag, T. Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains. *Nat. Chem.* **2022**, *14*, 196–207.

(68) Morris, O. M.; Torpey, J. H.; Isaacson, R. L. Intrinsically disordered proteins: Modes of binding with emphasis on disordered domains. *Open Biol.* **2021**, *11*.

(69) Vacic, V.; Oldfield, C. J.; Mohan, A.; Radivojac, P.; Cortese, M. S.; Uversky, V. N.; Dunker, A. K. Characterization of molecular recognition features, MoRFs, and their binding partners. *J. Proteome Res.* **2007**, *6*, 2351–2366.

(70) Van Roey, K.; Uyar, B.; Weatheritt, R. J.; Dinkel, H.; Seiler, M.; Budd, A.; Gibson, T. J.; Davey, N. E. Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* **2014**, *114*, 6733–6778.

(71) Basu, S.; Bahadur, R. P. Conservation and coevolution determine evolvability of different classes of disordered residues in human intrinsically disordered proteins. *Proteins: Structure, Function, and Bioinformatics* **2021**, *90*, 632–644.

(72) Brown, C. J.; Johnson, A. K.; Dunker, A. K.; Daughdrill, G. W. Evolution and disorder. *Current Opinion in Structural Biology* **2011**, *21*, 441–446.

(73) Zarin, T.; Strome, B.; Nguyen Ba, A. N.; Alberti, S.; Forman-Kay, J. D.; Moses, A. M. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *eLife* **2019**, *8*, e46883.

(74) Zarin, T.; Tsai, C. N.; Nguyen Ba, A. N.; Moses, A. M. Selection maintains signaling function of a highly diverged intrinsically disordered region. *Proceedings of the National Academy of Sciences* **2017**, *114*.

(75) Lu, A. X.; Lu, A. X.; Pritišanac, I.; Zarin, T.; Forman-Kay, J. D.; Moses, A. M. Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning. *PLOS Computational Biology* **2022**, *18*, e1010238.

(76) Cascarina, S. M.; Ross, E. D. Proteome-scale relationships between local amino acid composition and protein fates and functions. *PLOS Computational Biology* **2018**, *14*, e1006256.

(77) Kato, M.; Zhou, X.; McKnight, S. L. How do protein domains of low sequence complexity work? *RNA* **2021**, *28*, 3–15.

(78) Lukatsky, D. B.; Afek, A.; Shakhnovich, E. I. Sequence correlations shape protein promiscuity. *The Journal of Chemical Physics* **2011**, *135*.

(79) Chong, S.; Graham, T. G. W.; Dugast-Darzacq, C.; Dailey, G. M.; Darzacq, X.; Tjian, R. Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription. *Molecular Cell* **2022**, *82*, 2084–2097.e5.

(80) Chong, S.; Dugast-Darzacq, C.; Liu, Z.; Dong, P.; Dailey, G. M.; Cattoglio, C.; Heckert, A.; Banala, S.; Lavis, L.; Darzacq, X.; Tjian, R. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **2018**, *361*, eaar2555.