

Role of Strong Localized vs. Weak Distributed Interactions in Disordered Protein Phase Separation

Shiv Rekhi,[†] Dinesh Sundaravadivelu Devarajan,[†] Michael P. Howard,[‡] Young C. Kim,[¶] Arash Nikoubashman,[§] and Jeetain Mittal^{*,†,||}

[†]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, United States*

[‡]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, United States*

[¶]*Center for Materials Physics and Technology, Naval Research Laboratory, Washington, DC 20375, United States*

[§]*Institute of Physics, Johannes Gutenberg University Mainz, Staudingerweg 7, 55128 Mainz, Germany*

^{||}*Department of Chemistry, Texas A&M University, College Station, TX 77843, United States*

E-mail: jeetain@tamu.edu

Phone: (512) 699-4643

Abstract

Interaction strength and localization are critical parameters controlling the single-chain and condensed-state properties of intrinsically disordered proteins (IDPs). Here, we decipher these relationships using coarse-grained heteropolymers comprised of hydrophobic (H) and polar (P) monomers as model IDPs. We systematically vary the fraction of P monomers X_P and employ two distinct particle-based models that include either strong localized attractions between only H-H pairs (HP model) or weak distributed attractions between both H-H and H-P pairs (HP+ model). To compare different sequences and models, we first carefully tune the attraction strength for all sequences to match the single-chain radius of gyration. Interestingly, we find that this procedure produces similar conformational ensembles, nonbonded potential energies, and chain-level dynamics for single chains of almost all sequences in both models, with some deviations for the HP model at large X_P . However, we observe a surprisingly rich phase behavior for the sequences in both models that deviates from the expectation that similarity at the single-chain level will translate to similar phase-separation propensity. Coexistence between dilute and dense phases is only observed up to a model-dependent X_P despite the presence of favorable interchain interactions, which we quantify using the second virial coefficient. Instead, the limited number of attractive sites (H monomers) leads to the self-assembly of finite-sized clusters of different sizes depending on X_P . Our findings strongly suggest that models with distributed interactions favor the formation of liquid-like condensates over a much larger range of sequence compositions compared to models with localized interactions.

Introduction

Concentrated assemblies of biomolecules that exist without a lipid membrane, termed membraneless organelles (MLOs) or biomolecular condensates, play a key role in several cellular processes that include cell organization,¹ signaling,² and stress response³ in addition to their relevance in pathological studies.⁴ In general, MLOs have liquid-like characteristics and form *via* phase separation.^{1,5,6} It has been found that intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) within proteins primarily drive phase separation.^{7,8} A common hypothesis for the prevalence of IDPs and IDRs in MLOs is their ability to form multivalent contacts^{7,9} *via* hydrogen bonding,¹⁰ electrostatic interactions,^{11,12} cation- π interactions^{13,14} as well as sp^2/π interactions.^{15,16} However, quantifying the relative contributions of these interaction types to the phase separation process is non-trivial, thus impeding the accurate prediction of the phase behavior of proteins based on sequence.^{9,17}

Despite these challenges, significant progress has been made in elucidating the sequence determinants of phase separation through experimental and computational studies.^{10–22} Among these studies, the work on the low-complexity domain (LCD) of hnRNPA1^{21,23} and the FUS prion-like domain²² have shown that the critical temperature T_c and the saturation (dilute-phase) concentration c_{sat} were altered when the number of aromatic residues (tyrosine) was varied. In addition, the mutation of arginine to lysine while maintaining the tyrosine content in FUS-like proteins led to a significant increase in c_{sat} , thus implying a reduction in phase-separation propensity.^{13,14} Further work on the FUS-LCD, using a combination of nuclear Overhauser experiments and all-atom molecular simulations, demonstrated that residues such as serine, glycine, and glutamine also play an important role in the stabilization of liquid-like condensates through interactions with other residues in the sequence.^{18,24} Studies involving shuffling of charges in LAF-1 RGG¹¹ and DDX4¹² have shown that charge patterning, frequently quantified by the sequence charge decoration parameter,²⁵ influences the phase-separation propensity of a protein, highlighting the key role played by electrostatic interactions. These findings have led to different interpretations of

the interaction scenarios governing the phase separation of proteins.

One class of models considers only a few residue types responsible for phase separation and introduces strong localized interactions between certain amino acid pairs such as tyrosine–tyrosine or tyrosine–arginine.^{7,13,14,26} Such models have been used to investigate single-component and multi-component phase separation. These models are in the spirit of the HP model popularized by Dill and coworkers to investigate protein folding, in which the protein is composed of hydrophobic (H) and polar (P) monomers, and there are energetically favorable interactions (attractions) between only the H monomers.^{27–29} More recent simulations using an off-lattice HP variant studied the different assemblies that formed when varying the fraction of H monomers at fixed chain length and fixed interaction strength.³⁰

Another class of protein models considers the collective contribution of all amino acids in a protein sequence by employing weak distributed interactions, comprised of various interaction modes, distributed along the protein.^{10,18,31} Transferable coarse-grained models developed by some of us, using amino acid specific contact potentials based on hydrophobicity, are reflective of this consideration and have been used to successfully characterize the phase behavior of reference proteins like FUS-LCD, LAF-1 RGG, α -synuclein, and A1-LCD.^{19,23,32–34} An extension of the model applying an additional temperature-dependent parabolic scaling for the nonbonded interactions was shown to capture experimentally observed³⁵ upper critical or lower critical solution temperature transitions for a set of thermoresponsive model IDPs.³⁶

In this work, we investigate the differences between these two classes of models by systematically studying the phase behavior of a coarse-grained heteropolymer consisting of H and P monomers. We consider an off-lattice variant of the HP model, which has strong localized attractions between only H monomers, and an extension we call the HP+ model, which has weak distributed attractions between both H–H and H–P monomer pairs. To make the different sequences and models comparable, we carefully scale the interactions between H monomers for each sequence to achieve the same single-chain dimensions as a purely hydrophobic reference sequence. Despite similarity at the single-chain level, differences be-

tween sequences and models emerge in the condensed-state, emphasizing the importance of the nature of intra- and interchain interactions in the formation of a homogeneous liquid-like condensate. The implications of our findings are of particular interest to the experimental interpretation of distinct interaction scenarios driving the phase separation of a given disordered protein.

Molecular Model and Simulation Methodology

We used a coarse-grained polymer model with a single bead per residue (monomer) and the solvent modeled implicitly in the effective monomer–monomer pair interactions. The sequences comprised of two types of monomers, namely hydrophobic (H) and polar (P) monomers having the same diameter $\sigma = 0.5$ nm and mass 100 g/mol. We fixed the chain length to $N = 20$ and randomly generated 21 sequences with the fraction of polar beads X_P varying from purely hydrophobic, $X_P = 0$, to purely hydrophilic, $X_P = 1$ (Fig. 1a). In addition to our primary sequence set, we also tested three other randomly generated sequence sets (RS1, RS2, and RS3 in Fig. S1) as well as a highly patterned sequence set (RS4 in Fig. S1).

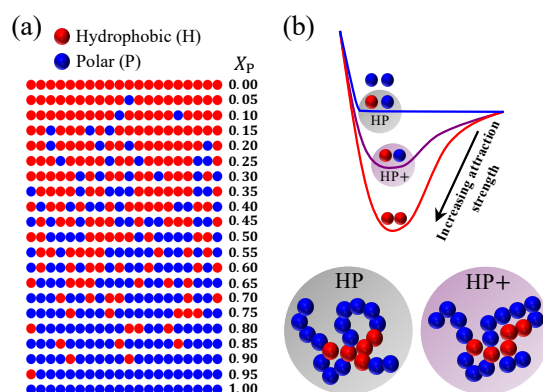


Figure 1: (a) Simulated sequences of length $N = 20$ with the fraction of polar beads X_P ranging from $X_P = 0$ (purely hydrophobic) to $X_P = 1$ (purely hydrophilic). (b) Schematic highlighting the different interactions used in the HP and HP+ models.

Interactions between bonded beads were modeled using a harmonic potential,

$$U_b(r) = \frac{k_b}{2}(r - r_0)^2, \quad (1)$$

with distance r between monomers, spring constant $k_b = 2000 \text{ kcal}/(\text{mol nm}^2)$, and equilibrium bond length $r_0 = 0.38 \text{ nm}$. Nonbonded interactions were modeled using a modified Lennard-Jones potential where the attractive contribution was scaled independently of the short-range repulsion by the pairwise hydrophathy λ_{ij} for monomers of type i and j ,^{19,37–39}

$$U_{nb}(r) = \begin{cases} U_{LJ}(r) + (1 - \lambda_{ij})\varepsilon, & r \leq 2^{1/6}\sigma \\ \lambda_{ij}U_{LJ}(r), & \text{otherwise} \end{cases}, \quad (2)$$

where U_{LJ} is the standard Lennard-Jones potential,

$$U_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (3)$$

and ε is the interaction strength. We note that Eq. (2) reduces to the purely repulsive Weeks–Chandler–Andersen potential,³⁹ which describes excluded volume interactions between monomers under good solvent conditions, for $\lambda_{ij} = 0$ and to Eq. (3), which also includes effective attraction between monomers as in a lower quality solvent, for $\lambda_{ij} = 1$.

To determine λ_{ij} , we accordingly assigned a nominal hydrophathy $\lambda_P = 0$ to the P monomers and a nonzero hydrophathy λ_H to the H monomers. (The specific value of λ_H was varied as described later.) We set $\lambda_{PP} = \lambda_P$ and $\lambda_{HH} = \lambda_H$ so that H–H monomer pairs were attracted to each other but P–P monomer pairs were not. We then proposed two different models (HP and HP+) for the cross-interactions between H–P monomer pairs controlled by λ_{HP} . The HP model has strong localized attractions between only a certain type of residue, *e.g.*, self-interactions between aromatic residues like tyrosine, so we set $\lambda_{HP} = 0$. The HP+ model also has weak distributed attractions with additional residues, *e.g.*, cross

interactions between aromatic residues and polar residues like glutamine, asparagine, serine, etc. (Fig. 1b), so we set $\lambda_{HP} = (\lambda_H + \lambda_P)/2 = \lambda_H/2$ to the mean hydropathy of the H and P monomers. In both models, we set $\varepsilon = 0.2 \text{ kcal/mol}$, which, in conjunction with other model parameters, was previously shown to accurately capture IDP properties like radius of gyration R_g .^{19,33}

We performed Langevin dynamics simulations in the low-friction limit [friction coefficient $0.1 \text{ g}/(\text{mol fs})$] at fixed temperature $T = 300 \text{ K}$ for a total duration of $1 \mu\text{s}$ using a timestep of 10 fs . We performed single-chain simulations using LAMMPS (29 October 2020 version),⁴⁰ while condensed-phase simulations were performed using HOOMD-blue (version 2.9.3)⁴¹ with features extended using azplugins (version 0.10.1).⁴² We simulated phase coexistence in rectangular simulation boxes ($10 \text{ nm} \times 10 \text{ nm} \times 75 \text{ nm}$) with 500 chains initially placed in a dense slab with surface normals aligned with the z -axis. We also performed simulations of the same number of chains in cubic boxes of edge length 40 nm to study the formation of finite-sized aggregates.

Results and Discussion

Single-chain heteropolymer ensembles are matched by scaling attractions

Since our model sequences have the same degree of polymerization ($N = 20$), their single-chain dimensions (*i.e.*, at infinite dilution) are mainly dictated by the fraction of polar monomers X_P and specific intramolecular interactions. Strong monomer attraction would lead to the collapse of the chain, mimicking poor solvent conditions, whereas pure repulsion would lead to chain expansion analogous to good solvent conditions.^{9,43} Single-chain compactness has been shown to correlate with the propensity to phase separate both experimentally and computationally for a wide range of IDPs.^{44–46} Here, we quantified single-chain compactness using the polymer’s radius of gyration R_g , which we computed from the

average of the trace of the gyration tensor.^{38,47} To facilitate comparison of our sequence- and model-dependent results, we used a purely hydrophobic polymer ($X_P = 0$) with $\lambda_H = 1$ and a purely hydrophilic polymer ($X_P = 1$) as reference sequences.

We first measured R_g for the different sequences and models while fixing $\lambda_H = 1$. As expected, we found that R_g increased with increasing X_P , indicating a reduced hydrophobic character (Figs. 2a and S2a). Furthermore, the probability distribution of R_g widened with increasing X_P , highlighting that the chains exhibited greater conformational fluctuations for large X_P (Fig. S2b). Since our primary goal is to study the effect of interaction model and sequence patterning on phase behavior, we tuned the monomer interactions so that all investigated polymers had a comparable single-chain R_g (and thus similar compaction at infinite dilution) as the purely hydrophobic reference sequence. To account for the decrease in the hydrophobicity of the chain with increasing X_P , we initially attempted to rescale $\lambda_H = 1/(1 - X_P)$ to maintain the same average hydropathy per monomer $\langle \lambda \rangle = N^{-1} \sum_{i=1}^N \lambda_i$ as the purely hydrophobic reference sequence. Scaling the interactions in this way did not, however, result in constant R_g , highlighting the need for fine tuning of the interaction strengths for both models (Fig. S2a).

Accordingly, we introduced an additional scaling factor a to modulate $\lambda_H = a/(1 - X_P)$. We selected a for each polymer by running a series of simulations to determine R_g as a function of a (Fig. 2b), then found the value of a that matched R_g to the hydrophobic reference sequence. We found $a > 1$ for the HP model, suggesting that the required monomer–monomer attractions are stronger compared to those needed to maintain the average hydropathy of the sequence (*i.e.*, $\langle \lambda \rangle > 1$). Conversely, we found $a < 1$ in the HP+ model and thus $\langle \lambda \rangle < 1$. With these scaling factors applied to the HP model, we observed that for all values of X_P , other than 0.90 and 0.95, R_g remained within 3% of that of the purely hydrophobic reference sequence (Fig. 2a). In the $X_P = 0.90$ sequence, strong attraction existed between the only two H monomers present in the sequence, and therefore their position within the sequence mainly dictated R_g (Figs. 2a and S3), while in the $X_P = 0.95$ case, the chain behaved as a

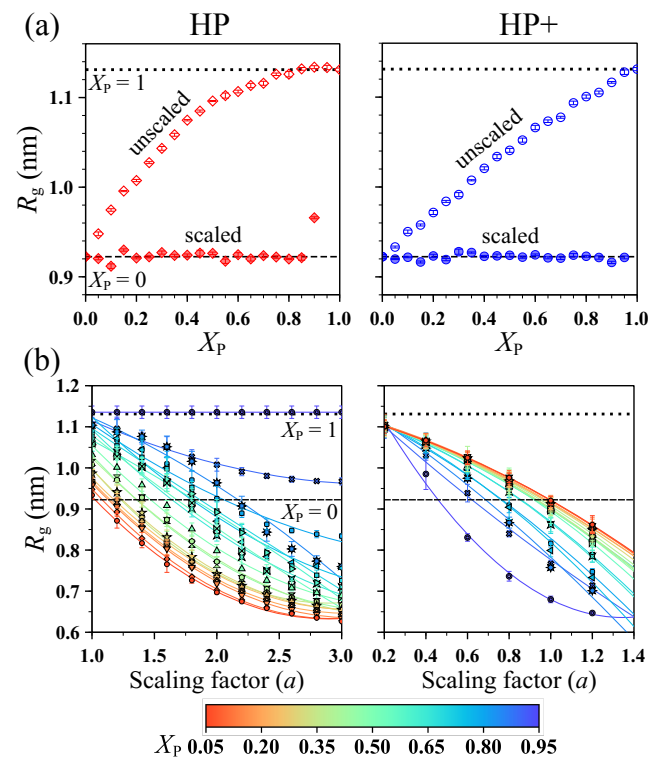


Figure 2: (a) Radius of gyration R_g as a function of X_P before scaling ($\lambda_H = 1$, open symbols) and after scaling ($\lambda_H = a/(1 - X_P)$, closed symbols) for the HP (red) and HP+ (blue) models. (b) R_g as a function of scaling factor a using the HP and HP+ models. Results for purely hydrophobic ($X_P = 0$) and purely hydrophilic ($X_P = 1$) reference sequences are shown as dashed and dotted black lines, respectively.

purely hydrophilic chain due to the presence of only one H monomer. For the HP+ model, we matched R_g for all X_P including the 0.90 and 0.95 cases due to the additional attractive interaction between H and P monomers present in the model.

To further quantify the similarity established at the single-chain level, we next investigated if the match in average R_g also extended to the probability distributions $P(R_g)$. Indeed, we found similar $P(R_g)$ in the HP model for $X_P \lesssim 0.5$, but the agreement quickly deteriorated for larger X_P (Fig. 3a), possibly due to the reduction in the number of available interaction sites (*i.e.*, H monomers). For the HP+ model, we found a near perfect overlap of $P(R_g)$ (Fig. 3b), implying that the single-chain conformations in the HP+ model were almost identical to those of a purely hydrophobic reference sequence for all investigated X_P values. To further test this similarity, we considered the shape anisotropy $\langle \kappa^2 \rangle$, computed using the eigenvalues of the gyration tensor;^{23,38} the nonbonded potential energies U_{nb} ; and the relaxation time τ_e of the end-to-end vector autocorrelation function³⁸ (Fig. 3c). Remarkably, we found for almost all X_P values $\langle \kappa^2 \rangle \approx 0.39$, similar to the numerically determined value of a three-dimensional random walk,⁴⁸ despite the very different sequence compositions. Again, we found that the HP+ model reproduced the reference state up to significantly larger X_P compared to the HP model. We observed similar trends in the nonbonded potential energy U_{nb} : for the HP model, U_{nb} remained almost the same as in the purely hydrophobic reference sequence up to $X_P \approx 0.5$, while in the HP+ model, it remained the same throughout the entire X_P range. Surprisingly, we did not see as pronounced a difference between the two models when considering the end-to-end vector relaxation times τ_e with varying X_P . To check the sensitivity of the employed models and single-chain indicators, we considered three additional sequence sets in the range from $X_P = 0$ to $X_P = 1$ (RS1, RS2, and RS3 in Fig. S1) but did not find any strong sequence dependence (Figs. S3 and S4). Additionally, we found that the scaling factor a remained similar for all sequence sets considered in this work (Fig. S5).

To summarize, tuning the attraction strength between H monomers gave similar average

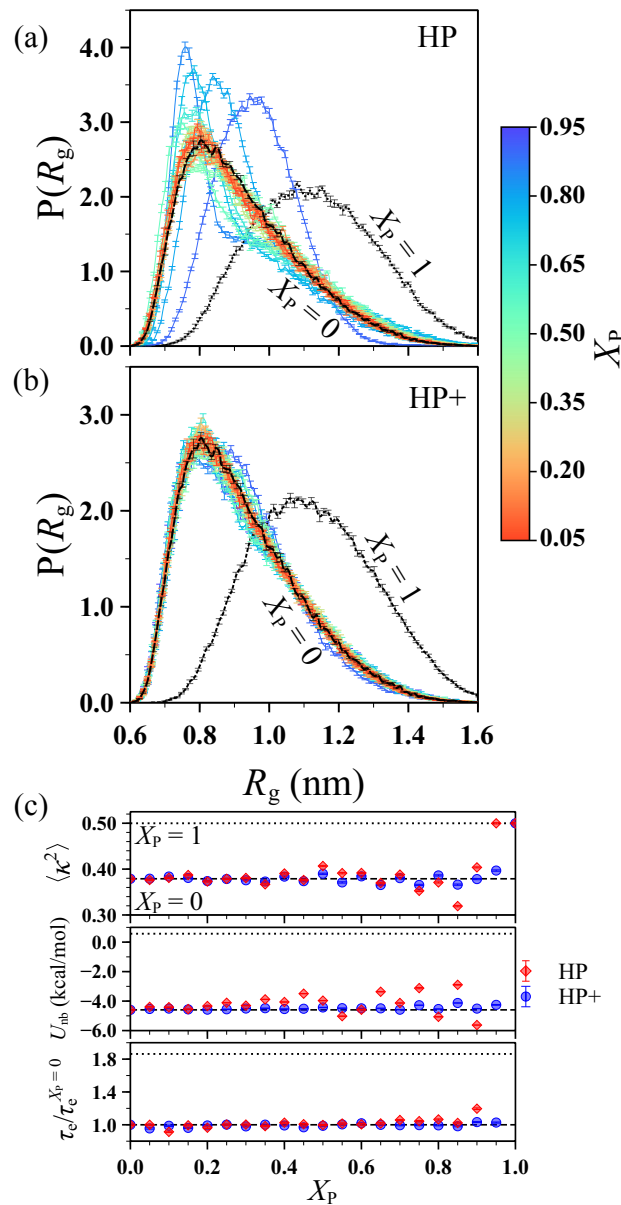


Figure 3: Probability distribution of radius of gyration R_g for the (a) HP and (b) HP+ models. (c) Shape anisotropy $\langle \kappa^2 \rangle$, nonbonded potential energy U_{nb} , and end-to-end vector relaxation time τ_e as functions of X_P for the HP and HP+ models. The last quantity in (c) is normalized by that obtained for the purely hydrophobic reference sequence ($X_P = 0$). Results for purely hydrophobic and purely hydrophilic ($X_P = 1$) reference sequences are shown as dashed and dotted black lines, respectively.

single-chain properties and $P(R_g)$ as the purely hydrophobic reference sequence across for all sequences in the HP+ model, but only about half of the sequences in the HP model. The better match over the entire X_P range is due to the additional attractive interaction between H and P monomers in the HP+ model. The HP+ model is less sensitive to changes in the number of attractive monomers as compared to the HP model at the single-chain level and hence might be expected to better capture the condensed-phase behavior of the purely attractive homopolymer for different values of X_P .

Localized interactions reduce phase-separation propensity

Having established similar single-chain properties for both models, we next investigated the implications for the condensed-state properties of the sequences studied. According to homopolymer-based solution theories, intrachain contacts are translated to interchain contacts in concentrated solutions, thus linking the behavior at the single-chain level to phase-separation propensity.^{9,46,49,50} Panagiotopoulos and coworkers⁵¹ showed *via* on-lattice Monte Carlo simulations that in the limit of infinitely long homopolymer chains, the critical temperature, the Boyle temperature, and the coil-to-globule transition temperature are equivalent. Recent simulations of heteropolymeric model IDPs and naturally occurring IDPs have revealed a similar correlation between the extent of collapse at the single-chain level and phase-separation propensity.⁴⁶ Extending the ideas above, corrections were proposed to account for the correlation between single-chain and condensed-state properties for charged heteropolymers.¹⁴ These prior works suggest that principles from homopolymer theory can be applied with reasonable accuracy when attractive monomers are distributed rather uniformly along the protein sequence.²³ With these considerations, we investigated whether the similarity established at the single-chain level by scaling monomer interactions also leads to similarities in phase behavior for our model sequences having strong localized or weak distributed interactions.

We performed direct coexistence simulations of the dilute and dense phases, computed

time-averaged concentration profiles (Fig. 4), and extracted the coexistence concentrations of the dilute and dense phases (Fig. 5). In the HP model, we found similar dilute-phase concentrations c_{sat} for $X_P \leq 0.10$, while beyond $X_P = 0.15$, we observed an increase in c_{sat} and a corresponding decrease in the phase-separation propensity (Figs. 4 and 5). We also found that the dense-phase concentration started to decrease beyond $X_P = 0.15$. For $X_P \geq 0.50$, we found parabolic concentration profiles centered at the origin, indicative of the formation of clusters rather than a bulk condensed-phase (see representative simulation snapshots⁵² for $X_P = 0.50$ and $X_P = 0.75$ in Fig. 4). From these observations, we concluded that the threshold value above which the sequences did not phase separate is roughly $X_P^* = 0.45$ for the HP model.

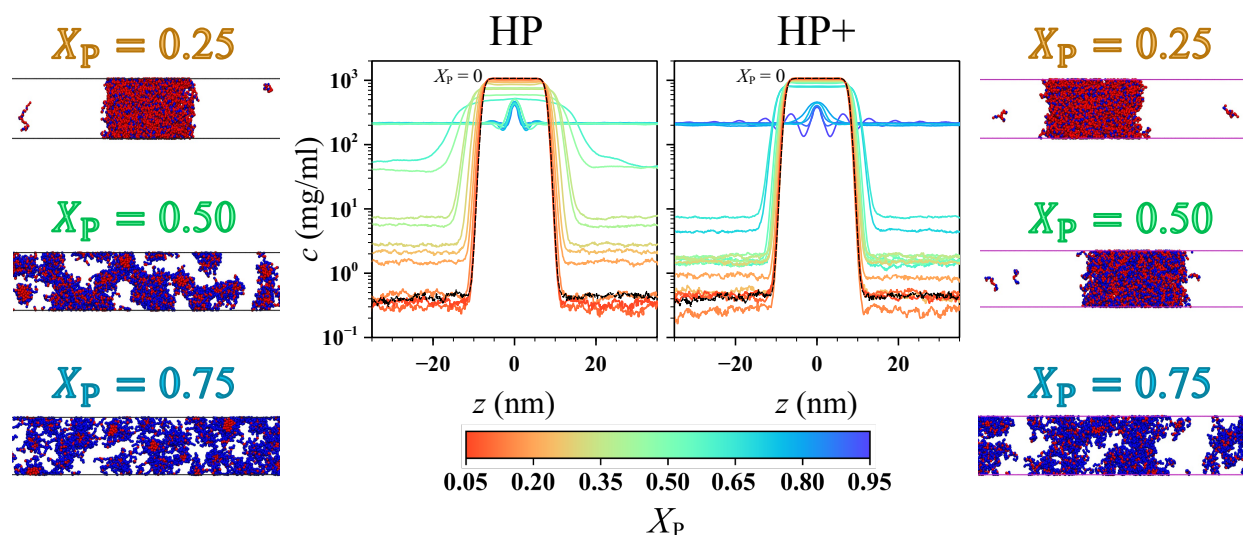


Figure 4: Concentration profiles estimated using direct coexistence simulations for the HP and HP+ models. The black line represents the purely hydrophobic reference sequence ($X_P = 0$). Representative simulation snapshots for $X_P = 0.25, 0.50$, and 0.75 are shown on the left (HP) and right (HP+) sides.

In the HP+ model, phase separation persisted for higher X_P compared to the HP model (Figs. 4 and 5), highlighting the stabilizing role played by the cross interactions between H and P monomers. We found that the coexistence concentrations remained almost the same only for $X_P \leq 0.15$. Beyond $X_P = 0.20$, c_{sat} increased, indicative of lower phase-separation propensity as compared to the purely hydrophobic reference sequence under the conditions

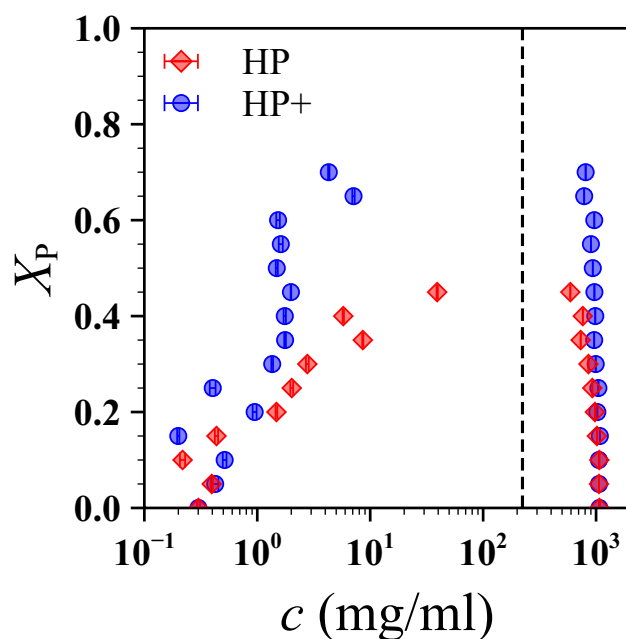


Figure 5: Phase diagrams for the HP and HP+ models. Concentrations of dilute and dense phases for all X_P values were extracted from the concentration profiles shown in Figure 4. Dashed line represents the total concentration of the coexistence simulation.

of this study. However, compared to the HP model, we observed cluster formation only for large X_P values ($X_P \geq 0.75$; Fig. 4). Based on this finding, we defined $X_P^* = 0.70$ for the HP+ model, which suggests that at least 30% of the sequence must consist of hydrophobic monomers to form a liquid-like condensate at the simulated temperature. We find that for both models, matching the single-chain dimensions through scaling of interactions leads to a dramatic increase in the threshold values (0.1 to 0.45 for HP and 0.25 to 0.70 for HP+) for phase separation when compared to the unscaled interactions (Fig. S6). This finding indicates that scaling the interactions between H monomers in order to match the single-chain dimensions did offset the effect of changes in composition in both models. However, the extent to which this similarity established at the single-chain level could compensate for changes in composition of the sequence was limited.

The sequences with X_P around the threshold X_P^* value ($X_P = 0.35$ and 0.40 in the HP model, and $X_P = 0.65$ and 0.70 in the HP+ model) formed a wider dense phase slab with its concentration about 75% of that of the purely hydrophobic reference sequence ($X_P = 0$). To

investigate whether this widening is due to the decrease in concentration or the formation of void volume within the slab, we computed the radial distribution function $g(r)$ between H monomers for these sequences (Fig. S7). We found only a marginal increase in the magnitude of the first peak in $g(r)$ compared to the $X_P = 0$ sequence, highlighting the absence of preferential patterning of H and P monomers in the dense phase of these sequences. Both H and P monomers were well mixed within the dense phase of all sequences up to X_P^* (Fig. 5), further substantiating their ability to form homogeneous liquid-like condensates.

We next probed whether the system could form a condensed-phase above the threshold X_P^* values (0.45 for the HP model and 0.70 for the HP+ model) by further increasing the attraction between H monomers. To this end, we ran simulations for all cases that did not form stable slabs using the HP+ model with a increased to 2, 5, and 10 times that needed to match the single-chain R_g . At such high interaction strengths, we observed the formation of highly patterned slabs as reflected in their concentration profiles (Fig. S8). Additionally, the concentrations of the dense phase of these sequences showed a system size dependence, analogous to those classified as aggregates by Panagiotopoulos and coworkers.⁵³ From these observations, we concluded that insufficient interaction strength was likely not the reason why the sequences above X_P^* did not form liquid-like condensates.

To investigate the sequence dependence of our observations, we also characterized the phase behavior of three additional sequence sets (RS1, RS2, and RS3 in Fig. S1). We found that the threshold value X_P^* showed a stronger sequence dependence than the single-chain behavior (Fig. S9). However, X_P^* for the HP+ model was always higher than that observed for the HP model for all sequence sets, consistent with our primary sequence set. Taken together, we found that matching single-chain properties did not prevent the decrease in the phase-separation propensity with increasing localization of interactions within the sequences, indicating that interaction strength alone did not dictate the formation of a homogeneous condensed-phase but rather a combination of interaction strength and the number of attractive monomers.

Scaled interactions lead to favorable interchain interactions for all sequence compositions

A possible reason for the reduction in phase-separation propensity with increasing X_P could be the decreasing favorability of interchain interactions despite favorable intrachain interactions. To quantify the strength of interaction between a pair of chains, we determined the chain–chain second virial coefficient B_{22} . Positive B_{22} indicates effective repulsion, while negative B_{22} indicates effective attraction. We used a combination of replica exchange Monte Carlo and umbrella sampling methods to compute the dilute chain–chain potential of mean force (PMF), $w(r)$. Then, B_{22} was computed as^{32,46}

$$B_{22} = 2\pi \int_0^\infty [1 - e^{-\beta w(r)}] r^2 dr. \quad (4)$$

The above method is known to estimate B_{22} with a high degree of statistical certainty.^{32,46} Fig. 6a shows the PMF for select X_P values (see Figs. S10 and S11 for the entire set), while the resulting B_{22} values for all X_P values are shown in Fig. 6b. For $X_P \lesssim 0.40$, both models showed highly similar PMF to that of the purely hydrophobic reference sequence in terms of both the location and depth of the attractive well. At intermediate to high values of X_P , we observed changes between the two models: PMFs of the HP+ model were shallower and closer to the purely hydrophobic reference sequence as compared to the HP model (Fig. 6a). This behavior implies that the effective chain-chain interactions in the HP model are stronger than in the HP+ model. For $X_P = 0.90$, the PMF obtained for the HP model is an order of magnitude more attractive than that of the HP+ model, possibly owing to the higher attraction strength needed to match the single chain dimensions using only two attractive H monomers.

This comparison shows that the similar behavior achieved at the single-chain level for all values of X_P translates to similar interactions at the two-chain level, with slightly negative B_{22} values up to roughly X_P^* . Beyond X_P^* , for both models, we see that the B_{22} values

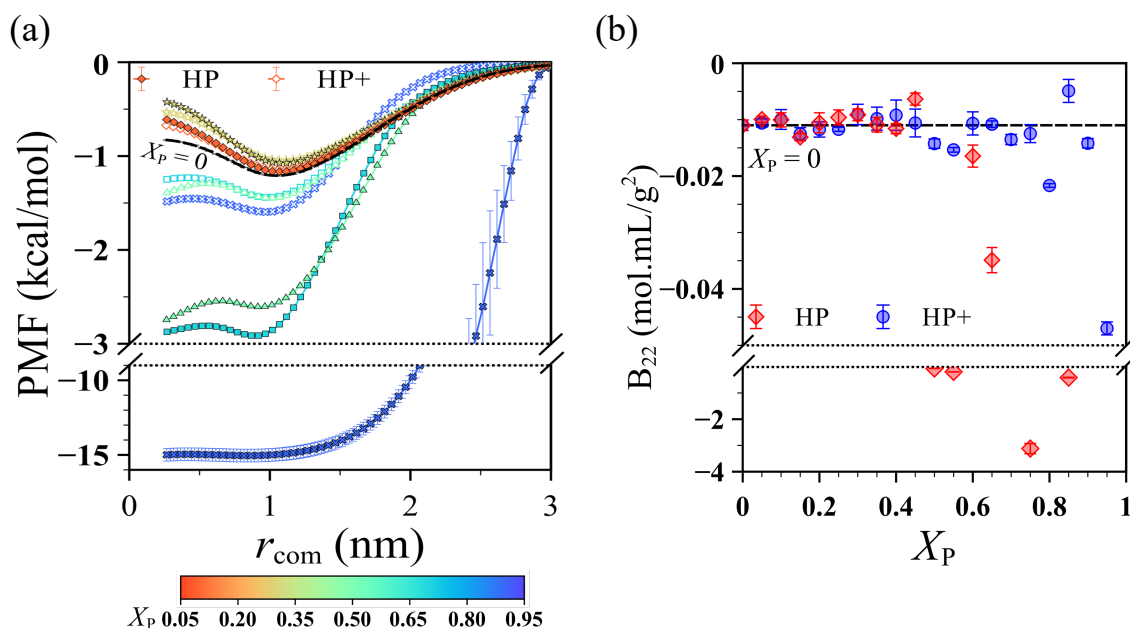


Figure 6: (a) Potential of mean force (PMF) as a function of center-of-mass distance r_{com} between two chains for select X_P values using the HP (filled symbols) and HP+ (open symbols) models. (b) Second virial coefficient B_{22} obtained from the PMFs for the HP and HP+ models.

fluctuate, with the HP model showing more negative values than the HP+ model indicating that the effective attraction between chains is stronger in the HP model than the HP+ model as the number of attractive sites decreases. In some cases ($X_P \geq 0.50$ for the HP model and $X_P = 0.80, 0.90$, and 0.95 for the HP+ model), the chain-chain interactions are more favorable compared to the purely hydrophobic case, which implies that from a pairwise interaction standpoint, these cases would be more prone to chain-chain association than the purely hydrophobic reference sequence. However, we did not observe the formation of a homogeneous condensate in those cases. This discrepancy prompts the question: what limits the system from undergoing phase separation if interactions remain favorable?

Beyond the threshold for phase separation, finite-sized clusters are formed

Motivated by our finding that sequences with $X_P > X_P^*$ do not form a condensed homogeneous phase despite exhibiting strong chain-chain attraction, we next investigated the energetically favorable assemblies formed by such sequences using both models. To this end, we let the chains self-assemble within a cubic simulation box, and then performed a clustering analysis of the resulting aggregates. Specifically, we considered two chains to be part of the same cluster if the distance between monomers of two chains was less than 1.5σ , and then computed the probability of finding a chain in a cluster of size N_c (Fig. 7). The probability distributions $P(N_c)$ were relatively insensitive to the distance criteria chosen for analysis (Fig. S12). To rule out possible finite-size effects resulting from different box sizes and geometries (slab vs. cubic),^{46,54} we first performed the analysis on sequences that undergo phase separation (*i.e.*, $X_P < X_P^*$; Fig. S13). Consistent with the results from slab simulations, we found that the chains predominantly resided in a single large cluster for these sequences.

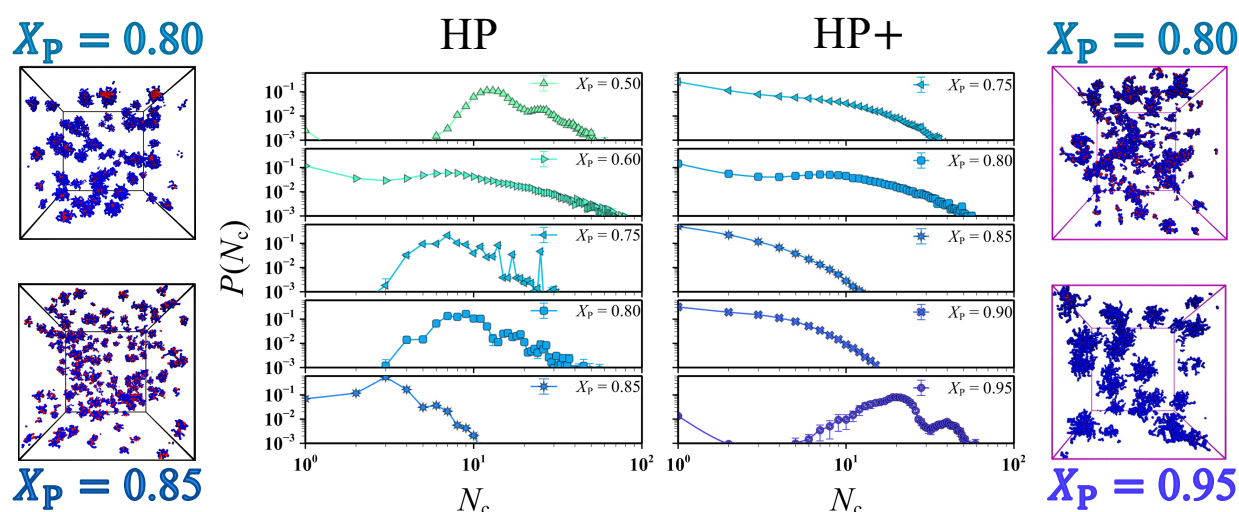


Figure 7: Probability of finding a chain in a cluster of size N_c for select X_P values that did not undergo phase separation in the HP and HP+ models. Representative simulation snapshots for $X_P = 0.80$ and 0.85 on the left side correspond to the HP model, while those for $X_P = 0.80$ and 0.95 on the right side correspond to the HP+ model.

Beyond X_P^* , the probability of finding chains within smaller clusters was nonzero for both models (Fig. 7), indicating that interchain interactions are still favored in the limit of multiple chains. Interchain contacts are more favored in the case of the HP model as compared to the HP+ model, reflected by the higher probability of observing clusters of chains as compared to single chains in solution. The attraction between H and P monomers in the HP+ model allowed isolated chains to form intrachain contacts for all but $X_P = 0.95$, which formed micelles with their cores consisting of H monomers from multiple chains (Fig. 7). We hypothesized that the different behavior exhibited for $X_P = 0.95$ in the HP+ model could be due to the rather uniform patterning of H monomers. Indeed, in the case of a highly patterned sequence set (RS4 in Fig. S1), we found that all systems beyond X_P^* preferably formed small micellar clusters, thus highlighting the role of the arrangement of H monomers in dictating the formation of interchain or intrachain contacts (Fig. S14).³⁰ Despite some differences between the specific cluster morphologies in the two models, it is clear that the general emergence of smaller clusters beyond X_P^* is model-independent. This finding further highlights that the formation of a single high concentration phase is governed by the interplay between favorability of interactions and the number of available interaction sites.^{55–58} Beyond X_P^* , the system is limited by the number of available interaction sites (or valence), resulting in its inability to form a single continuous stable condensed-phase.

Conclusions

In this computational study, we comprehensively investigated the interplay between the localization and strength of interactions in dictating the phase behavior of model disordered proteins consisting of hydrophobic (H) and polar (P) monomers. We systematically varied the fraction of P monomers in the sequences, X_P , and considered two model classes, representing interaction scenarios with either strong localized or weak distributed attractions. To establish a common ground for the large variation in IDP sequences, we carefully scaled

the attraction strength in each case to match the single-chain dimensions of an attractive homopolymer, thereby providing all of the model proteins an equal opportunity to phase separate, from a single-chain perspective. To our surprise, scaling the interactions this way also led to similar conformational ensembles, nonbonded potential energies, and chain-level dynamics in both models for almost all investigated sequences. Based on the expectation that intramolecular interactions will translate directly to the intermolecular level, we probed the propensity of the IDPs to phase separate. We found similar phase-separation propensity for IDPs with sufficiently low X_P , where a significant scale up in the attraction strength counterbalanced the decreasing number of H monomers. As X_P increased, however, the phase-separation propensity dramatically declined because of limited number of attractive monomers. Beyond a certain X_P , the model IDPs did not form a continuous condensed-phase anymore, but instead self-assembled into finite-sized aggregates. These deviations from the reference attractive homopolymer were much more pronounced for the model with strong localized interactions, suggesting that weak distributed interactions between multiple residue types may better stabilize the liquid-like condensed-phase of disordered proteins.^{10,18,24,31} We believe that our work, performed at multiple scales, will aid in the mechanistic understanding of biomolecular phase separation and in the development of theoretical models to accurately capture the interactions of disordered proteins.

Acknowledgments

This material is based on the work supported by the National Institute of General Medical Science of the National Institutes of Health under the grant R01GM136917 and the Welch Foundation under the grant A-2113-20220331. AN acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through project 470113688. Y.C.K. is supported by the Office of Naval Research through the U.S. Naval Research Laboratory base program. Further, we acknowledge the Texas A&M High Performance Research

Computing (HPRC) for providing the computational resources required to complete this work.

References

- (1) Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, eaaf4382.
- (2) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F. et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483*, 336–340.
- (3) Shih, J.-W.; Wang, W.-T.; Tsai, T.-Y.; Kuo, C.-Y.; Li, H.-K.; Wu Lee, Y.-H. Critical roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response. *Biochemical Journal* **2012**, *441*, 119–129.
- (4) Alberti, S.; Dormann, D. Liquid-liquid phase separation in disease. *Annu. Rev. Genet* **2019**, *53*, 171–194.
- (5) Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol* **2014**, *30*, 39–58.
- (6) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **2009**, *324*, 1729–1732.
- (7) Martin, E. W.; Holehouse, A. S. Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof. *Emerging topics in life sciences* **2020**, *4*, 307–329.
- (8) Borchers, 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.

- (9) 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.
- (10) Murthy, A. C.; Dignon, G. L.; Kan, Y.; Zerze, G. H.; Parekh, S. H.; Mittal, J.; Fawzi, N. L. Molecular interactions underlying liquid–liquid phase separation of the FUS low-complexity domain. *Nature structural & molecular biology* **2019**, *26*, 637–648.
- (11) Schuster, B. S.; Dignon, G. L.; Tang, W. S.; Kelley, F. M.; Ranganath, A. K.; Jahnke, C. N.; Simpkins, A. G.; Regy, R. M.; Hammer, D. A.; Good, M. C. et al. Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of the National Academy of Sciences* **2020**, *117*, 11421–11431.
- (12) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D. et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **2017**, *114*, E8194–E8203.
- (13) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D. et al. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **2018**, *174*, 688–699.
- (14) Bremer, A.; Farag, M.; Borchers, 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. *Nature Chemistry* **2022**, *14*, 196–207.

- (15) 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*, e31486.
- (16) Das, S.; Lin, Y.-H.; Vernon, R. M.; Forman-Kay, J. D.; Chan, H. S. Comparative roles of charge, π , and hydrophobic interactions in sequence-dependent phase separation of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2020**, *117*, 28795–28805.
- (17) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **2019**, *176*, 419–434.
- (18) Murthy, A. C.; Tang, W. S.; Jovic, N.; Janke, A. M.; Seo, D. H.; Perdikari, T. M.; Mittal, J.; Fawzi, N. L. Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads. *Nature structural & molecular biology* **2021**, *28*, 923–935.
- (19) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS computational biology* **2018**, *14*, e1005941.
- (20) Mohanty, P.; Kapoor, U.; Sundaravadivelu Devarajan, D.; Phan, T. M.; Rizuan, A.; Mittal, J. Principles governing the phase separation of multidomain proteins. *Biochemistry* **2022**, *61*, 2443–2455.
- (21) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694–699.
- (22) Lin, Y.; Currie, S. L.; Rosen, M. K. Intrinsically disordered sequences enable modulation

- p of protein phase separation through distributed tyrosine motifs.
- Journal of Biological Chemistry*
- 2017**
- ,
- 292*
- , 19110–19120.
- (23) Bauer, D. J.; Stelzl, L. S.; Nikoubashman, A. Single-chain and condensed-state behavior of hnRNPA1 from molecular simulations. *The Journal of Chemical Physics* **2022**, *157*, 154903.
 - (24) Zheng, W.; Dignon, G. L.; Jovic, N.; Xu, X.; Regy, R. M.; Fawzi, N. L.; Kim, Y. C.; Best, R. B.; Mittal, J. Molecular details of protein condensates probed by microsecond long atomistic simulations. *The Journal of Physical Chemistry B* **2020**, *124*, 11671–11679.
 - (25) Firman, T.; Ghosh, K. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *The Journal of chemical physics* **2018**, *148*, 123305.
 - (26) Zhang, Y.; Xu, B.; Weiner, B. G.; Meir, Y.; Wingreen, N. S. Decoding the physical principles of two-component biomolecular phase separation. *eLife* **2021**, *10*, e62403.
 - (27) Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985**, *24*, 1501–1509.
 - (28) Lau, K. F.; Dill, K. A. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* **1989**, *22*, 3986–3997.
 - (29) Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133–7155.
 - (30) Statt, A.; Casademunt, H.; Brangwynne, C. P.; Panagiotopoulos, A. Z. Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *The Journal of chemical physics* **2020**, *152*, 075101.
 - (31) Holehouse, A. S.; Ginell, G. M.; Griffith, D.; Böke, E. Clustering of aromatic residues in prion-like domains can tune the formation, state, and organization of biomolecular condensates. *Biochemistry* **2021**, *60*, 3566–3581.

- (32) Kim, Y. C.; Hummer, G. Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *Journal of molecular biology* **2008**, *375*, 1416–1433.
- (33) Regy, R. M.; Thompson, J.; Kim, Y. C.; Mittal, J. Improved coarse-grained model for studying sequence dependent phase separation of disordered proteins. *Protein Science* **2021**, *30*, 1371–1379.
- (34) Das, S.; Muthukumar, M. Microstructural organization in α -synuclein solutions. *Macromolecules* **2022**, *55*, 4228–4236.
- (35) Quiroz, F. G.; Chilkoti, A. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nature materials* **2015**, *14*, 1164–1171.
- (36) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J. Temperature-controlled liquid–liquid phase separation of disordered proteins. *ACS central science* **2019**, *5*, 821–830.
- (37) Ashbaugh, H. S.; Hatch, H. W. Natively unfolded protein stability as a coil-to-globule transition in charge/hydrophathy space. *Journal of the American Chemical Society* **2008**, *130*, 9536–9542.
- (38) Sundaravadivelu Devarajan, D.; Rekhi, S.; Nikoubashman, A.; Kim, Y. C.; Howard, M. P.; Mittal, J. Effect of charge distribution on the dynamics of polyampholytic disordered proteins. *Macromolecules* **2022**, *55*, 8987–8997.
- (39) Weeks, J. D.; Chandler, D.; Andersen, H. C. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of chemical physics* **1971**, *54*, 5237–5247.
- (40) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D. et al.

- LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comp. Phys. Comm.* **2022**, *271*, 108171.
- (41) Anderson, J. A.; Lorenz, C. D.; Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of computational physics* **2008**, *227*, 5342–5359.
- (42) <https://github.com/mphowardlab/azplugins>.
- (43) Michels, J. J.; Brzezinski, M.; Scheidt, T.; Lemke, E. A.; Parekh, S. H. Role of solvent compatibility in the phase behavior of binary solutions of weakly associating multivalent polymers. *Biomacromolecules* **2021**, *23*, 349–364.
- (44) Lin, Y.-H.; Chan, H. S. Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophysical Journal* **2017**, *112*, 2043–2046.
- (45) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **2012**, *109*, 16155–16160.
- (46) Dignon, G. L.; Zheng, W.; Best, R. B.; Kim, Y. C.; Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2018**, *115*, 9929–9934.
- (47) Rubinstein, M.; Colby, R. H. *Polymer physics*; Oxford University Press, NY, 2003.
- (48) Rudnick, J.; Gaspari, G. The Shapes of Random Walks. *Science* **1987**, *237*, 384–389.
- (49) Flory, P. J. Thermodynamics of High Polymer Solutions. *The Journal of Chemical Physics* **1942**, *10*, 51–61.
- (50) Huggins, M. L. Some properties of solutions of long-chain compounds. *The Journal of Physical Chemistry* **1942**, *46*, 151–158.

- (51) Panagiotopoulos, A. Z.; Wong, V.; Floriano, M. A. Phase equilibria of lattice polymers from histogram reweighting Monte Carlo simulations. *Macromolecules* **1998**, *31*, 912–918.
- (52) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.
- (53) Rana, U.; Brangwynne, C. P.; Panagiotopoulos, A. Z. Phase separation vs aggregation behavior for model disordered proteins. *The Journal of Chemical Physics* **2021**, *155*, 125101.
- (54) Nilsson, D.; Irbäck, A. Finite-size shifts in simulated protein droplet phase diagrams. *The Journal of Chemical Physics* **2021**, *154*, 235101.
- (55) Sun, S.; Yang, S.; Xin, H. L.; Nykypanchuk, D.; Liu, M.; Zhang, H.; Gang, O. Valence-programmable nanoparticle architectures. *Nature communications* **2020**, *11*, 1–10.
- (56) Markova, O.; Alberts, J.; Munro, E.; Lenne, P.-F. Bond flexibility and low valence promote finite clusters of self-aggregating particles. *Physical Review Letters* **2012**, *109*, 078101.
- (57) Russo, J.; Tartaglia, P.; Sciortino, F. Reversible gels of patchy particles: role of the valence. *The Journal of chemical physics* **2009**, *131*, 014504.
- (58) Rovigatti, L.; Sciortino, F. Designing enhanced entropy binding in single-chain nanoparticles. *Physical Review Letters* **2022**, *129*, 047801.