Phase separation versus aggregation behavior for model disordered proteins

Ushnish Rana,[1] Clifford P. Brangwynne,[1,2] and Athanassios Z. Panagiotopoulos[1,a)]
[1)] *Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544*
[2)] *Howard Hughes Medical Institute, Chevy Chase, MD 20815*

(Dated: 15 June 2021)

Liquid-liquid phase separation (LLPS) is widely utilized by the cell to organize and regulate various biochemical processes. Although the LLPS of proteins is known to occur in a sequence dependent manner, it is unclear how sequence properties dictate the nature of the phase transition and thereby influence condensed phase morphology. In this work, we have utilized grand canonical Monte Carlo simulations for a simple coarse-grained model of disordered proteins to systematically investigate how sequence distribution, sticker fraction and chain length influence the phase behavior and regulate the formation of finite-size aggregates preempting macroscopic phase separation for some sequences. We demonstrate that a normalized sequence charge decoration (SCD) parameter establishes a "soft" criterion for predicting the underlying phase transition of a model protein. Additionally, we find that this order parameter is strongly correlated to the critical density for phase separation, highlighting an unambiguous connection between sequence distribution and condensed phase density. Results obtained from an analysis of the order parameter reveals that at sufficiently long chain lengths, the vast majority of sequences are likely to phase separate. Our results predict that classical LLPS should be the dominant phase transition for disordered proteins and suggests a possible reason behind recent findings of widespread phase separation throughout living cells.

PACS numbers: Valid PACS appear here
Keywords: Suggested keywords

I. INTRODUCTION

Liquid-liquid phase separation (LLPS) of proteins is understood to be a universal biophysical mechanism for the organization and regulation of the intracellular environment.[1–3] Phase separated assemblies of proteins and RNA/DNA, also known as biological condensates, have been implicated in many key biomolecular processes such as cellular signalling[4], ribosomal assembly[5] and transcription of genes[6]. LLPS is often driven by multivalent proteins which act as poylmeric scaffolds that enable the formation of weakly transient networks of noncovalent bonds.[7] Disorder in protein conformation is also known to play a major role in the formation of these condensates and a large majority of phase separating proteins are known to have intrinsically disordered regions (IDRs).[8,9] The underlying driving forces include hydrophobic[10] or electrostatic[11] interactions and can be regulated by changes in temperature[12], pH[13], RNA concentration[14], salt concentration[15] as well as the surrounding intracellular environment.

Protein phase separation is highly sensitive to changes in the underlying protein sequence. Performing point mutations at key residue sites is known to disrupt phase separation.[16,17] Additionally, the sequence patterning of the protein is relevant to its phase separation propensity.[11] Both analytical theory[18–21] and explicit chain simulations[22–24] have been utilized to investigate this sequence-dependent phase behavior. Different sequence-based order parameters such as the sequence charge decoration[25], sequence hydropathy decoration[26] and κ parameter[27] have been proposed which correlate the protein sequence and its structural properties (radius of gyration) or phase behavior (critical temperature). In addition to forming through phase separation, biological condensates are also known to form via gelation[28] or aggregation.[29] Although the sequence determinants driving protein phase separation have been the subject of extensive investigation, it remains unclear how protein sequence dictates the formation of these alternative phase morphologies, a question of potential significance for both native and de novo engineered condensates.[30]

Recently, highly coarse-grained simulations, in which the protein is modelled as an associative polymer, have emerged as a powerful tool for probing the general principles underlying the sequence dependent phase separation of proteins.[31–35] Associative polymers can have strongly sequence-dependent phase behavior; depending on their architecture, they may also form a variety of different finite-size aggregates ranging from near-spherical micelles to bilayers, instead of exhibiting classic first order phase separation.[36–39] Despite the huge diversity of protein sequences, examples of such aggregates maybe quite rare in healthy cells – phase separated condensates appear to be vastly more common.[40] The reason behind this apparent preponderance of phase separated protein morphologies in biology remains unexplained.

In this work, we have utilized grand canonical Monte Carlo (GCMC) simulations to investigate the connections between protein sequence and the type of phase transition that occurs. GCMC simulations, used alongside standard histogram reweighting techniques, can un-

[a)] email: azp@princeton.edu

ambiguously characterize the nature of a phase transition and distinguish macroscopic phase separation from the formation of finite-size aggregates.[41] Using a coarse-grained lattice model of proteins with purely hydrophobic interactions, we study the influence of sequence composition, patterning and chain length on the nature of the phase transition. By characterizing a data set of 100 model sequences, we show that a suitably normalized sequence charge decoration metric (SCD) works remarkably well at predicting the nature of the transition. For a range of different sequence compositions and chain lengths, we map out the critical value of the normalized SCD and show that phase separation becomes the dominant mode of phase transition for sufficiently long chains. We hypothesize that this size effect could be the reason behind the ubiquity of biological phase separation. Finally, we demonstrate that the normalized SCD is strongly correlated to the critical density, illustrating a fundamental connection between sequence patterning and condensed phase properties.

II. MODEL AND METHODS

A. Model for Proteins

In this work, we use a coarse-grained lattice model where the proteins are represented as polymers comprised of two types of entities, hydrophobic/sticky beads which have a net attractive interaction and repulsive hydrophilic beads. Each bead can only occupy a single lattice site and any unoccupied lattice sites are considered to filled by an implicit solvent. Similar lattice models have been extensively used for investigating protein folding and self-assembly.[42,43] In accordance with conventions used in surfactant literature, we refer to the hydrophobic/sticker segments as "tail" (T) beads and hydrophilic segments as "head" (H) beads. In subsequent figures, hydrophobic beads are represented with red circles and hydrophilic beads with blue circles. Both bonded and non-bonded interactions between neighboring beads can be along the relative position vectors (0,0,1), (0,1,1), (1,1,1) and vectors generated by symmetry operations on this set along the principal axes. This produces a lattice with a coordination number of $Z = 26$. We set the hydrophobic tail beads to have an attractive interaction of $\epsilon_{TT} = -1$, which also sets the energy scale for the temperature. All other interactions (specifically HH and HT) are set to zero.

B. Histogram Reweighting Monte Carlo Simulations

GCMC simulations with histogram reweighting were used to investigate the phase behavior of model sequences. Initial runs were performed at a chosen set of temperatures and chemical potentials to obtain the energy and density histograms at those conditions. For a simulation performed at a temperature T and chemical potential μ in a system with volume V, the entropy function at these conditions can be written in terms of the probability of occurrence $f(N,U)$ of N particles with a total energy U in the system up to a run specific additive constant C.

$$S(N,V,U)/k_B = \ln f(N,U) - \beta\mu N + \beta U + C \qquad (1)$$

Multiple histograms can be combined using the Ferrenberg-Swendsen[44,45] method to determine the entropy function of the system across a range of temperatures and chemical potentials. This global entropy function can be utilized to obtain thermodynamic properties of the system at any temperature and chemical potential, given the initial simulation data spanned the range of energies and densities relevant for the new conditions.

C. Distinguishing phase separation and aggregation

To characterize the nature of the transition, we utilized the system-size dependence of the calculated coexistence curves.[46] Sequences which undergo a conventional first order phase transition into macroscopic liquid phases have a coexistence curve which is independent of the system size (upper half of Fig. 1). However, for sequences which aggregate, the apparent coexistence curve shows a strong system size dependence (lower half of Fig. 1). Upon increasing the size of the simulation box, there is an apparent decrease in dense phase concentration. This apparent system size dependence can be attributed to the fact that the system forms finite-sized aggregates. Thus, when the system size is increased, the size of the aggregate formed remains unchanged, leading to an apparent reduction in density. This signature of finite aggregation can also be observed from the probability histograms of the density at coexistence as shown in Fig. S1 of the supplementary material.

We note that aggregation in our model refers to a morphological transformation leading to the formation of a finite aggregate and does not imply irreversibility. Fig. 2 shows snapshots of the dense phase morphology for a phase separating and an aggregating sequence, taken at the same temperature. Importantly, the snapshots illustrate that while the underlying transitions are fundamentally different, there are no major morphological differences between the two dense phases particularly when operating at relatively small system sizes. Thus, it is hard or impossible to distinguish between true phase separation and formation of finite-size aggregates by visual inspection of the simulation box contents alone.

D. Estimating critical parameters

For phase separating sequences, we obtained the critical temperature and density using mixed-field finite-size

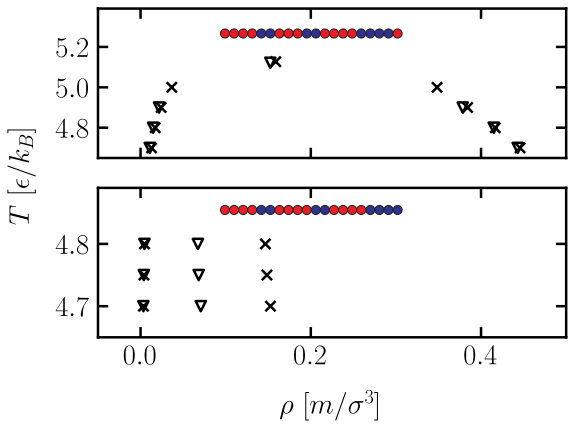

FIG. 1. Dense and dilute phase concentrations for $T_4H_2T_3H_2T_4H_4T$ (top) and $T_4H_2T_4H_2T_4H_4$ (bottom) with simulations performed in systems of size $L = 20\sigma$ (shown as crosses) and $L = 30\sigma$ (shown as inverted triangles). A common density axis is used to highlight the difference in the dense phase concentrations for the two sequences.

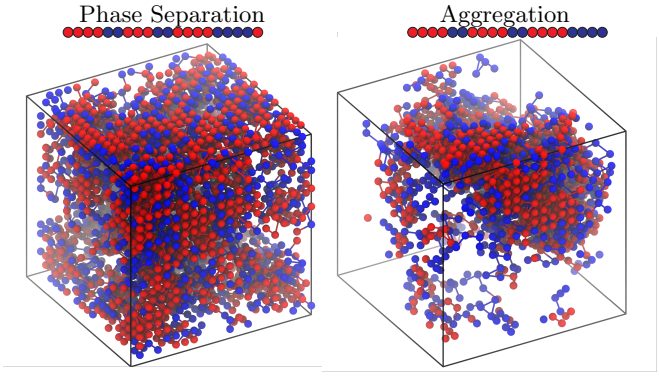

FIG. 2. Snapshots of the dense phase morphologies for two sequences, $T_4H_2T_3H_2T_4H_4T$ (phase separates) and $T_4H_2T_4H_2T_4H_4$ (aggregates). Both of these snapshots were taken at a reduced temperature of $T = 4.8$ in a box of size $20\sigma \times 20\sigma \times 20\sigma$. The corresponding phase diagrams for these sequences are shown in Fig. 1. Hydrophobic tail beads are shown in red while hydrophilic head beads are colored blue. The snapshots were generated using VMD.[50]

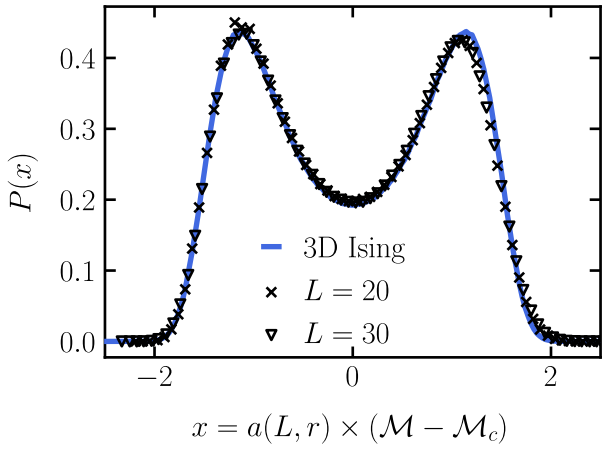

FIG. 3. The scaled order parameter distribution from simulation data (shown in symbols) is matched to the universal curve for the 3D Ising universality class (shown in solid line). The sequence used here is $T_4H_2T_3H_2T_4H_4T$, with simulations performed in two different system sizes.

scaling methods.[47–49] In this approach, an ordering operator $\mathcal{M} = N - sE$ is defined which couples the number of particles N to the configurational energy E using the field-mixing parameter s. For a fixed system size, at criticality, the probability distribution of the scaled ordering parameter $x = a(L, r) \times (\mathcal{M} - \mathcal{M}_c)$ assumes a universal form that depends on the universality class of the underlying first order transition; liquid-liquid phase separation belongs to the three-dimensional Ising universality class. The non-universal parameter $a(L, r)$ is set to rescale the distribution to unit variance. To obtain the probability distribution of the ordering parameter, we perform a set of GCMC runs near the critical point which are then combined using the Ferrenberg-Swendsen method.[44] These distributions can be fitted (shown in Fig. 3) to the universal distribution to obtain an estimate for the critical temperature T_c and the critical chemical potential μ_c.

III. RESULTS AND DISCUSSION

A. Effects of sequence on phase behavior

To investigate the influence of the polymer sequence on the nature of the transition, we first characterized the phase behavior of chains with a fixed sticker fraction f_T and chain length r but having distinct sequence patterning. Five different values of $f_T = 0.4, 0.5, 0.6, 0.75, 0.8$ were considered with chain length $r = 20$. The sequences studied for this chain length are listed in Table I.

We observed that as the degree of dispersion of stickers in the sequence was reduced by clustering them together into longer blocks, the propensity to phase separate decreased. When the dispersion of stickers is reduced beyond a certain point, sequences start showing aggregation behavior and lose the ability to phase separate. These findings are consistent with experimental results which show that phase separation propensity is weakened upon clustering hydrophobic or aromatic "sticky" residues.[51,52]

Furthermore, we found that the transition from phase separation to aggregation depends sensitively on the specific patterning of a sequence. A particularly striking example is seen at $f_T = 0.6$ for the two sequences $T_4H_2T_3H_2T_4H_4T$ and $T_4H_2T_4H_2T_4H_4$. These two sequences have near identical patterns with the only difference being the position of a single T bead. The dilute and dense phase concentrations for these two sequences as a function of temperature are shown in Fig. 1 and

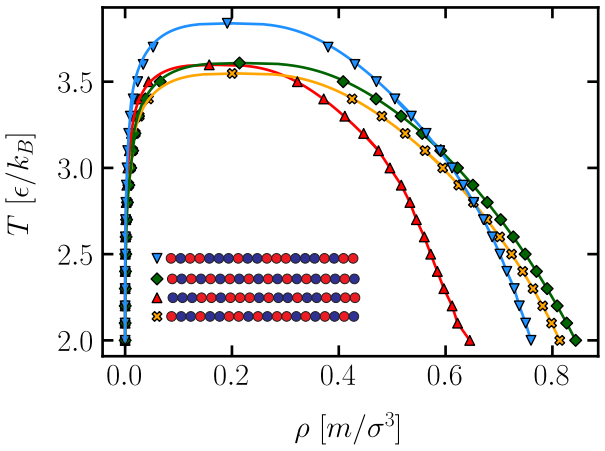

FIG. 4. Coexistence curves for sequences with chain length $r = 20$ and sticker fraction $f_T = 0.5$. The lines are obtained by fitting the near critical coexistence data to the law of rectilinear diameters and the universal scaling relation for densities.[53]

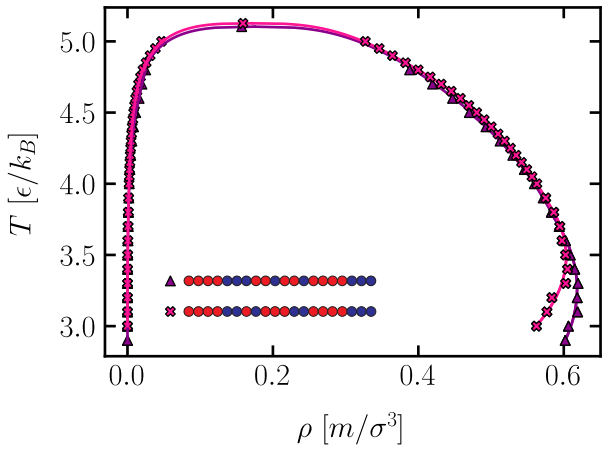

FIG. 5. Coexistence curves for reentrant sequences with chain length $r = 20$ and sticker fraction $f_T = 0.6$. The lines connecting the coexistence points are obtained by fitting the near critical coexistence data to the law of rectilinear diameters and the universal scaling relation for densities. The lines connecting the reentrant points are obtained from a quadratic fit.

snapshots of the dense phase morphologies at $T = 4.8$ are shown in Fig. 2. We also found that among the set of the phase separating sequences for a given (f_T, r), the sequence patterning also influences the critical properties and the shape of the phase envelope (shown in Fig. 4). Additional coexistence data is shown in Fig S2 of the supplementary information.

In addition to sequences showing conventional phase separation and aggregation behavior, we also observed that for $f_T = 0.6$, certain sequences show a "reentrant" transition at low temperatures, with the concentration of the dense phase *decreasing* at lower temperatures, as shown in Fig. 5. Thus, for these reentrant sequences, the condensed phase density reaches a maximum at some intermediate temperature. We find that this anomalous decrease is driven by microphase separation of the hydrophobic sticky blocks at colder temperatures leading to the emergence of voids in the condensed phase; similar behavior has been observed in continuum chain models involving stickers and spacers.[24,34] In Fig. 6, dense phase morphologies are shown at two different temperatures for the reentrant sequence $T_4H_3T_2HT_2HT_4H_3$. At $T = 4.0$, the dense phase is observed to be relatively homogeneous with no clear substructure. However, at $T = 3.0$, we see the formation of a lamellar morphology with clear evidence of microphase separation.

Recent experimental results have implicated reentrant phase transitions for driving the formation of core-shell type morphologies commonly seen in biological condensates.[54,55] Although direct analogies cannot be made due to the inherent simplicity of our model, we speculate that the underlying principles might be similar. While it would be interesting to investigate what happens to the density of the condensed phase of these reentrant sequences as we continue to lower the temperature, equilibration becomes extremely difficult at even lower temperatures; thus we restrict ourselves to temperatures at which we are able to equilibrate our systems with certainty.

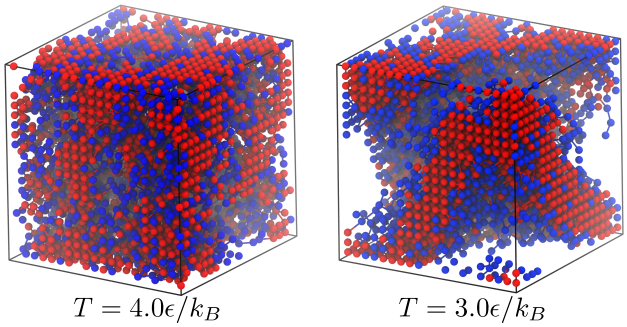

FIG. 6. Dense phase morphologies for the reentrant sequence $T_4H_3T_2HT_2HT_4H_3$ at T=$4.0\epsilon/k_B$ and $T = 3.0\epsilon/k_B$. These snapshots were taken in a box of size $20\sigma \times 20\sigma \times 20\sigma$. The corresponding phase diagram for these sequences is shown in Fig. 5 (legend symbol: pink cross). The snapshots were generated using VMD.[50]

B. Normalized SCD: distinguishing phase separation and aggregation

Although there is a clear empirical connection between the sequence patterning and resulting phase behavior, we sought to develop a more quantitative link by establishing a predictive order parameter for the nature of the transition. To do this, we first augmented our data set by further characterizing the phase behavior of

sequences with chain length $r = 40$ having sticker fractions $f_T = 0.4, 0.5, 0.6, 0.8$ and $r = 100$ with $f_T = 0.5$. Data for the sequences studied are shown in Table S1 and S2 in supplementary information. We then tested a set of order parameters previously proposed in protein and polymer literature which have been correlated with structural or condensed phase properties, specifically: 1) sequence charge decoration (SCD)[25], 2) κ parameter[27], 3) sequence hydropathy decoration[56] and 4) the mean square fluctuation of block hydrophobicity Ψ[57].

Among the tested parameters, we observed that the sequence charge decoration (SCD) metric was the one most capable of distinguishing the nature of the transition, performing well across different chain lengths and overall sticker fractions. The SCD was originally developed to capture the effect of charge patterning in polyampholytes and measures the degree of dispersion of a residue in a protein sequence.[25] In this work we have adapted it to measure the patterning of hydrophobic residues instead. The SCD is defined as:

$$SCD = \frac{1}{N}\sum_{i=2}^{N-1}\sum_{j=1}^{i-1}\sigma_i\sigma_j\sqrt{i-j} \tag{2}$$

where N is the total length of the chain, i and j refer to positions along the chain and σ_i is the identity of the i-th bead. In this work, we have used $\sigma = 1$ for a hydrophobic bead and $\sigma = -1$ for a hydrophilic bead. Using this definition, we find that for each (f_T, r) pair, there exists a "soft" threshold value of the SCD beyond which aggregation becomes the dominant behavior.

An undesirable feature of the SCD parameter is that the range of possible SCD values is a strong function of (f_T, r) making global comparisons difficult. To enable comparison across different values of sticker fraction f_T and chain length r, we modified the SCD parameter by normalizing it according to the definition:

$$\Omega = \frac{SCD_{max}(f_T, r) - SCD}{SCD_{max,r}(f_T, r) - SCD_{min,r}(f_T, r)} \tag{3}$$

where $SCD_{max}(f_T, r)$ and $SCD_{min}(f_T, r)$ are the maximum and minimum possible SCD values for sequences with sticker fraction f_T and chain length r. The sequences having the maximum and minimum SCD values are the least and most "blocky" sequences, respectively. This definition simply rescales the value of the SCD between 0 and 1 with the most blocky sequence having $\Omega = 1$ and the least blocky sequence having $\Omega = 0$.

As previously mentioned, even though Ω performs reasonably well as a predictive order parameter, it is not a perfect metric. Ω is invariant with residue inversion i.e changing all T beads to H beads and vice-verse will not affect Ω. This becomes prominent for $f_T = 0.5$ sequences, since every sequence with 50% stickers also has a complement, both of which have identical Ω values. Thus Ω cannot distinguish between a $f_T = 0.5$ sequence with terminal tail beads and its complement which has terminal head beads. This is problematic because sequences having terminal tail beads are known to have a stronger propensity to phase separate. We observe this near the phase separation threshold where the effect of the terminal beads becomes most pronounced. We also find that Ω has slightly weaker performance at $f_T = 0.4$ with occasional mispredictions seen even far away from the threshold.

C. Phase separation thresholds

1. Influence of sticker fraction and chain length

Having established Ω as a soft order parameter, we proceed to define a threshold value for the onset of aggregation. The threshold value Ω^* was defined as the average Ω of the two sequences on either side of the transition. Intuitively, we expected that at very low values of the sticker fraction, only the most dispersed chains will be able to phase separate and thus $\Omega^* \approx 0$ for low f_T. Conversely, at high values of f_T, all but the most blocky sequences will phase separate, so $\Omega^* \approx 1$. In Fig. 7 we show that the variation of Ω^* with sticker fraction has the expected scaling near the end points. Additionally, for intermediate sticker fractions, we find that the Ω^* has a roughly quadratic dependence on f_T.

The qualitative dependence of Ω^* on f_T is robust across chain length for different sticker fractions but the exact dependence of Ω^* on chain length remains unclear. To probe this, we computed Ω^* for sequences having $f_T = 0.5$ across chain lengths $r = 20, 40$ and 100. We find that Ω^* decreases monotonically with r and reaches an asymptotic (non-zero) value as $1/r \to 0$, shown in Fig. 8. From a linear regression, we obtained a $\Omega^* = 0.008 \pm 0.003$ at the infinite chain limit. Taken together, our results establish a global predictive order parameter for the phase behavior of protein sequences for this model, which is robust across different chain lengths and sticker fractions.

2. Sequence space statistics

Given that we now have an understanding of the threshold $\Omega = \Omega^*$ between phase separation and aggregation, an interesting question to pose is what fraction of possible sequences at a given (f_T, r) lie below the aggregation threshold Ω^*. To do this, we first obtained the sequence-space statistics by generating the probability distribution of Ω for different combinations of chain length and sticker fraction. Probability distributions were estimated by generating 10^6 random sequences for each (f_T, r) and computing the corresponding Ω for each sequence. The fraction of sequences that phase separate was then estimated by integrating the probability distribution up to Ω^*.

TABLE I. Sequence architecture, sticker fraction, normalized SCD Ω and phase separation capability for sequences of length $r = 20$

f_T	Sequence	Ω	PS
0.4	$HTH[TH_2]_4[TH]_2T$	0.060	✓
	$T_2H_4TH_2T_2H[TH]_3H_2$	0.069	×
	$[TH_2]_5[TH]_2T$	0.074	✓
	$[HT]_2[H_2T]_5T$	0.082	✓
	$[TH]_3[HT]_2H_4[TH]_2HT$	0.091	✓
	$HTH_5T_4HT_2H_3TH_2$	0.092	×
	$TH_3T_3H[TH_2]_2HTH_3T$	0.097	×
	$H_2T_3[HT]_2H_5THTH_2T$	0.108	×
	$H_2TH_3TH_2THT_5H_4$	0.115	×
	$HT_3H_2TH_3THTH_4THT$	0.122	✓
0.5	$[TH]_{10}$	0.000	✓
	$[TH_2]_2T_2HTHT_2[HT]_3H$	0.018	✓
	$THT_2H_3TH_2T_3H_3THT_2$	0.054	✓
	$H_3T_2HT_3HTH_2TH_2THT_2$	0.071	✓
	$H_2T_2H_2TH_2THT_2HTH_2T_3$	0.094	✓
	$HTH_3T_3H_2T_3HT_3H_3$	0.099	×
	$H_2T_3H_2TH_2TH_2TH_2T_4$	0.120	×
	$T_4H_2T_2H_3TH_2T_2H_2TH$	0.145	×
	$T_4H_3T_4H_5TH_2T$	0.218	×
	$THT_2HT_4HTHTH_5TH$	0.266	×
0.6	$T_3H_3T_3H_2T_3H_3T_3$	0.045	✓
	$THTH_3T_4HT_3HTHT_2H$	0.107	✓
	$HTHTH_2T_2HTHT_4H_2T_3$	0.123	✓
	$T_4H_2T_2HT_3HTH_2THTH$	0.136	✓
	$T_4H_3T_2HT_2HT_4H_3$	0.147	✓
	$T_4H_2THT_3H_2T_4H_3$	0.156	✓
	$T_4H_2T_3H_2T_4H_4T$	0.167	✓
	$HT_2HTHT_7H_4THT$	0.190	×
	$T_3H_2T_5HT_2H_2TH_2TH$	0.207	×
	$TH_4T_2HT_7H_2THT$	0.220	×
	$HTH_4T_5HT_2HTHT_3$	0.237	×
	$T_6H_2T_3H_2TH_2THTH$	0.244	×
0.75	$T_4HT_2HT_2HTHT_2HT_4$	0.000	✓
	$THT_3HTHT_2HT_8H$	0.193	✓
	$THTHT_9HTHT_3H$	0.290	✓
	$T_2HT_2HT_9HT_2H_2$	0.382	✓
	$HTHT_{12}HTH_2T$	0.472	✓
	$T_{10}HT_3HTHTH_2$	0.566	×
	$H_2THTHT_2HT_{11}$	0.603	×
	$H_2TH_2T_4HT_{10}$	0.621	×
	$H_2THTHTHT_{12}$	0.649	×
0.8	$T_8H_4T_8$	0.098	✓
	$H_2T_2HT_{13}HT$	0.627	✓
	$H_2T_{16}H_2$	0.760	✓
	$H_3T_2HT_{14}$	0.847	×
	H_4T_{16}	1.000	×

Fig. 9 shows the probability distributions of Ω for chain length $r = 40$. Vertical arrows in the figure indicate the threshold values Ω^* for the different fraction of stickers f_T. The fraction of phase separating sequences increased monotonically with sticker fraction going from 33% at $f_T = 0.4$ to 78% at $f_T = 0.6$, in line with the expectation that phase separation should be enhanced with the addition of sticky residues. We then investigated how increasing the length of the chain (at fixed $f_T = 0.5$) influences the propensity to phase separate. For $r = 20, 40$ and 100, we compute the fraction of phase separating sequences as 49.9%, 43.6% and 57.0% respectively (Fig. 10). The apparent decrease at $r = 40$ is likely due to inaccuracy in our measurement of Ω^*, and we hypothesize that the fraction of phase separating sequences increases monotonically with chain length. To test our hypothesis, we used the chain length dependence of Ω^* at $f = 0.5$,

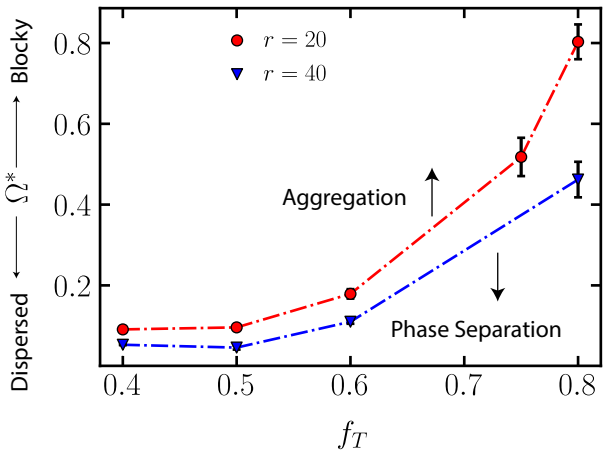

FIG. 7. Scaling of the threshold for phase separation Ω^* with sticker fraction at a fixed chain length. The dashed lines are meant to guide the eye and demarcate the regimes of phase separation and aggregation for a given chain length. Uncertainties were estimated by measuring differences in the Ω between the two sequences at the boundary between phase separation and aggregation.

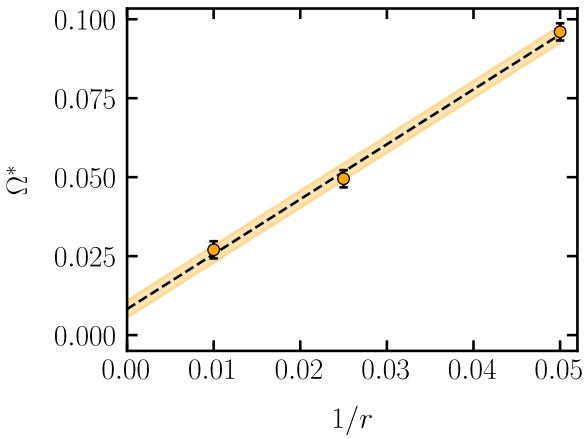

FIG. 8. Scaling of the threshold for phase separation Ω^* with the inverse chain length for sequences with constant sticker fraction $f_T = 0.5$. The dotted line represents a linear fit, which is extrapolated to infinite chain length. The shaded area represents the statistical uncertainty measured as the standard error of the fit.

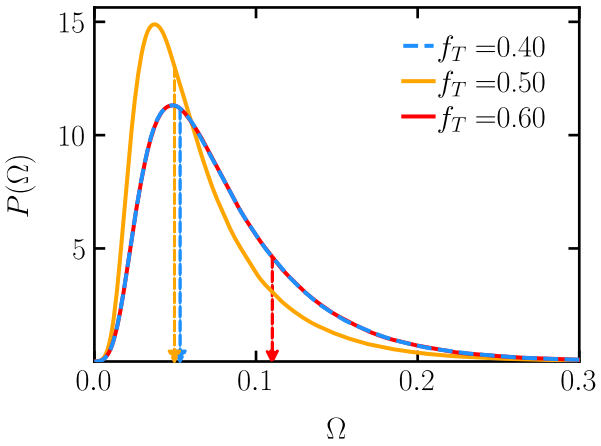

FIG. 9. Probability distributions of the normalized SCD Ω for different f_T at chain length $r = 40$. The vertical arrows indicate the threshold value Ω^*. Due to the symmetry of Ω, the distributions for $f_T = 0.4$, shown in blue, and $f_T = 0.6$, shown in red, are identical but their corresponding Ω^* is different.

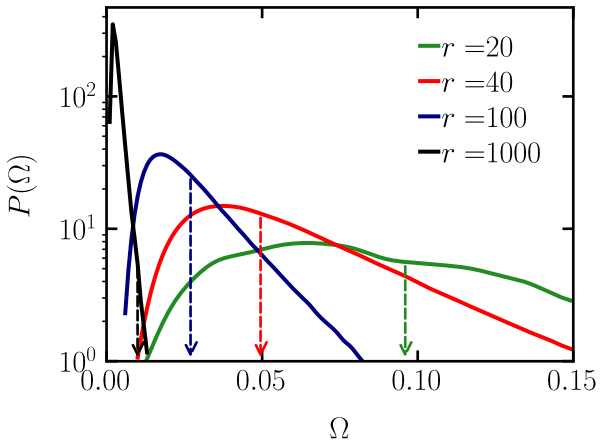

FIG. 10. Probability distributions of the normalized SCD Ω for different r at sticker fraction $f_T = 0.5$. The vertical arrows indicate the threshold value Ω^*.

as shown in Fig. 8, to obtain the fraction of phase separating sequences at a chain length of $r = 1000$. We find that $97.4\% \pm 2.3\%$ of all sequences having $r = 1000$ and $f_T = 0.5$ are expected to phase separate. Thus, there is a very clear increase in the fraction of phase separating sequences for longer chains, with the vast majority of possible sequences capable of phase separating.

This remarkable and unexpected result of a sharp increase in the fraction of phase separating sequences at long chain lengths could have important biological consequences. The ubiquity of biological condensates has been a rather puzzling question. For chains lengths comparable to typical proteins in the cells, our results predict that phase separated morphologies should be overwhelmingly common, with only a tiny fraction of sequences showing aggregation into finite clusters. Our result is consistent with existing experimental evidence and could have important implications regarding the possible phase separation of other long biopolymers such DNA and RNA.

D. Dependence of critical properties on sequence

In the previous section, we demonstrated that Ω, the normalized SCD parameter, performs well for distinguishing the phase behavior of model sequences. Additionally, for sequences that phase separate, our findings in Sec. III A showed that their coexistence curves were strongly sequence dependent. Here, we investigate whether the dependence of the critical temperature and density on sequence composition and patterning can be rationalized using Ω as the control parameter.

In Fig. 11 and Fig. 12, we show the critical temperatures and densities of sequences having chain length

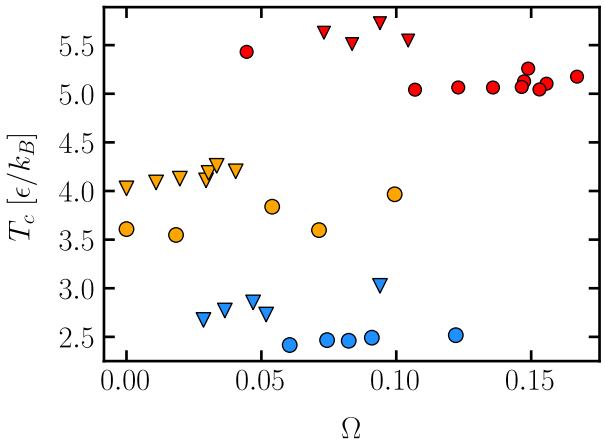

FIG. 11. Scaling of the critical temperature with the normalized SCD Ω. The symbol shape represents the chain length: circle $r = 20$, inverted triangle $r = 40$. The color of the symbol is used to represent the fraction of stickers in the chain: $f_T = 0.6$ in red, $f_T = 0.5$ in orange and $f_T = 0.4$ in blue.

$r = 20$ and 40 with sticker fractions $f_T = 0.4, 0.5$ and 0.6 against the normalized SCD Ω. We find that the critical temperature of sequences is largely decided by the sticker fraction with the precise distribution of stickers in the sequence only seeming to cause small perturbations around this average value. In addition, we also observed sequences at the very edge of phase separation, $\Omega \approx \Omega^*$, have a systematically higher T_c than sequences further away from the aggregation threshold.

In contrast, the critical density shows a strong negative correlation with the normalized SCD. For both $r = 20$ and $r = 40$, we observe the critical density decreases almost monotonically with Ω until the threshold Ω^* is reached. Additionally, for a fixed sticker fraction, the critical density decreases linearly with Ω as seen in Fig.

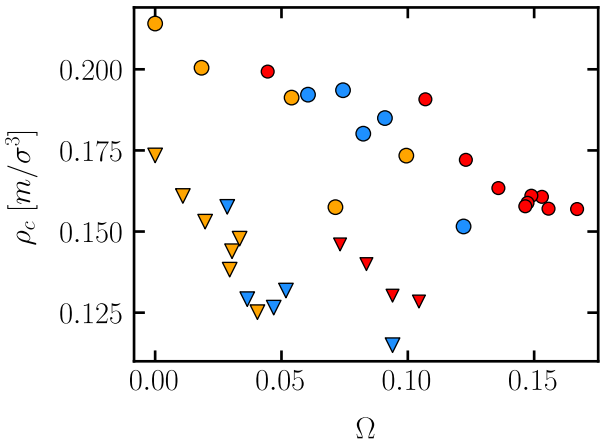

FIG. 12. Scaling of the critical density with the normalized SCD Ω. The symbol shape represents the chain length: circle $r = 20$, inverted triangle $r = 40$. The color of the symbol is used to represent the fraction of stickers in the chain: $f_T = 0.6$ in red, $f_T = 0.5$ in orange and $f_T = 0.4$ in blue.

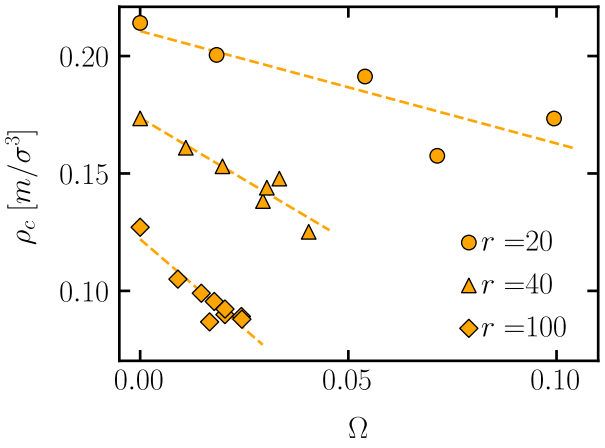

FIG. 13. Scaling of the critical density with the normalized SCD Ω for sequences with constant sticker fraction $f_T = 0.5$ across different chain lengths. The symbol shape represents the chain length.

13. The slope of this line, the fractional change in ρ_c with Ω, increases upon going from $r = 20$ to $r = 40$ and then stays approximately constant upon increasing the chain length further to $r = 100$. Thus we conclude that as the blockiness of the sequence is increased, the density of the condensed phase decreases monotonically until it reaches a minimum value at $\Omega = \Omega^*$. Below this threshold, the sequence becomes prone to aggregation into finite structures.

IV. CONCLUSIONS

In this work, we have investigated how the sequence patterning of model proteins influences their phase behavior. We found that model proteins can either phase separate or aggregate into clusters of finite extent, depending sensitively on the precise sequence patterning. GCMC simulations combined with histogram reweighting and mapping of a normalized order parameter distribution to the universal Ising curve were found to be sensitive tools to discriminate between phase separation and aggregation and to obtain precise values of the critical parameters. Among the phase separating sequences, we observed that certain sequences exhibit a reentrant transition, with the concentration of protein in the dense phase decreasing as temperature is lowered. This behavior is associated with microphase separation within the condensed phase.

From the characterized phase behavior of 100 different sequences, we found that a normalized sequence charge decoration metric Ω is able to broadly distinguish phase separating from aggregating sequences of the model proteins. Thus, there exists a threshold value Ω^*, beyond which the ability to phase separate into a macroscopic phase is lost and sequences become aggregation prone. Although we have focused on the relation between sequence blockiness and finite-size aggregation in this work, experiments also suggest a potential link between clustering of residues and the propensity of forming irreversible protein aggregates.[58] Further theoretical and experimental efforts will be needed to investigate this connection.

Using the normalized SCD Ω, we found that at a constant chain length, the threshold normalized SCD Ω^* has an approximately quadratic dependence on the sticker fraction (hydrophobicity) f_T. At a constant sticker fraction, our results show that Ω^* scales linearly with inverse chain length and reaches an asymptotic non-zero value at infinite chain length. Since the Ω is intrinsically related to the overall blockiness of the sequence, our result establishes a robust connection between blockiness in the sequence patterning and its underlying phase behavior. In addition to hydrophobic patterning, charge patterning is also known to play an important role in driving protein LLPS. However, unlike hydrophobic residues, clustering of charges is seen to enhance phase separation tendency.[11] Investigating the cumulative effects of charge and hydrophobic patterning will be necessary to develop a complete picture of sequence dependent protein phase behavior.

To estimate what fraction of sequences of a certain length and sticker fraction are likely to phase separate, we obtained the sequence space statistics by calculating the distribution of Ω for a given (r, f_T) and utilized this distribution . Our results show that the fraction of phase separating sequences increases monotonically with sticker fraction at a constant chain length. The variation with chain length was found to be nearly monotonic with a relatively minor change in the fraction of phase separat-

ing sequences when going from chain length $r = 20$ to $r = 100$. However, for $r = 1000$, we found a dramatic increase in the fraction phase separated, with 98% of possible sequences predicted to phase separate.

From our results, we conclude that the phase separation propensity increases rapidly as a function of chain length. Our findings suggest that at sufficiently long chain lengths, the vast majority of possible sequences will phase separate irrespective of sticker fraction or sequence patterning. We hypothesize that the ubiquity of biological phase separation may simply be tied to the fact that most biologically relevant proteins are sufficiently long to be in the regime where phase separation becomes dominant. This would also explain why finite aggregation behavior is relatively rare in biology despite the huge diversity of possible protein sequences.

ACKNOWLEDGMENTS

This research was primarily supported by the Princeton Center for Complex Materials (PCCM), a U.S. National Science Foundation Materials Research Science and Engineering Center (Grants No. DMR- 1420541 and DMR-2011750). The authors thank Antonia Statt for valuable comments and discussions. Simulations were performed using computational resources provided by the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]Y. Shin and C. P. Brangwynne, "Liquid phase condensation in cell physiology and disease," Science **357** (2017).

[2]S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, "Biomolecular condensates: Organizers of cellular biochemistry," (2017).

[3]S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. Van Den Bosch, P. Tompa, and M. Fuxreiter, "Protein phase separation: A new phase in cell biology," Trends Cell Biol. **28**, 420–435 (2018).

[4]X. Su, J. A. Ditlev, E. Hui, W. Xing, S. Banjade, J. Okrut, D. S. King, J. Taunton, M. K. Rosen, and R. D. Vale, "Phase separation of signaling molecules promotes T cell receptor signal transduction," Science **352**, 595–599 (2016).

[5]J. A. Riback, L. Zhu, M. C. Ferrolino, M. Tolbert, D. M. Mitrea, D. W. Sanders, M.-T. Wei, R. W. Kriwacki, and C. P. Brangwynne, "Composition-dependent thermodynamics of intracellular phase separation," Nature **581**, 209–214 (2020).

[6]L. Peng, E.-M. Li, and L.-Y. Xu, "From start to end: Phase separation and transcriptional regulation," Biochim. Biophys. Acta Gene Regul. Mech. **1863**, 194641 (2020).

[7]P. Li, S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. X. Jiang, B. T. Nixon, and M. K. Rosen, "Phase transitions in the assembly of multivalent signalling proteins," Nature **483**, 336–340 (2012).

[8]H.-X. Zhou, V. Nguemaha, K. Mazarakos, and S. Qin, "Why do disordered and structured proteins behave differently in phase separation?" Trends Biochem. Sci. **43**, 499–516 (2018).

[9]A. Garaizar, I. Sanchez-Burgos, R. Collepardo-Guevara, and J. R. Espinosa, "Expansion of intrinsically disordered proteins increases the range of stability of Liquid–Liquid phase separation," Molecules **25**, 4705 (2020).

[10]A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal, and N. L. Fawzi, "Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain," Nat. Struct. Mol. Biol. **26**, 637–648 (2019).

[11]T. J. Nott, E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, J. D. Forman-Kay, and A. J. Baldwin, "Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles," Mol. Cell **57**, 936–947 (2015).

[12]J. A. Riback, C. D. Katanski, J. L. Kear-Scott, E. V. Pilipenko, A. E. Rojek, T. R. Sosnick, and D. A. Drummond, "Stress-Triggered phase separation is an adaptive, evolutionarily tuned response," Cell **168**, 1028–1040.e19 (2017).

[13]T. M. Franzmann, M. Jahnel, A. Pozniakovsky, J. Mahamid, A. S. Holehouse, E. Nüske, D. Richter, W. Baumeister, S. W. Grill, R. V. Pappu, A. A. Hyman, and S. Alberti, "Phase separation of a yeast prion protein promotes cellular fitness," Science **359** (2018).

[14]M.-T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. C.-H. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, and C. P. Brangwynne, "Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles," Nat. Chem. **9**, 1118–1125 (2017).

[15]G. Krainer, T. J. Welsh, J. A. Joseph, J. R. Espinosa, S. Wittmann, E. de Csilléry, A. Sridhar, Z. Toprakcioglu, G. Gudiškytė, M. A. Czekalska, W. E. Arter, J. Guillén-Boixet, T. M. Franzmann, S. Qamar, P. S. George-Hyslop, A. A. Hyman, R. Collepardo-Guevara, S. Alberti, and T. P. J. Knowles, "Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions," Nat. Commun. **12**, 1085 (2021).

[16]D. Bracha, M. T. Walls, M.-T. Wei, L. Zhu, M. Kurian, J. L. Avalos, J. E. Toettcher, and C. P. Brangwynne, "Mapping local and global liquid phase behavior in living cells using Photo-Oligomerizable seeds," Cell **175**, 1467–1480.e13 (2018).

[17]A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman, and S. Alberti, "A Liquid-to-Solid phase transition of the ALS protein FUS accelerated by disease mutation," Cell **162**, 1066–1077 (2015).

[18]Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan, "Sequence-Specific polyampholyte phase separation in membraneless organelles," Phys. Rev. Lett. **117**, 178101 (2016).

[19]Y.-H. Lin and H. S. Chan, "Phase separation and Single-Chain compactness of charged disordered proteins are strongly correlated," Biophys. J. **112**, 2043–2046 (2017).

[20]T. Firman and K. Ghosh, "Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins," J. Chem. Phys. **148**, 123305 (2018).

[21]X. Zeng, A. S. Holehouse, A. Chilkoti, T. Mittag, and R. V. Pappu, "Connecting Coil-to-Globule transitions to full phase diagrams for intrinsically disordered proteins," Biophys. J. **119**, 402–418 (2020).

[22]G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, "Sequence determinants of protein phase behavior from a coarse-grained model," PLoS Comput. Biol. **14** (2018).
[23]G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, "Relation between single-molecule properties and phase behavior of intrinsically disordered proteins," Proc. Natl. Acad. Sci. U. S. A. **115**, 9929–9934 (2018).
[24]M. K. Hazra and Y. Levy, "Biophysics of phase separation of disordered proteins is governed by balance between short- and Long-Range interactions," J. Phys. Chem. B **125**, 2202–2211 (2021).
[25]L. Sawle and K. Ghosh, "A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins," J. Chem. Phys. **143**, 085101 (2015).
[26]W. Zheng, G. Dignon, M. Brown, Y. C. Kim, and J. Mittal, "Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins," J. Phys. Chem. Lett. **11**, 3408–3415 (2020).
[27]R. K. Das and R. V. Pappu, "Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues," Proc. Natl. Acad. Sci. U. S. A. **110**, 13392–13397 (2013).
[28]A. Putnam, M. Cassani, J. Smith, and G. Seydoux, "A gel phase promotes condensation of liquid P granules in caenorhabditis elegans embryos," Nat. Struct. Mol. Biol. **26**, 220–226 (2019).
[29]T. Yamazaki, T. Yamamoto, H. Yoshino, S. Souquere, S. Nakagawa, G. Pierron, and T. Hirose, "Paraspeckles are constructed as block copolymer micelles," EMBO J. , e107270 (2021).
[30]M. Dzuricky, B. A. Rogers, A. Shahid, P. S. Cremer, and A. Chilkoti, "De novo engineering of intracellular condensates using artificial disordered proteins," Nat. Chem. **12**, 814–825 (2020).
[31]J.-M. Choi, A. S. Holehouse, and R. V. Pappu, "Physical principles underlying the complex biology of intracellular phase transitions," Annu. Rev. Biophys. **49**, 107–133 (2020).
[32]T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, "Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins," Elife **6** (2017).
[33]J.-M. Choi, F. Dar, and R. V. Pappu, "LASSI: A lattice model for simulating phase transitions of multivalent proteins," PLoS Comput. Biol. **15**, e1007028 (2019).
[34]A. Statt, H. Casademunt, C. P. Brangwynne, and A. Z. Panagiotopoulos, "Model for disordered proteins with strongly sequence-dependent liquid phase behavior," J. Chem. Phys. **152**, 075101 (2020).
[35]Y. Zhang, B. Xu, B. G. Weiner, Y. Meir, and N. S. Wingreen, "Decoding the physical principles of two-component biomolecular phase separation," Elife **10** (2021).
[36]M. A. Floriano, E. Caponetti, and A. Z. Panagiotopoulos, "Micellization in model surfactant systems," Langmuir **15**, 3143–3151 (1999).
[37]G. Srinivas, D. E. Discher, and M. L. Klein, "Self-assembly and properties of diblock copolymers by coarse-grain molecular dynamics," Nat. Mater. **3**, 638–644 (2004).
[38]M. E. Gindy, R. K. Prud'homme, and A. Z. Panagiotopoulos, "Phase behavior and structure formation in linear multiblock copolymer solutions by monte carlo simulation," J. Chem. Phys. **128**, 164906 (2008).
[39]V. Hugouvieux, M. A. V. Axelos, and M. Kolb, "Amphiphilic multiblock copolymers: From intramolecular pearl necklace to layered structures," Macromolecules **42**, 392–400 (2009).
[40]M. Hardenberg, A. Horvath, V. Ambrus, M. Fuxreiter, and M. Vendruscolo, "Widespread occurrence of the droplet state of proteins in the human proteome," Proc. Natl. Acad. Sci. U. S. A. **117**, 33254–33262 (2020).
[41]A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano, "Phase equilibria of lattice polymers from histogram reweighting monte carlo simulations," Macromolecules **31**, 912–918 (1998).
[42]E. M. O'Toole and A. Z. Panagiotopoulos, "Monte carlo simulation of folding transitions of simple model proteins using a chain growth algorithm," J. Chem. Phys. **97**, 8644–8652 (1992).
[43]E. M. O'Toole and A. Z. Panagiotopoulos, "Effect of sequence and intermolecular interactions on the number and nature of low-energy states for simple model proteins," J. Chem. Phys. **98**, 3185–3190 (1993).
[44]A. M. Ferrenberg and R. H. Swendsen, "New monte carlo technique for studying phase transitions," Phys. Rev. Lett. **61**, 2635–2638 (1988).
[45]A. M. Ferrenberg and R. H. Swendsen, "Optimized monte carlo data analysis," Phys. Rev. Lett. **63**, 1195–1198 (1989).
[46]A. Z. Panagiotopoulos, M. A. Floriano, and S. K. Kumar, "Micellization and phase separation of diblock and triblock model surfactants," Langmuir **18**, 2940–2948 (2002).
[47]A. D. Bruce and N. B. Wilding, "Scaling fields and universality of the liquid-gas critical point," Phys. Rev. Lett. **68**, 193–196 (1992).
[48]N. B. Wilding, "Critical-point and coexistence-curve properties of the Lennard-Jones fluid: A finite-size scaling study," Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics **52**, 602–611 (1995).
[49]N. B. Wilding, M. Müller, and K. Binder, "Chain length dependence of the polymer–solvent critical point parameters," J. Chem. Phys. **105**, 802–809 (1996).
[50]W. Humphrey, A. Dalke, and K. Schulten, "VMD – Visual Molecular Dynamics," Journal of Molecular Graphics **14**, 33–38 (1996).
[51]E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, and T. Mittag, "Valence and patterning of aromatic residues determine the phase behavior of prion-like domains," Science **367**, 694–699 (2020).
[52]M. A. Bowman, J. A. Riback, A. Rodriguez, H. Guo, J. Li, T. R. Sosnick, and P. L. Clark, "Properties of protein unfolded states suggest broad selection for expanded conformational ensembles," Proc. Natl. Acad. Sci. U. S. A. **117**, 23356–23364 (2020).
[53]K. S. Silmore, M. P. Howard, and A. Z. Panagiotopoulos, "Vapour–liquid phase equilibrium and surface tension of fully flexible Lennard–Jones chains," Mol. Phys. **115**, 320–327 (2017).
[54]I. Alshareedah, M. M. Moosa, M. Raju, D. A. Potoyan, and P. R. Banerjee, "Phase transition of RNA-protein complexes into ordered hollow condensates," Proc. Natl. Acad. Sci. U. S. A. **117**, 15650–15658 (2020).
[55]T. Kaur, M. Raju, I. Alshareedah, R. B. Davis, D. A. Potoyan, and P. R. Banerjee, "Sequence-encoded and composition-dependent protein-RNA interactions control multiphasic condensate morphologies," Nat. Commun. **12**, 872 (2021).
[56]W. Zheng, G. Dignon, M. Brown, Y. C. Kim, and J. Mittal, "Hydropathy patterning complements charge patterning to describe conformational preferences of disordered proteins," J. Phys. Chem. Lett. **11**, 3408–3415 (2020).
[57]A. Irbäck, C. Peterson, and F. Potthast, "Evidence for nonrandom hydrophobicity structures in protein chains," Proc. Natl. Acad. Sci. U. S. A. **93**, 9533–9538 (1996).
[58]T. R. Peskett, F. Rau, J. O'Driscoll, R. Patani, A. R. Lowe, and H. R. Saibil, "A liquid to solid phase transition underlying pathological huntingtin exon1 aggregation," Mol. Cell **70**, 588–601.e6 (2018).

V. SUPPLEMENTARY MATERIAL

A. Distinguishing phase separation and finite aggregation from density histograms

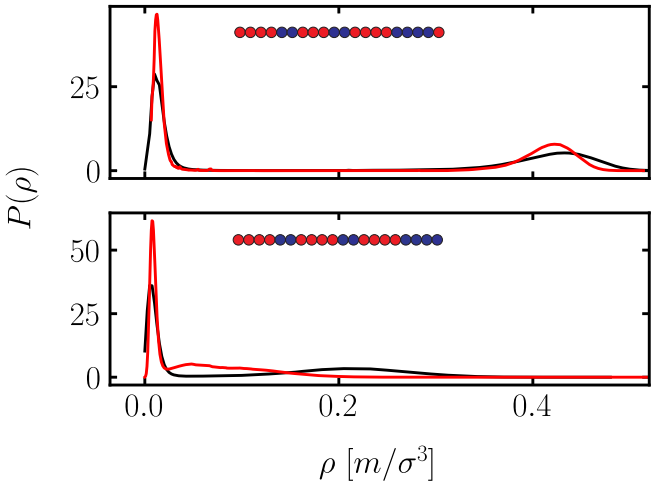

FIG. S1. Histograms of the density at $T = 4.8$ and coexistence chemical potential for $T_4H_2T_3H_2T_4H_4T$ (top) and $T_4H_2T_4H_2T_4H_4$ (bottom) with simulations performed in systems of size $L = 20\sigma$ (black) and and $L = 30\sigma$ (red). For the phase separating sequence $T_4H_2T_3H_2T_4H_4T$, the dilute and dense peaks are invariant with system size. In contrast, there is a shift in the location of the dense phase peak for aggregating sequence $T_4H_2T_4H_2T_4H_4$.

B. Phase diagrams of sequences with $f_T = 0.6$

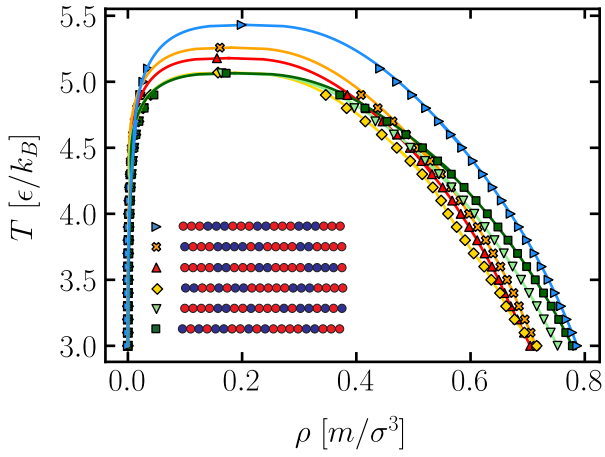

FIG. S2. Coexistence curves for sequences with chain length $r = 20$ and sticker fraction $f_T = 0.6$. The lines are obtained by fitting the near critical coexistence data to the law of rectilinear diameters and the universal scaling relation for densities.

TABLE S1. Sequence architecture, sticker fraction, normalized SCD Ω and phase separation capability for sequences of length $r = 40$

f_T	Sequence	Ω	PS
0.4	$HTHTH_2TH_2TH_2TH_2THTHTHTHTH_2TH_2TH_2TH_2THTHT$	0.028	✓
	$THTH_7T_4H_4THTHTH_2T_2H_2T_2H_3T_2HTH_2$	0.036	✓
	$H_2TH_2TH_2TH_2T_2HTH_3TH_2THTH_4T_3HT_3HTH_3$	0.041	×
	$THTHTH_7T_2HT_2HT_2H_4T_2HTH_5T_3HTH_2$	0.047	✓
	$HTHTHTHT_2HTH_2TH_2T_2H_3TH_5T_3H_3TH_2THTH$	0.051	✓
	$THT_3H_5THTHTH_3T_3H_2T_2H_3THTHTHTH_5$	0.054	×
	$HTH_3THTHTH_2TH_2TH_3THT_2HTH_4TH_3THT_2HT_2$	0.072	×
	$HT_2HT_2HTH_3TH_2TH_2THT_3HTHT_2H_5TH_5TH$	0.081	×
	$THTH_4T_4H_6TH_3TH_3THT_2H_3THT_3H_2T$	0.094	✓
	$HT_2HTHT_2H_2T_2HTHTHTH_3T_2HTH_4T_2H_5TH_3$	0.105	×
	$T_2H_3TH_3TH_8T_3H_2THTHTH_2T_3HTH_2THT$	0.115	×
0.5	$[TH]_{20}$	0.000	✓
	$TH2THT_2H_2T_3H_2THTH_2TH_2T_2HTH_2T_2HTH_2T_3HTH$	0.011	✓
	$T_2H_3T_2H_3T_3H_2THT_2H_3T_2HT_2HT_2H_2T_3H_4T$	0.020	✓
	$H_2T_3HTHT_2H_4TH_2THT_4HTH_2T_4H_4THTHT$	0.029	✓
	$HTH_2T_3HT_3H_2TH_2T_3H_5THT_2H_2TH_2T_3HTHT$	0.030	✓
	$THT_2H_2THT_2H_2T_2H_4TH_3T_4H_3THT_2HTHTHT_2$	0.033	✓
	$THTHT_2HTH_3T_2HT_3H_5T_2H_3THT_6HTH_3$	0.040	✓
	$H_2T_3H_2T_3H_5T_2H_2TH_2T_2H_2THT_2HTHTHTHT_3$	0.047	✓
	$H_2THTH_2THTH_2T_4H_2TH_3T_2HTH_2T_3H_3T_2HT_3$	0.048	✓
	$H_2TH_3T_2HTHTHTHTHT_2HTH_2T_2HT_2H_2TH_4T_5$	0.050	×
	$H_3T_2HTH_2T_2H_3THT_2H_2THTH_2T_5HTHT_3H_3T$	0.056	×
	$H_3T_2HTH_3T_4HTHT_3HT_3HT_2H_5T_2HTHTH_2$	0.061	×
	$T_7H_6THTHTH_2T_2H_2THT_2HTH_2T_2H_3T_2H$	0.071	×
	$TH_3T_2H_2THT_4HT_3HT_2HTHTH_8THT_4H$	0.081	×
	$T_5H_2T_3HTHTH_3TH_4THTHTH_2T_2HTHT_3H_3$	0.089	×
0.6	$H_2THT_2HTHT_2HT_2HT_2HTHT_2H_2T_4HT_4H_3THT_2$	0.062	✓
	$THT_2HTH_2T_2HTH_2T_5HTHTHTHT_6H_2T_3H_3$	0.073	✓
	$TH_2THTHT_2HT_2H_2T_4HTHT_4H_2T_2HT_4HT_2H_3$	0.084	✓
	$HT_3H_2T_2HTHT_2HT_7HT_3H_3TH_4T_3H_2T_2$	0.094	✓
	$TH_2T_3H_5THT_6HTHT_4HTHT_2HT_3HT_2H_2$	0.104	✓
	$T_2HT_2HTHTHT_2HT_4HT_3HT_2HTHTH_2T_4H_2TH_3$	0.115	×
	$H_2TH_2TH_2T_2HT_2HT_4H_2T_4H_4THT_5HT_4$	0.126	×
	$T_2HT_5HT_2HT_4H_2T_2H_2T_2HT_2HTHTHTH_4THT$	0.136	×
	$H_3THTHT_7H_2T_3H_2T_6H_2T_4H_2THTH_2$	0.141	×
0.8	$HTHT_2HT_3HT_{17}HT_4H_2T_2HT_3$	0.351	✓
	$HTH_2T_7HT_{17}HT_3HTH_2T_3$	0.407	✓
	$T_6HT_{16}HT_5H_2T_3H_3T_2H$	0.450	×
	$H_2TH_2T_3HTHT_4HT_{10}HT_{13}$	0.494	×
	$HTH_3T_2HTHT_{22}HT_4HT_2$	0.556	×
	$HTHTH_3THT_{26}HTHT_2$	0.593	×
	$T_2HT_{25}HT_2HTHTH_3TH$	0.648	×

TABLE S2. Sequence architecture, sticker fraction, normalized SCD Ω and phase separation capability for sequences of length $r = 100$

. All sequences have $f_T = 0.5$.

Sequence	Ω	PS
$[TH]_{50}$	0.000	✓
$[TH]_2T_2HTH_3T_2HT_3H_5T_2H_3THT_6HTH_3THTHT_2HTH_3T_2HT_3H_5T_2H_3THT_6HTH_3THT_2H_3TH_2T_3H_3THT_2$	0.009	✓
$T_2HTHTH_2T_3HTHT_2H_3T_4H_5T_3H_2THT_3HTHT_2H_6TH_2T_2H_4T_3HT_3HTH_3TH_2THTHT_7H_2THTH_5THT_3H$	0.015	✓
$TH_3TH_4T_4HT_5HT_2H_2T_3H_3THT_2H_2T_2H_4THTH_2T_2H_3T_5HT_2HTHTHT_2H_2THT_2H_3T_2HTH_5T_2H_2T_4HTH_2TH_2$	0.017	✓
$T_4H_2T_4H_3THTH_3TH_3T_4HT_2H_2THTHTH_4T_3H_3TH_5THT_3H_2THTHT_5HTH_3TH_3THT_5HTH_3TH_2THT_4H$	0.018	✓
$TH_3TH_2T_2H_5T_2HTHT_3H_3T_3H_2T_2HT_3HT_4H_4T_2HTH_2T_2HTHT_4HT_4H_5T_2HTHT_2H_5T_3HTHT_2H_2TH_2TH_2TH$	0.020	✗
$T_3H_2TH_5TH_5T_2HTHT_3HT_3HTHT_4HTHT_2HT_2H_2TH_4THTH_2THTHT_3H_2T_4H_3T_3HT_3H_8T_2HTHT_2H_2THT_2$	0.020	✓
$TH_2T_2HTHT_4HT_3HT_2H_2TH_2TH_2TH_2THT_5H_5THTH_2T_2H_3T_3HT_2H_4T_2H_5T_3HTH_2TH_3THTH_3THT_3H_2T_4HT_2$	0.024	✓
$H_3T_3HT_4H_5THTHTH_2TH_2T_2HTH_5T_2H_4TH_3T_4HT_7HT_2H_3THTHT_2HT_2HT_3H_2T_2HTHTH2THTHT_2H_5T_3$	0.030	✗
$HT_3HT_7HT_3H_2TH_5T_3H_3TH_2T_3H_4TH_4THTHT_2HTHT_3H_2T_2H_3T_2HTH_4TH_2TH_2TH_3T_2HT_3H_2THTHT_4HT$	0.034	✗
$TH_5TH_5T_3H_2T_3H_2THTHT_5HTHT_5H_2T_2H_2T_3HT_2HTH_2T_2HT_4HT_2H_4TH_2THTH_5T_3HT_2H_3T_2HTHTH_3TH$	0.037	✗
$T_2HTH_6THTH_2TH_3T_2H_2T_5H_3T_2HTH_2T_6HT_5HT_4H_3TH_2THT_3H_2THTHT_2H_3THT_2HT_2H_3THT_2H_3TH_4TH$	0.041	✗