Model for disordered proteins with strongly sequence-dependent liquid phase behavior

Antonia Statt [iD], Helena Casademunt [iD], Clifford P. Brangwynne [iD], and Athanassios Z. Panagiotopoulos [iD]

View Online   Export Citation   CrossMark

Model for disordered proteins with strongly sequence-dependent liquid phase behavior

View Online   Export Citation   CrossMark

Antonia Statt,[1,a] Helena Casademunt,[2,b] Clifford P. Brangwynne,[1,3]
and Athanassios Z. Panagiotopoulos[1,c]

AFFILIATIONS

[1] Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA

[2] Department of Physics, Princeton University, Princeton, New Jersey 08544, USA

[3] Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA

[a] **Present address:** Materials Science and Engineering, Grainger College of Engineering, University of Illinois, Urbana-Champaign, IL 61801, USA.
[b] **Present address:** Department of Physics, Harvard University, Cambridge, MA 02138, USA.
[c] **Author to whom correspondence should be addressed:** azp@princeton.edu

ABSTRACT

Phase separation of intrinsically disordered proteins is important for the formation of membraneless organelles or biomolecular condensates, which play key roles in the regulation of biochemical processes within cells. In this work, we investigated the phase separation of different sequences of a coarse-grained model for intrinsically disordered proteins and discovered a surprisingly rich phase behavior. We studied both the fraction of total hydrophobic parts and the distribution of hydrophobic parts. Not surprisingly, sequences with larger hydrophobic fractions showed conventional liquid–liquid phase separation. The location of the critical point was systematically influenced by the terminal beads of the sequence due to changes in interfacial composition and tension. For sequences with lower hydrophobicity, we observed not only conventional liquid–liquid phase separation but also re-entrant phase behavior in which the liquid phase density decreases at lower temperatures. For some sequences, we observed the formation of open phases consisting of aggregates, rather than a normal liquid. These aggregates had overall lower densities than the conventional liquid phases and exhibited complex geometries with large interconnected string-like or membrane-like clusters. Our findings suggest that minor alterations in the ordering of residues may lead to large changes in the phase behavior of the protein, a fact of significant potential relevance for biology.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5141095

I. INTRODUCTION

Liquid–liquid phase separation of intrinsically disordered proteins (IDPs) in cells is known to be crucial for a number of biological functions.[1–4] Membraneless organelles (such as the nucleolus, stress granules, and P-bodies[5]) partition cellular components into distinct regions, which play an important role in regulating biochemical processes.[6]

IDPs adopt many different chain conformations, much like synthetic polymers. This directly contributes to phase separation properties.[7] The unfolded conformational states are hypothesized to enable the formation of transient interaction networks,[8] which make phase separation possible at far lower concentrations than for folded proteins.[9]

The precise importance of protein disorder remains unclear. Here, we address the question of how much the sequence of the protein matters for phase separation. In order to better understand the biological relevance of IDPs and their phase separation, we attempt to make a connection between their sequence, conformational distributions, and resulting phase behavior.[10]

The substructure and properties of aggregates of IDPs are especially relevant to their biological function. Certain membrane-less organelles (like those formed by FUS, an RNA-binding protein[11]) are known to be liquid-like, while others (e.g., TDP-43, a

DNA-binding protein involved in splicing regulation[12]) have a more gel-like structure. For certain IDPs, changes in sequence-dependent phase behavior have been shown to drive pathological aggregation. This is the case for mutations in FUS and TDP-43 that are associated with amyotrophic lateral sclerosis (ALS).[13] Aggregate formation is commonly observed and well characterized in block copolymers,[14–17] where self-assembly is driven by the microphase separation of the different blocks, but it is less well understood for proteins.

Re-entrant phase behavior, where the concentration of the dense phase first increases, reaches a maximum, and then decreases again, is proposed to be an important regulatory mechanism for dissolving membraneless organelles,[18] as well as for the formation of dynamic droplet substructures or vacuoles.[19] This type of phase behavior is found in many different systems, such as patchy particles,[20] network forming systems,[21,22] and proteins.[18,19,23–27] In proteins, it can be driven by temperature,[23] RNA concentration,[18,19] or salt/ion concentration.[24,25,27]

To investigate the phase behavior of IDPs, both theoretical calculations[28–30] and molecular dynamics simulations on the atomistic[9,31–35] and the coarse-grained level[28,36–39] have recently been performed. Proteins can also be modeled with lattice models,[40] or as simple patchy particles or multi-component mixtures of patchy particles.[41–44] Recently, there has been significant progress in using the sticker and spacer model,[45,46] which preserves the polymeric nature of proteins. Field theory based methods[28,47,48] have been used to show the effects of charge patterning on the phase behavior of an IDP.

In this work, we focus on the sequence dependence of phase and aggregation behavior rather than the bulk self-assembly of chains, which has been studied extensively for various architectures of synthetic polymers,[49–52] including multiblock copolymers[53,54] and tapered blocks.[55] Aggregation behavior has also been extensively studied for dilute systems in solution, specifically with di- and tri-blocks[15,17] and multiblocks,[56] where the focus was mainly to describe finite-size aggregates such as micelles and vesicles, and gelation.[57]

In the present work, we use a simplified model of an IDP, where each section of the protein has either favorable or unfavorable interactions with the surrounding solvent and itself, which we call hydrophobic or hydrophilic hereafter. Because of the computational efficiency of the model, we are able to systematically study the influence of both the overall level of hydrophobicity and the distribution of hydrophobic/hydrophilic regions on liquid–liquid phase separation, as well as the characteristic of the aggregates that form. When the sequence has a substantial amount of both hydrophobic and hydrophilic beads, the large number of possible different sequences allows us to investigate the influence of the bead distribution on the phase behavior. We note that none of the sequences studied in this work corresponds to a specific protein. Instead, we aim at uncovering and understanding systematic trends in the phase behavior of these model disordered proteins.

In the following, we first describe the model and simulation details in Sec. II and then investigate the influence of hydrophobicity on the phase separation in Sec. III A, as well as the role of the end of the chain in Sec. III B. We study the effect of the distribution of hydrophobic parts in Sec. III C, where we observe re-entrant phase behavior and large-scale aggregation. We then investigate a number of previously proposed order parameters in Sec. III D. Finally, we conclude with discussion and outlook in Sec. IV.

II. MODEL AND METHODS

In this work, we study the phase behavior of a simplified model for IDPs using classical molecular dynamics (MD) simulations. We investigate the influence of the distribution of hydrophobic/hydrophilic regions and the degree of hydrophobicity on the resulting phase behavior. Thus, we only use two types of regions ("beads" in the model), namely, hydrophobic and hydrophilic, without resolving specific chemical interactions. Because of the coarse-grained nature of this model, each bead corresponds to multiple amino-acids in a hydrophobic or hydrophilic region along the chain in the protein.

For computational efficiency, we use an implicit-solvent model, so the vapor and liquid phases correspond to the dilute and condensed liquid phases of an IDP solution. Each chain consists of $M = 20$ bonded beads of mass m each. To ensure equilibration within computationally feasible time scales, we chose the length $M = 20$ ensuring that the chains are not entangled.

Inspired by surfactant models, we name the hydrophilic beads H and the hydrophobic beads T. The hydrophobic, attractive T beads interact through the Lennard-Jones (LJ) potential,

$$U_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right], \tag{1}$$

where r is the distance between two beads, ϵ is the energy well depth, and σ determines the interaction range. For computational efficiency, we applied a smoothing function to gradually decrease both the force and potential to zero at a cutoff of $r = 3\sigma$. The functional form is given in the supplementary material. The pair interaction of hydrophilic H beads was modeled with a purely repulsive Weeks–Chandler–Anderson (WCA) potential,[58]

$$U_{WCA}(r) = \begin{cases} U_{LJ}(r) + \epsilon, & r < 2^{1/6}\sigma \\ 0, & r \geq 2^{1/6}\sigma. \end{cases} \tag{2}$$

Cross-interactions between hydrophilic and hydrophobic beads were also described by the WCA potential. The total fraction of attractive, hydrophobic beads along the chain is denoted by f_T. Bonds between subsequent beads in the chain are described by the Finite Extensible Nonlinear Elastic (FENE) potential,

$$U_{b}(r) = -\frac{KR_0^2}{2} \ln\left[1 - \left(\frac{r}{R_0} \right)^2 \right] + U_{WCA}(r), \tag{3}$$

where $R_0 = 1.5\sigma$ is the maximum extension of a bond and $K = 30\epsilon/\sigma^2$ is the spring constant.

All simulations were performed using the HOOMD-blue (version 2.6.0) simulation package[59] on graphics processing units. The equations of motion were integrated using the velocity-Verlet algorithm with a time step of 0.005τ, where $\tau = \sqrt{m\sigma^2/\epsilon}$ is the unit of time. A weakly coupled Langevin thermostat with a friction constant of $0.1 m/\tau$ was employed in the NVT simulations to keep temperature constant. In the case of NpT simulations, a MTK barostat–thermostat[60] with coupling constants $\tau = 0.5$ and $\tau_P = 0.5$ was used.

In the following, ϵ is used as the energy unit, σ as the unit of length, and the mass m of a single bead as the unit of mass.

To obtain the coexistence properties, we used the established direct coexistence method.[61] Coexisting dense and dilute phases were simulated in an elongated box with dimensions $L_x \times L_y \times L_z$, where $L_z > L_x = L_y$. The two interfaces present in the simulation were oriented perpendicular to z to minimize surface energy. By recording density profiles along z, the coexistence densities were estimated from the density of the bulk regions sufficiently far away from the interfaces.

To check for finite size effects, we systematically varied both the cross-sectional area $A = L_x^2$ and the length of the simulation box ranging from $L_x = 10\sigma$ to 50σ and $L_z = 3L_x$ to $6L_x$ for selected sequences at different temperatures. We found that a box size of $L_x = L = 30\sigma$ and $L_z = 5L$ with $N = 1000$ chains of length $M = 20$ is sufficient to obtain reliable values for the coexistence densities. We excluded any simulations where either bulk phase occupied less than 10σ in the z direction and repeated them in a larger box of size $50\sigma \times 50\sigma \times 250\sigma$ containing $N = 4000$ chains, increasing the volume of both bulk phases.

All simulations were run on a single Graphics Processing Unit (GPU) (NVIDIA P100 or NVIDIA GeForce GTX 1080) for at least $25\,000\tau$ for equilibration and then another $50\,000\tau$ for measuring the density histograms from which the coexistence densities were obtained. The critical points (ρ_c, T_c) were estimated using the universal scaling of the coexistence densities near the critical point and the law of rectilinear diameters,

$$\rho_L - \rho_v = \Delta\rho_0\left(1 - T/T_c\right)^\beta, \qquad (4)$$

$$\frac{\rho_L + \rho_v}{2} = \rho_c + A(T_c - T), \qquad (5)$$

where $\beta \approx 0.325$ is the three-dimensional Ising model critical exponent.[61] A and $\Delta\rho_0$ are system specific fitting parameters. Because Eq. (4) is only valid close to the critical point, we only fitted coexistence densities up to approximately 30% below the critical point. Any simulations close to the critical temperature where the standard deviations of the coexistence densities were larger than the difference between the two densities were also excluded.

To estimate the statistical uncertainties in the critical points, we used the statistical error in the coexistence densities. Each run was divided into ten equal parts, and the coexistence densities were determined for each of the parts independently. Then, we fitted Eqs. (4) and (5) 300 times with a randomly selected coexistence density out of the ten parts for each measured point.[62]

III. RESULTS AND DISCUSSION

A. Influence of the fraction of hydrophobic beads on the phase behavior

We first determined the phase diagrams for a number of regular sequences with different degrees of hydrophobicity, starting from the fully hydrophobic chain. We systematically varied the fraction of hydrophobic beads from $f_T = 1$ to $f_T = 0.6$ by adding repulsive beads evenly distributed along the chain. The measured phase diagrams of selected sequences are shown in Fig. 1. In this and subsequent figures, sequences are depicted with hydrophobic, attractive T beads in red circles and hydrophilic, repulsive H beads in blue circles.
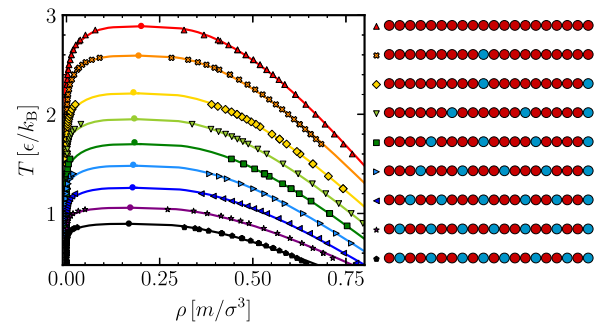

FIG. 1. Coexistence densities of the dilute and dense phases for regular sequences with varying f_T, from $f_T = 1.0$ (red) to $f_T = 0.6$ (black), as indicated. The locations of the critical points are shown by circles, and the lines show the fit of Eqs. (4) and (5) to the upper part of the phase envelope. The legend shows the corresponding sequences (sorted from high to low T_c) with red filled circles for T beads and blue filled circles for H beads. Statistical uncertainties are smaller than symbol size.

All sequences studied and their respective critical points are given in Table S1 of the supplementary material.

As shown in Fig. 1, the critical temperature decreases with decreasing fraction f_T and the phase envelopes become flatter and narrower, as expected from the phase behavior of long chains.[62] Note that all simulated chains in this work are of the same length $M = 20$. The shifts in phase envelope and critical points are purely due to the fraction of attractive beads and their distribution along the chain.

As expected from Flory–Huggins scaling,[63] the critical temperature T_c shows a quadratic dependency on the fraction of hydrophobic, attractive beads f_T, as shown in Fig. 2(a). A systematic decrease with decreasing f_T can be observed in the critical densities ρ_c as well [Fig. 2(b)].

Increasing the fraction of hydrophobic, attractive beads in our model is similar to increasing a protein's number of stickers, representing folded binding domains.[64] Studies have shown that performing mutations at sticker sites, effectively rendering them nonfunctioning, reduces the propensity of a protein to phase separate,[65] which is consistent with our observations.
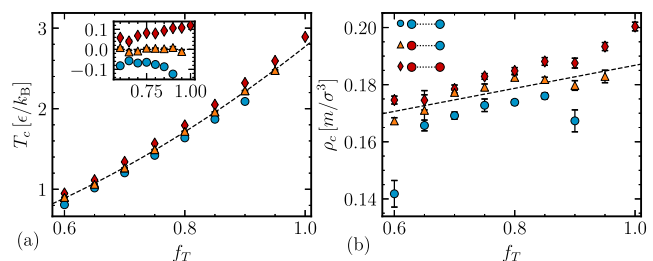

FIG. 2. Scaling of the critical temperature T_c (a) and critical density ρ_c (b) with f_T. As shown in the legend, red diamonds indicate sequences where both terminal beads are T, blue circles indicate sequences with two H ends, and orange triangles indicate the sequences with one H and one T end. The dashed line in (a) indicates a quadratic fit and a linear fit to the data in (b). The inset in (a) shows the deviation of T_c from the quadratic fit.

B. Influence of the terminal bead type on the location of the critical point

The findings above show that the fraction of hydrophobic, attractive beads is important for defining the phase boundary, but it is unclear if the precise sequence of beads matters. For each $f_T \leq 0.9$, we computed the phase boundaries and critical point for (1) a sequence with two attractive, hydrophobic terminal beads, (2) a sequence with two repulsive, hydrophilic terminal beads, and (3) a sequence with mixed terminal beads. All sequences are given in Table S1 of the supplementary material.

Attractive ends shifted both T_c and ρ_c up, whereas repulsive ends resulted in a lowered T_c and ρ_c, with the mixed end sequences in between. This effect appeared to be general for all investigated sequences up to $f_T = 0.6$, as shown in Fig. 2.

As previously suggested, the radius of gyration R_g of a single chain measured at a fixed temperature could be used as a predictor of the critical point.[37,48,66] It is especially useful because the radius of gyration is an experimentally accessible quantity.[67,68] For the model investigated here, we indeed found that both critical temperature and density scale roughly linearly with R_g, as reported in Fig. S1 of the supplementary material. However, R_g did not capture the influence of the terminal bead type, suggesting that this effect results from some other mechanisms.

To systematically study the effect of terminal bead type, we simulated a series of chains with only one repulsive bead ($f_T = 0.95$) and varied its position along the chain from the middle to the end. The phase diagrams and scaling of the critical point are shown in Fig. S2 of the supplementary material. For sequences where the repulsive bead was about six positions away from the end or more, almost no difference in critical point was measured. For sequences where the repulsive bead was near the end of the chain, we observed that the critical point decreased as the bead moved closer to the end. As displayed in Fig. 3, the decrease in the critical point was ~5% when moving the repulsive, hydrophilic H bead from the middle to the end to the chain, and ~3% when moving the H bead to the second outermost position of the chain.

The type of terminal bead had a systematic influence on the interfacial composition, as shown in Fig. S3 of the supplementary material. A slight excess of chain ends at the dilute-dense interface

is expected because of entropic effects, even for the homopolymer case.[69] We observed an enhancement in the concentration of both end beads and repulsive beads in the interfacial region for sequences where the repulsive H bead was near the end of the chain. The interface composition influences the value of the interfacial tension, which scales according to

$$\gamma = \gamma_0 (1 - T/T_c)^\mu, \tag{6}$$

with the distance to the critical point, where $\mu \approx 1.26$, the relevant exponent for the three-dimensional Ising universality class.[70] Because the critical point and the interfacial tension are connected, we speculate that any changes in interfacial composition also lead to a change in the critical point.

Due to computational limitations, we have not measured the interfacial tension directly in this work but instead report an estimated interfacial tension $\hat{\gamma} = k_B T/(2\pi w^2) \ln L$ from the interface width w, as shown in the inset of Fig. 3. As expected from Eq. (6), we observed a collapse of $\hat{\gamma}$ onto the same curve for the different sequences, when plotted against $1 - T/T_c$. The investigation of the precise relationship between sequence, critical point, and interfacial tension is left for future work.

Thus, these findings show that the exact sequence can have a significant effect on the phase boundary and signal the importance of the terminal bead, a result which to our knowledge has not been reported before. It would be interesting to study experimentally if mutations at the ends of an IDP have a more pronounced effect on the phase behavior than mutations in the middle of the chain. To our knowledge, the effect of the relative position of a sticker site on the phase behavior of a protein has not been experimentally probed.

C. Influence of the distribution of attractive beads on the phase behavior for $f_T = 0.6$

In contrast to the sequences with higher fraction of attractive beads, the chains with $f_T = 0.6$, the lowest value studied, displayed not only conventional phase separation between a dilute and a dense phase but also re-entrant phase behavior. This behavior is characterized by a density of the condensed phase that first increases, reaches a maximum, and then decreases as temperature decreases (see Fig. 4). We also found sequences that do not seem to form conventional liquid phases at all but form large-scale aggregates instead. In the following, we describe results for each of these three different cases.

1. Phase separation and re-entrant phase behavior

We have found some sequences with $f_T = 0.6$ [shown in Fig. S6(a) of the supplementary material] that exhibit conventional phase separation into a dilute and a dense phase, following Eqs. (4) and (5). By conventional phase separation, we mean that it involves a first-order phase transition with a discontinuous change in density. Because we did not exhaustively investigate the space of possible sequences, we expect that there are more sequences with conventional phase separation. The glass transition for the purely attractive homopolymer is $T_g \approx 0.4\epsilon/k_B$,[71] and we therefore did not attempt to simulate temperatures below $0.45\epsilon/k_B$.

Some sequences with $f_T = 0.6$ showed re-entrant phase behavior, where the density of the condensed phase first increased,
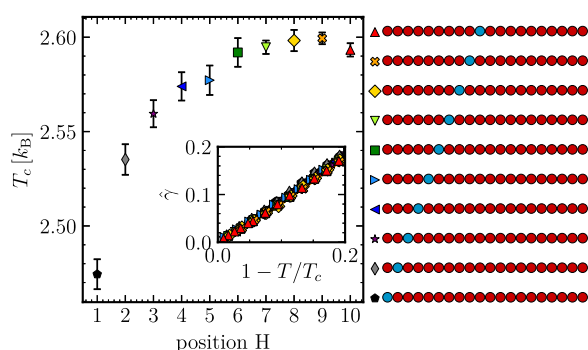

FIG. 3. Scaling of the critical temperature as a function of the position of the repulsive H (blue circles) bead from the end of the chain to the middle. The inset shows the scaling of $\hat{\gamma}$ with the distance $1 - T/T_c$ to the critical point.
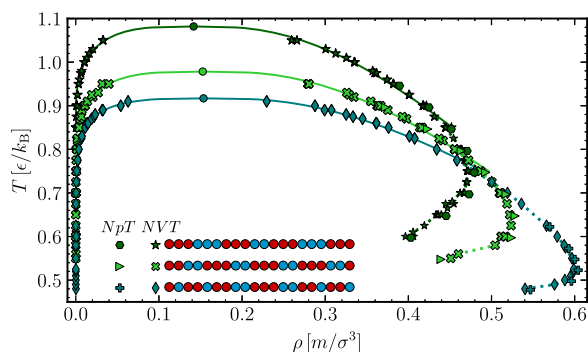
FIG. 4. Coexistence densities for selected sequences with $f_T = 0.6$, which show re-entrant behavior. The coexistence densities were measured in *NVT* and *NpT* ensembles as indicated. The locations of the critical points are marked by circles, and the statistical uncertainties are smaller than the symbol size. The solid lines show the fit of Eqs. (4) and (5) to the upper part of the phase envelope determined in *NVT* conditions. The dashed lines are guides to the eye only.

reached a maximum, and then decreased with decreasing temperature. Examples of the phase envelopes are displayed in Fig. 4; the rest of the sequences are given in Fig. S3(b) of the supplementary material. First, a conventional phase separation into a dense and a dilute phase occurred as the system was cooled down. This part if the phase envelope can be fitted by Eqs. (4) and (5). Upon further cooling, the dense phase developed sub-structures. We observed

small H and T rich regions, with large voids in between them. This microphase separation of the condensed liquid resulted in a lower overall density of the dense phase.

For even colder temperatures, we found the formation of large-scale aggregates. A typical example for the sequence $T_3H_3T_3H_2T_3H_3T_3$ is shown in Fig. 5, and more examples at different temperatures are shown in Fig S7 of the supplementary material. These large-scale aggregates typically had fairly low densities compared to the liquid densities observed for conventional phase separation into a disordered condensed phase. The structure and properties of the aggregates are discussed in Sec. III C 2.

We confirmed the results of the *NVT* direct coexistence measurements with *NpT* simulations of the dense phase (shown in Fig. 4). For the *NpT* simulations, we estimated $p \approx 0$ from the ideal gas law $p \approx (\rho_v/Mm)k_BT$. Accordingly, we only performed *NpT* simulations for low temperatures, where $\rho_v \approx 0$.

For the reported temperature range, we did not observe a difference in the structure or density of the phases between *NVT* and *NpT* or between different independent simulation runs. For sufficiently cold temperatures, we were unable to equilibrate the large-scale aggregates reliably and excluded the data.

The critical temperatures and densities of the re-entrant sequences in Fig. 4 were comparable to those of the sequences in Fig. 1. Even though the re-entrant onset in the liquid branch occurred at different temperatures for each sequence, the relative distance T/T_c to the critical temperature is similar for all of them, roughly 60%–70% below T_c, suggesting a common underlying mechanism. However, the highest density value for the liquid branch varies greatly between 0.45 and $0.6 m/\sigma^3$, as shown in Fig. 4.
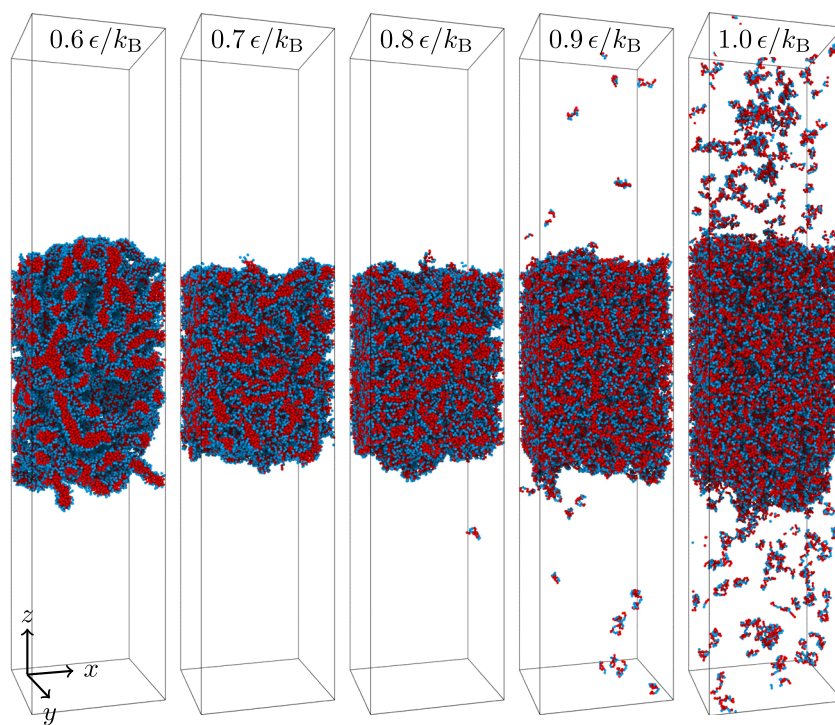

FIG. 5. Typical configurations of the sequence $T_3H_3T_3H_2T_3H_3T_3$ (d1) at different temperatures. The snapshots were taken from the direct coexistence *NVT* simulations in a box of size $50\sigma \times 50\sigma \times 250\sigma$ with $N = 4000$ chains. Hydrophobic, attractive T beads are shown in red circles and hydrophilic H beads are shown in blue circles. The critical point is $T_c = (1.109 \pm 0.006)\varepsilon/k_B$, $\rho_c = (0.146 \pm 0.002)m/\sigma^3$ and the phase envelope (green stars) of this sequence is shown in Fig. 4. Snapshots were generated using OVITO.[72]

We observed limitations with respect to our ability to equilibrate the systems at low temperatures: (1) near the glass transition, the dynamics slowed down drastically, (2) the results depended on the initial configuration and ensuring that large-scale structures were properly equilibrated became increasingly more difficult, and (3) for some large-scale aggregates, it was not clear how to unambiguously define a bulk density. Therefore, we limited each reported phase diagram to the temperature range where the mentioned limitations were not severe.

The sequences that exhibited re-entrant phase behavior have an overall more "blocky" distribution of hydrophilic and hydrophobic sections, with the longest section being three beads long. This suggests that "blockiness" plays an important role in the ability to form structured liquids at low temperatures, in agreement with the work of Nott et al.,[73] who showed the relevance of blocky patterned electrostatic interactions to phase separation. We found all three possible combinations of terminal end beads among the sequences as well as different terminal block lengths. Because we investigated a limited subset of possible sequences, future work will be needed to solidify the systematic connection between sequences and phase behavior.

The driving mechanism for re-entrant phase behavior is the competition between self-assembly of large-scale ordered structures and condensation into a conventional dense liquid, similar to what has been observed in patchy particles[20] and network-forming fluids.[21] We observed the formation of large-scale microphase separated structures at low temperatures, which led to a lower density of the condensed phase. It is therefore reasonable to assume that the temperature at which we observe re-entrant phase behavior will be connected to the order-to-disorder transition temperature[55,56] of the sequence, which determines the temperature dependence of the microphase separation.

2. Formation of large-scale aggregates and their structure

In addition to conventional dense-dilute phase separation and re-entrant phase behavior, we also found some sequences which only formed large-scale aggregates. These sequences are listed in Table S2 of the supplementary material. In contrast to the previously discussed sequences in Sec. III C 1, the sequences discussed here did not show dense-dilute phase coexistence at any given temperature. Instead, the large-scale aggregates fell apart into smaller isolated aggregates when the temperature was increased sufficiently.

The structure of the large-scale aggregates varied strongly with sequence. We observed fibril-like or string-like clusters that were interconnected, as displayed in Fig. 6. This type of aggregate is characterized by many large voids. We also found membrane-like structures for some sequences (also in Fig. 6), including empty vesicles, flat membranes, and layered, onion-like structures, often observed in the same simulation. Similar aggregate structures are reported in the literature for multi-block polymers.[15,17,56,57,74,75]

In addition to the large-scale fibril-like aggregates and membrane-like aggregates, we also found the expected finite sized aggregates from the multi-block polymer literature: worm-like interconnected micelles for the tri-block chain $T_6H_8T_6$, and spherical micelles for $H_4T_{12}H_4$ and the block–copolymer $T_{12}H_8$. Some examples for those morphologies are shown in Fig. 6. None of the measured short range structural properties of the chains in those aggregates and the liquid configurations of the previous sections were different enough to distinguish between the different behaviors, as shown in Sec. G of the supplementary material.

In Fig. 7, we report the apparent coexistence densities of the fibril-like aggregates that did not exhibit a conventional dilute-dense phase coexistence at any temperature. Each density was obtained
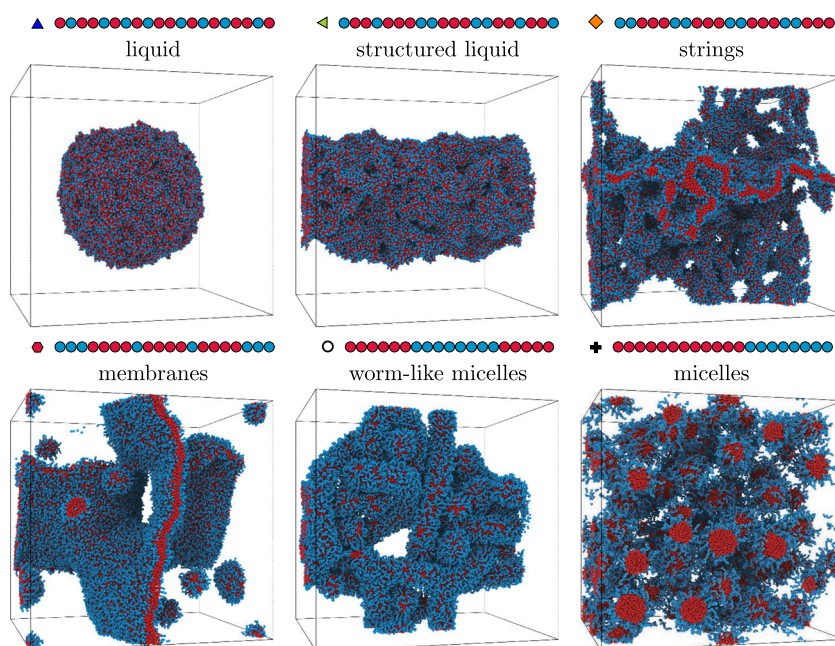

FIG. 6. Typical configurations at $\rho_{tot} = 0.1 m/\sigma^3$ and $T = 0.65\epsilon/k_B$ of selected sequences with $f_T = 0.6$ (see main text for identification criteria). The volume of each system was $100\sigma^3$. Snapshots were generated using OVITO.[72]
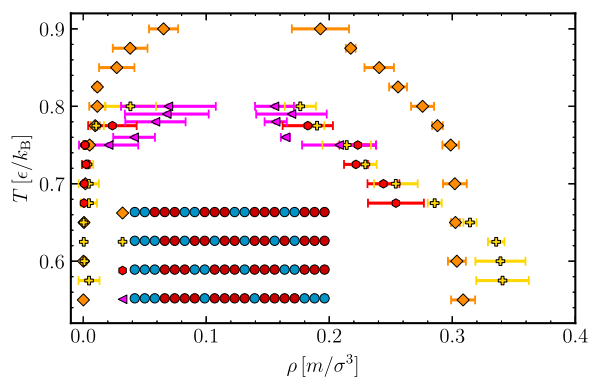
FIG. 7. Apparent densities of the dilute and dense phase of sequences with $f_T = 0.6$, which do not show conventional phase separation. In the legend, hydrophobic T beads are shown in red circles and hydrophilic H beads are shown in blue circles.

FIG. 8. Genus G and average area-to-volume aspect ratio $\langle A/V \rangle$ of the surfaces determined from the configurations with $\rho_{tot} = 0.1 m/\sigma^3$ and $T = 0.55\epsilon/k_B$. The labeled light gray circles are drawn for a better visualization of the different morphologies as identified by eye from snapshots of configurations, so boundaries are not rigorous. The gray crosses indicate measured G and $\langle A/V \rangle$ values for sequences with $f_T > 0.6$. Corresponding typical snapshots are shown in Fig. 6.

by averaging over five independent direct coexistence runs with $N = 4000$ each, started from different initial configurations. Because of the presence of large voids, the observed density of the dense phase was much lower than for the liquid phases studied in Secs. III A and III B. Due to sampling limitations at low temperatures, we only report fibril-like large-scale aggregate densities in Fig. 7, where we were able to obtain a consistent density value from independent simulation runs.

In comparison to conventional liquids, we also observed a much higher heterogeneity within the aggregates, including the formation of large voids and holes as well as microphase separation into H and T rich regions. Consequently, the variance of the local density within the dense phase was much higher for a fibril-like aggregate than for a conventional dense liquid. For a normal liquid, the variance decreases with decreasing temperature, but this trend is inverted for the fibril-like aggregates. For re-entrant phase behavior, the variance decreases and then increases with decreasing temperature, as shown in Fig. S12 of the supplementary material.

We determined the surface to volume ratio of the clusters $\langle A/V \rangle$ with a surface mesh method,[76] using a probe sphere radius of $R = 2\sigma$. The resulting surface is a triangulated mesh enclosing the dense aggregates or liquid phase. As expected, the liquids have the lowest average surface to area ratio, as shown in Fig. 8. The large-scale aggregates are less dense and have holes or voids, leading to a higher surface area. We note that the results were shifted for different temperatures, but the relative ordering of the sequences remained unchanged.

From the triangulated mesh, we were able to calculate the genus of the surface, $G = 1 - \chi/2$, where $\chi = N_T - N_E + N_V$ is the Euler characteristic.[77] N_T is the number of faces, N_E is the number of edges, and N_V is the number of vertices defining the surface. By determining the genus of the surfaces, we effectively counted the holes (positive G) or internal voids (negative G). This property, in conjunction with $\langle A/V \rangle$, was able to distinguish between the different large-scale aggregates, as shown in Fig. 8.

There are no sharp boundaries between the different large-scale aggregate types and some sequences exhibit different types of behavior depending on temperature (e.g., re-entrant phase behavior or a
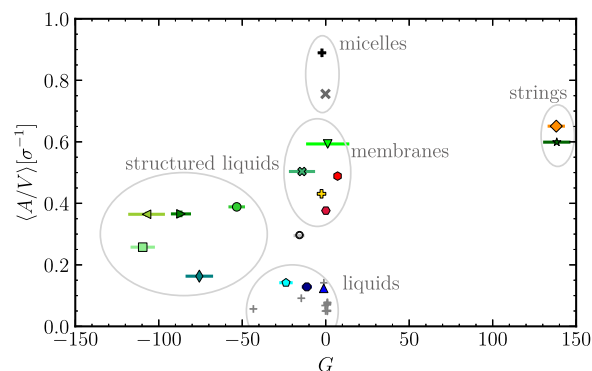
string-like to membrane-like transition in $H_3[T_4H]_3H_2$). Therefore, the boundaries drawn in Fig. 8 serve as guidelines only. We observe that, in contrast to the conventional polymer systems, the sequences investigated here are less regular, so relatively small changes in the sequence have a significant effect on the aggregation behavior.

Analysis of biological condensates using fluorescence microscopy can provide information regarding the localization of IDPs, but microstructure and conformational states of the aggregates are often difficult to access in experimental studies. However, these might be of biological relevance. The microstructure in intracellular condensates has been reported in both P granules,[78] where gel and liquid phases are co-assembled, and stress granules,[79] where a network of protein–protein interactions is formed. It has been suggested that microstructures are essential elements for biomolecular condensates,[78] pointing toward potential biological relevance of the large-scale aggregates that form in our simulation.

Pathological protein aggregation also plays an important role in diseases such as ALS and Alzheimer's.[3,80] However, the precise mechanism explaining how point mutations in disordered proteins can give rise to pathological aggregates is not fully understood.[3,80] We speculate that a systematic classification of the different connections between sequence and aggregate types may lead to advances in both understanding the diseases and, potentially, new drug development.

D. Order parameter

As shown in Secs. III A–III C, this simple model for IDPs shows diverse phase behavior. We observed conventional dense-dilute phase separation, re-entrant phase behavior, and large-scale aggregate formation. To distinguish between all of them and to estimate the critical point location, we would like to have a predictive order parameter.

We have tested commonly used order parameters: (1) the average length L_T of the hydrophobic segments in the sequence,[81] (2) the normalized mean-square fluctuation of the block hydrophobicity $\Psi^{(s)}$ of a sub-section of length s,[82,83] (3) the sequence charge

decoration SCD,[32] (4) the order parameter κ,[31] and (5) the corrected probability of finding a T segment after a T segment P_{TT} $- f_T$.[84,85] The SCD and κ parameters are used frequently for proteins,[32] whereas the other order parameters are usually applied to co-polymer systems. Details are given in Sec. H of the supplementary material.

All of the tested order parameters are based only on single chain sequence, which is desirable for predictive capability. Unfortunately, most of them perform poorly (Fig. S13 of the supplementary material), with SCD showing the best correlation with T_c, as shown in Fig. 9(a). None of the order parameters in the literature take the distance of each bead to the end of the chain into account and thus cannot predict the strong effect of changing the terminal bead type. For SCD, changing the end bead seems to cause a constant offset, which might allow us to incorporate the terminal bead type effect to this order parameter in the future.

To capture the variation in the effect size of changing the bead type based on the position of the bead in the chain, we defined a new order parameter that acts as an effective reweighted f_T^*. The weights of T beads are determined based on the critical temperature of sequences with $f_T = 0.95$ [as reported in Fig. S2(b) of the supplementary material], while H beads are given zero weight. The full definition is given in the supplementary material. We have not considered the impact of H beads on entropic effects related to chain configurations, and we neglect chain length effects that may impact phase separation.

With this definition, we can achieve a fairly linear correlation between f_T^* and T_c for all investigated sequences which had a critical point, as shown in Fig. 9(b). In fact, it seems to follow the quadratic trend expected for f_T (as in Fig. 2). This definition is specific to the model investigated here and it is not purely based on the sequence alone. Regardless of its limitations, this order parameter illustrates the significance of the beads near the end of the chain for the location of the critical point.

None of the order parameters or single-chain properties tested here, including f_T^*, R_g, and T_Θ, were able to predict the different observed types of phase behaviors (e.g., re-entrant) and aggregations (e.g., membrane-like structures), as discussed in Secs. III C 1 and III C 2.

From the previous results, it is clear that the overall fraction of hydrophobic beads f_T, the types of terminal beads, and the

"blockiness" of the sequence are the most important factors to consider for an order parameter. Low blockiness sequences with sufficient fraction of T beads appeared to undergo conventional phase separation. Increasing blockiness seemed to lead to a higher propensity to show re-entrant phase behavior and eventually to the formation of large-scale aggregates. Even higher blockiness resulted in finite sized aggregates such as micelles.

IV. CONCLUSIONS

In this work, we determined the phase behavior of 37 different coarse-grained model IDP sequences. We investigated the influence of both the fraction of hydrophobic beads f_T and the distribution of hydrophobic beads along the chain for $f_T = 0.6$. We have found phase behavior ranging from conventional dense-dilute phase separation to re-entrant phase behavior, as well as large-scale aggregate formation.

Our results show that the types of beads located at the end of the chain (hydrophobic or hydrophilic) have a systematic effect on the location of the critical point, where hydrophobic terminal beads increase both T_c and ρ_c and hydrophilic terminal beads decrease T_c and ρ_c. We then showed that the systematic shift is related to the composition of the interfacial region and speculated that it is intimately connected to the interfacial tension. To our knowledge, this systematic trend has not been reported before.

For many sequences with $f_T = 0.6$, we observed re-entrant phase behavior, where the density of the liquid phase first increased and then decreased with decreasing temperature. This intriguing complex phase boundary behavior was due to the emerging order in the dense phase upon cooling. For protein systems, this is proposed to be an important mechanism for dissolving membrane-less organelles[18] and for forming vacuoles.[19] While the simple one-component model in this work showed re-entrant phase behavior, it cannot be easily mapped onto the more complex experimental systems that are usually driven by salt[24,25,27] or RNA concentration.[18,19] Establishing a qualitative link to biologically relevant systems by introducing some of these effects is outside the scope of this work.

We also found some sequences that form large-scale aggregates instead of conventional condensed phases. These aggregates differ in their morphology and can be distinguished by their area-to-volume ratio and their number of holes and voids. We observed membrane-like configurations, interconnected fibril-like networks, as well as traditional spherical and worm-like micelles. Seemingly minor changes in the sequence of the model protein led to large changes in the phase behavior. Many intracellular condensates may potentially exhibit rich substructures,[78,79] which likely reflects one aspect of the complex phase behavior captured in the present simulations.

To our knowledge, there is no predictive order parameter, which allows us to distinguish dense-dilute phase separation, re-entrant phase behavior, and the different large-scale aggregation types observed. Our order parameter f_T^*, single chain properties R_{ij} (or T_θ) and R_g, and SCD could predict the critical point location to an extent. However, they were all unable to differentiate between conventional liquid–liquid phase separation and other aggregation behavior. Future efforts will be directed toward understanding which sequence features determine phase behavior and toward developing an order parameter.
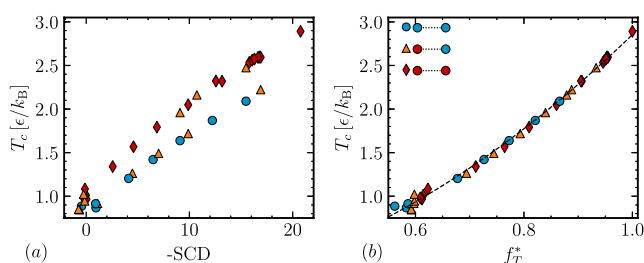

FIG. 9. Critical temperature T_c as a function of (a) sequence charge decoration SCD and (b) reweighted f_T^* for all sequences with a critical point. Red diamonds indicate sequences where both terminal beads are T, blue circles indicate sequences with two H ends, and orange triangles indicate the sequences with one H and one T end. The dashed line in (b) is a quadratic fit.

Characterizing the influence of sequence on phase behavior is key to understanding the biological function of phase separation of proteins. The rich structural behavior we observed in this work may be linked to pathological protein aggregation found in diseases such as ALS and Alzheimer's.[3,80] Thus, establishing a link between sequence and aggregation behavior, as done in this study, might lead to insights into the connection between point mutations and pathological aggregation. Further research efforts are needed to better characterize the relation between sequence and phase behavior for more realistic models, which could help understand neurodegenerative diseases and offer insights into drug development.

It would be especially useful to develop a predictive order parameter for biologically relevant protein sequences. With a solid understanding of which sequence features play an important role in phase separation and aggregate formation, engineered condensates could potentially be used for medical applications, as we may be able to predict how specific mutations might change the phase behavior of a given protein.

SUPPLEMENTARY MATERIAL

See the supplementary material for additional information on simulation details, the scaling of the radius of gyration and the Θ-temperature with the critical point, phase diagrams and interface compositions of the sequences with $f_T = 0.95$, and additional phase diagrams for sequences with $f_T = 0.6$. The structure of the liquid and large-scale aggregates, and the order parameters used in this work are also discussed. Two tables with all 37 sequences investigated and their critical points are provided.

AUTHOR'S CONTRIBUTIONS

A.S. and H.C. contributed equally to this work.

ACKNOWLEDGMENTS

REFERENCES

[1] J. Berry, C. P. Brangwynne, and M. Haataja, Rep. Prog. Phys. **81**, 046601 (2018).
[2] Y. Shin, Y.-C. Chang, D. S. Lee, J. Berry, D. W. Sanders, P. Ronceray, N. S. Wingreen, M. Haataja, and C. P. Brangwynne, Cell **175**, 1481 (2018).
[3] Y. Shin and C. P. Brangwynne, Science **357**, eaaf4382 (2017).
[4] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, Nat. Rev. Mol. Cell Biol. **18**, 285 (2017).
[5] L. Zhu and C. P. Brangwynne, Curr. Opin. Cell Biol. **34**, 23 (2015).
[6] J. Berry, S. C. Weber, N. Vaidya, M. Haataja, and C. P. Brangwynne, Proc. Natl. Acad. Sci. U. S. A. **112**, E5237 (2015).
[7] V. N. Uversky, Curr. Opin. Struct. Biol. **44**, 18 (2017).
[8] H.-X. Zhou, V. Nguemaha, K. Mazarakos, and S. Qin, Trends Biochem. Sci. **43**, 499 (2018).
[9] 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, Nat. Chem. **9**, 1118 (2017).
[10] P. E. Wright and H. J. Dyson, Nat. Rev. Mol. Cell Biol. **16**, 18 (2015).
[11] K. Burke, A. Janke, C. Rhine, and N. Fawzi, Mol. Cell **60**, 231 (2015).
[12] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. Kanagaraj, H. Kim, T. Mittag, and J. Taylor, Cell **163**, 123 (2015).
[13] W. M. Babinchak, R. Haider, B. K. Dumm, P. Sarkar, K. Surewicz, J.-K. Choi, and W. K. Surewicz, J. Biol. Chem. **294**, 6306 (2019).
[14] C. Koch, A. Z. Panagiotopoulos, F. L. Verso, and C. N. Likos, Soft Matter **11**, 3530 (2015); M. A. Floriano, E. Caponetti, and A. Z. Panagiotopoulos, Langmuir **15**, 3143 (1999).
[15] S. Li, C. Yu, and Y. Zhou, Sci. Chin. Chem. **62**, 226 (2019).
[16] P. Posocco, M. Fermeglia, and S. Pricl, J. Mater. Chem. **20**, 7742 (2010).
[17] D. Dolgov, T. Grigor'ev, A. Kulebyakina, E. Razuvaeva, R. Gumerov, S. Chvalun, and I. Potemkin, Polym. Sci., Ser. A **60**, 902 (2018).
[18] A. N. Milin and A. A. Deniz, Biochemistry **57**, 2470 (2018).
[19] P. R. Banerjee, A. N. Milin, M. M. Moosa, P. L. Onuchic, and A. A. Deniz, Angew. Chem., Int. Ed. **56**, 11354 (2017).
[20] J. R. Espinosa, A. Garaizar, C. Vega, D. Frenkel, and R. Collepardo-Guevara, J. Chem. Phys. **150**, 224510 (2019).
[21] J. Russo, J. M. Tavares, P. I. C. Teixeira, M. M. Telo da Gama, and F. Sciortino, Phys. Rev. Lett. **106**, 085703 (2011).
[22] A. Zilman and S. Safran, Phys. Rev. E **66**, 051107 (2002).
[23] M. Tempel, G. Isenberg, and E. Sackmann, Phys. Rev. E **54**, 1802 (1996).
[24] F. Zhang, M. Skoda, R. Jacobs, S. Zorn, R. A. Martin, C. Martin, G. Clark, S. Weggler, A. Hildebrandt, O. Kohlbacher et al., Phys. Rev. Lett. **101**, 148101 (2008).
[25] F. Zhang, S. Weggler, M. J. Ziller, L. Ianeselli, B. S. Heck, A. Hildebrandt, O. Kohlbacher, M. W. A. Skoda, R. M. J. Jacobs, and F. Schreiber, Proteins: Struct., Funct., Bioinf. **78**, 3450 (2010).
[26] J. Möller, S. Grobelny, J. Schulze, S. Bieder, A. Steffen, M. Erlkamp, M. Paulus, M. Tolan, and R. Winter, Phys. Rev. Lett. **112**, 028101 (2014).
[27] E. Jordan, F. Roosen-Runge, S. Leibfarth, F. Zhang, M. Sztucki, A. Hildebrandt, O. Kohlbacher, and F. Schreiber, J. Phys. Chem. B **118**, 11365 (2014).
[28] J. McCarty, K. T. Delaney, S. P. Danielsen, G. H. Fredrickson, and J.-E. Shea, J. Phys. Chem. Lett. **10**, 1644 (2019).
[29] J. Lee, Y. O. Popov, and G. H. Fredrickson, J. Chem. Phys. **128**, 224908 (2008).
[30] L. Sawle and K. Ghosh, J. Chem. Phys. **143**, 085101 (2015).
[31] R. K. Das and R. V. Pappu, Proc. Natl. Acad. Sci. U. S. A. **110**, 13392 (2013).
[32] S. Das, A. N. Amin, Y.-H. Lin, and H. S. Chan, Phys. Chem. Chem. Phys. **20**, 28558 (2018).
[33] A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, and R. V. Pappu, Proc. Natl. Acad. Sci. U. S. A. **107**, 8183 (2010).
[34] A. Vitalis and R. V. Pappu, J. Comput. Chem. **30**, 673 (2009).
[35] G. H. Zerze, R. B. Best, and J. Mittal, J. Phys. Chem. B **119**, 14622 (2015).
[36] G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, ACS Cent. Sci. **5**, 821 (2019).
[37] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, Proc. Natl. Acad. Sci. U. S. A. **115**, 9929 (2018).
[38] S. Qin and H.-X. Zhou, J. Phys. Chem. B **120**, 8164 (2016).
[39] C. Miller, Y. Kim, and J. Mittal, Biophys. J. **111**, 28 (2016).
[40] E. M. O'Toole and A. Z. Panagiotopoulos, J. Chem. Phys. **97**, 8644 (1992).
[41] V. Nguemaha and H.-X. Zhou, Sci. Rep. **8**, 6728 (2018).
[42] P. Sarangapani, S. Hudson, R. Jones, J. Douglas, and J. Pathak, Biophys. J. **108**, 724 (2015).
[43] H. Liu, S. K. Kumar, and F. Sciortino, J. Chem. Phys. **127**, 084902 (2007).
[44] A. Ghosh, K. Mazarakos, and H.-X. Zhou, Proc. Natl. Acad. Sci. U. S. A. **116**, 19474 (2019).
[45] T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, Elife **6**, e30294 (2017).
[46] J. Wang, J.-M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, and A. A. Hyman, Cell **174**, 688 (2018).

[47]Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan, Phys. Rev. Lett. **117**, 178101 (2016).

[48]Y.-H. Lin and H. S. Chan, Biophys. J. **112**, 2043 (2017).

[49]M. W. Matsen, Macromolecules **45**, 2161 (2012).

[50]J. Zhang, R. Deubler, M. Hartlieb, L. Martin, J. Tanaka, E. Patyukova, P. D. Topham, F. H. Schacher, and S. Perrier, Macromolecules **50**, 7380 (2017).

[51]C. M. Bates and F. S. Bates, Macromolecules **50**, 3 (2017).

[52]W. G. Levine, Y. Seo, J. R. Brown, and L. M. Hall, J. Chem. Phys. **145**, 234907 (2016).

[53]F. S. Bates, M. A. Hillmyer, T. P. Lodge, C. M. Bates, K. T. Delaney, and G. H. Fredrickson, Science **336**, 434 (2012).

[54]L. Wu, E. W. Cochran, T. P. Lodge, and F. S. Bates, Macromolecules **37**, 3360 (2004).

[55]T. Pakula and K. Matyjaszewski, Macromol. Theory Simul. **5**, 987 (1996).

[56]M. E. Gindy, R. K. Prud'homme, and A. Z. Panagiotopoulos, J. Chem. Phys. **128**, 164906 (2008).

[57]V. Hugouvieux, M. A. V. Axelos, and M. Kolb, Macromolecules **42**, 392 (2009); Soft Matter **7**, 2580 (2011).

[58]J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).

[59]J. Glaser, T. D. Nguyen, J. A. Anderson, P. Lui, F. Spiga, J. A. Millan, D. C. Morse, and S. C. Glotzer, Comput. Phys. Commun. **192**, 97 (2015); J. A. Anderson, C. D. Lorenz, and A. Travesset, J. Comput. Phys. **227**, 5342 (2008).

[60]G. J. Martyna, D. J. Tobias, and M. L. Klein, J. Chem. Phys. **101**, 4177 (1994).

[61]J. S. Rowlinson and B. Widom, *Molecular Theory of Capillarity* (Clarendon Press, Oxford, 1982).

[62]K. S. Silmore, M. P. Howard, and A. Z. Panagiotopoulos, Mol. Phys. **115**, 320 (2017).

[63]P. J. Flory, J. Chem. Phys. **10**, 51 (1942); M. L. Huggins, J. Phys. Chem. **46** 151 (1942).

[64]A. S. Holehouse and R. V. Pappu, Biochemistry **57**, 2415 (2018).

[65]D. Bracha, M. T. Walls, M.-T. Wei, L. Zhu, M. Kurian, J. L. Avalos, J. E. Toettcher, and C. P. Brangwynne, Cell **175**, 1467 (2018).

[66]Y.-H. Lin, J. P. Brady, J. D. Forman-Kay, and H. S. Chan, New J. Phys. **19**, 115003 (2017).

[67]H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **109**, 16155 (2012).

[68]J. A. Riback, C. D. Katanski, J. L. Kear-Scott, E. V. Pilipenko, A. E. Rojek, T. R. Sosnick, and D. A. Drummond, Cell **168**, 1028 (2017).

[69]E. Helfand, S. M. Bhattacharjee, and G. H. Fredrickson, J. Chem. Phys. **91**, 7200 (1989).

[70]J. S. Rowlinson and B. Widom, *Molecular Theory of Capillarity* (Courier Corporation, 2013).

[71]T. S. Jain and J. J. de Pablo, Phys. Rev. Lett. **92**, 155505 (2004).

[72]A. Stukowski, Modell. Simul. Mater. Sci. Eng. **18**, 015012 (2009).

[73]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, Mol. Cell **57**, 936 (2015).

[74]R. Wu, M. Deng, B. Kong, and X. Yang, J. Phys. Chem. B **113**, 15010 (2009).

[75]J. Kuldová, P. Košovan, Z. Limpouchová, and K. Procházka, Macromol. Theory Simul. **22**, 61 (2013).

[76]A. Stukowski, JOM **66**, 399 (2014).

[77]J. V. Sheth, V. Sahni, S. F. Shandarin, and B. S. Sathyaprakash, Mon. Not. R. Astron. Soc. **343**, 22 (2003).

[78]A. Putnam, M. Cassani, J. Smith, and G. Seydoux, Nat. Struct. Mol. Biol. **26**, 220 (2019).

[79]S. Jain, J. Wheeler, R. Walters, A. Agrawal, A. Barsic, and R. Parker, Cell **164**, 487 (2016).

[80]F. Chiti and C. M. Dobson, Annu. Rev. Biochem. **75**, 333 (2006).

[81]G. Pandav, V. Pryamitsyn, K. C. Gallow, Y.-L. Loo, J. Genzer, and V. Ganesan, Soft Matter **8**, 6471 (2012).

[82]A. Irbäck and E. Sandelin, Biophys. J. **79**, 2252 (2000).

[83]A. Irbäck, C. Peterson, and F. Potthast, Proc. Natl. Acad. Sci. U. S. A. **93**, 9533 (1996).

[84]P. J. Flory, Trans. Faraday Soc. **51**, 848 (1955).

[85]C. L. P. Shan and L. G. Hazlitt, Macromol. Symp. **257**, 80 (2007).