

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

Antonia Statt,^{1, 2, a)} Helena Casademunt,^{3, 4, a)} Clifford P. Brangwynne,^{1, 5} and Athanassios Z. Panagiotopoulos^{1, b)}

¹⁾*Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544*

²⁾*Present address: Materials Science and Engineering, Grainger College of Engineering, University of Illinois, Urbana-Champaign, IL, 61801*

³⁾*Department of Physics, Princeton University, Princeton, NJ 08544*

⁴⁾*Present address: Department of Physics, Harvard University, Cambridge, MA 02138*

⁵⁾*Howard Hughes Medical Institute, Chevy Chase, MD 20815*

(Dated: 19 January 2020)

^{a)}H. Casademunt and A. Statt contributed equally to this work.

^{b)}email:azp@princeton.edu

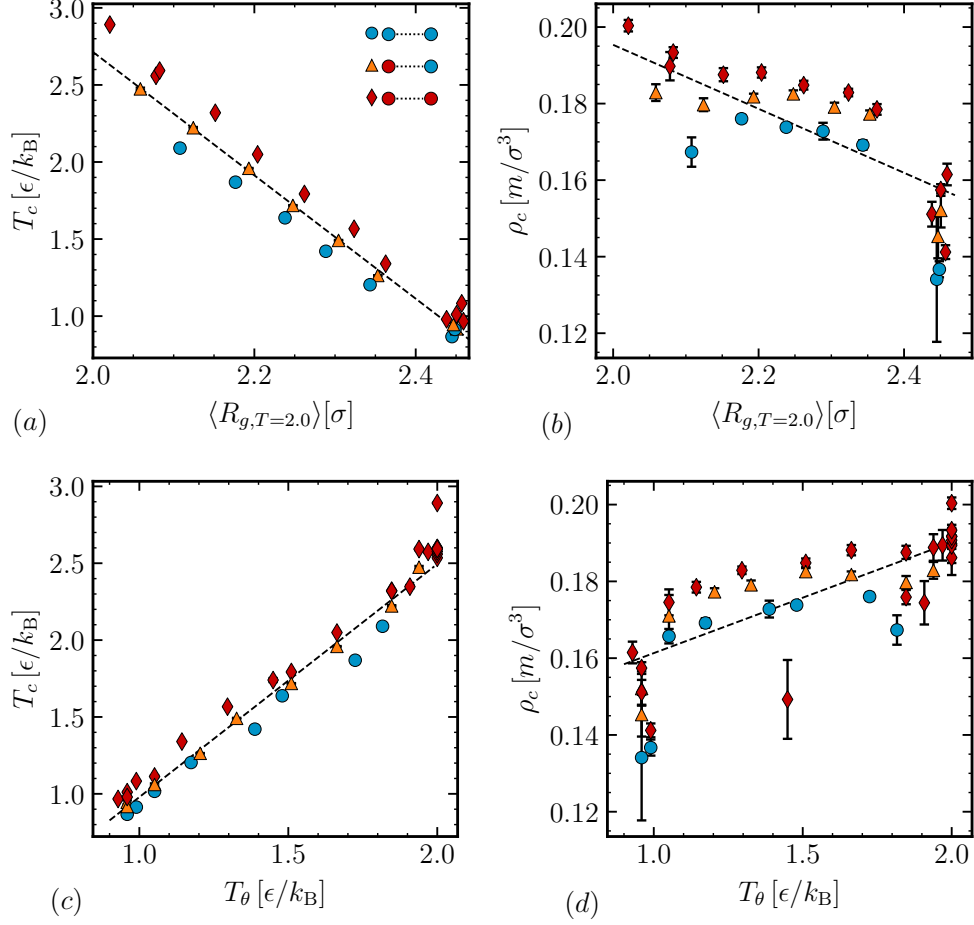


FIG. S1. (a) Scaling of the critical temperature T_c with the radius of gyration of a single chain $\langle R_g \rangle$ measured at a fixed temperature of $T = 2.0\epsilon/k_B$. (b) Same scaling as in (a) for the critical density ρ_c . (c) Scaling of T_c with T_θ . (d) Scaling of ρ_c with T_θ . Red diamonds ($\blacklozenge$) indicate sequences where both terminal beads are T, blue circles ($\bullet$) mark sequences with two H ends, and orange triangles ($\blacktriangle$) indicate the sequences with mixed terminal beads. The dashed lines are linear fits to the data for sequences with mixed terminal beads.

A. Smoothing function for the pair potential

For the LJ pair potential, we applied a smoothing polynomial $s(r)$

$$s(r) = \begin{cases} 1 & r \leq r_s, \\ \frac{(r_c^2 - r^2)^2 (r_c^2 + 2r^2 - 3r_s^2)}{(r_c^2 - r_s^2)^2} & r_s < r \leq r_c, \\ 0 & r > r_c, \end{cases} \quad (1)$$

which ensures that the pairwise potential and forces transition smoothly to zero at the truncation radius r_c . In this work, we chose $r_c = 3\sigma$ and began smoothing from $r_s = 2.5\sigma$.

B. Scaling of single-chain properties with f_T

As previously suggested, the single chain radius of gyration at a fixed temperature¹⁻³ should scale with the critical temperature and density. As shown in Figs. S1(a) and (b), this holds true for our model as well. While the critical temperature T_c scaled rather well with R_g , the correlation with the critical density ρ_c was less pronounced.

Since the critical temperature T_c and the θ -temperature T_θ are also connected^{1,4} and a better scaling relationship with T_θ for chains of different lengths is expected, we calculated T_θ . We determined the coil-to-globule transition from the average bead distance R_{ij} between bead i and j along the chain. The temperature where the coil-to-globule occurs is the θ -temperature. For details on this method we refer the reader to Dignon *et al.*¹ The results are shown in Figs. S1(c) and (d). Similar to R_g , R_{ij} showed roughly linear scaling, with better results for T_c than ρ_c .

Both R_g and R_{ij} can provide an estimate of the critical point location; however, they fail to capture the systematic influence of the terminal beads found in our simulations. As illustrated in section IIIC2, it is also not possible to predict if a certain sequence will phase separate or not by using R_g or R_{ij} .

C. Sequences with $f_T = 0.95$

Fig. S2(a) shows the phase envelopes of all sequences with $f_T = 0.95$, and Figs. S2(b) and (c) show the scaling of the critical point with the position of the repulsive hydrophilic head, measured as the distance from the end of the chain.

As shown in Figs. S3(a) and (b), the interfacial composition changes significantly with both temperatures and position of H along the chain. The interface location was found by fitting a tanh profile to the density histogram. We defined the interfacial region based on the location where the tanh reached 10% and 90% of its bulk density value. For each sequence and at each temperature, the interface of the liquid had a unique composition of H and T beads, as well as a unique relative percentage of end beads. When the hydrophilic H bead is located at the end of the chain, a significant increase in the number of end beads in the interfacial region in comparison to a random distribution was observed below the critical point. When the hydrophobic bead was closer to the center of the chain, we found only a weak enhancement of end beads in the interfacial region. For $T_{19}H$, $T_{18}HT$, and $T_{17}HT_2$, we even observed fewer end beads in the interfacial region than expected from a purely random distribution below the critical point.

As a rough estimate of the surface tension, we determined $\hat{\gamma} = \frac{k_B T}{2\pi w^2} \ln L$ from the interfacial width w , as determined by the tanh fit.⁵ Results are shown in Fig. S3(c). Here, L is the box size in x and y dimensions. In order to determine the true interfacial tension, we would need the bulk correlation length l_b , which is of the order of the molecular size σ . The value of $\hat{\gamma}$ varies with temperature and sequence, but it can be rescaled onto one master curve. All sequences collapsed when plotted against the distance to the critical point $1 - T/T_c$ of each sequence, as shown in Fig. S3(d). This is expected from the scaling law $\gamma = \gamma_0(1 - T/T_c)^\mu$, with $\mu \approx 1.26$ from the Ising Universality class,⁶ and it illustrates the importance of interface composition. The position of the H bead changes the interfacial composition at a fixed temperature, which, in turn, changes $\hat{\gamma}$ at that temperature. Because T_c and the interfacial tension are connected, a change in interfacial composition leads to a change in the critical temperature.

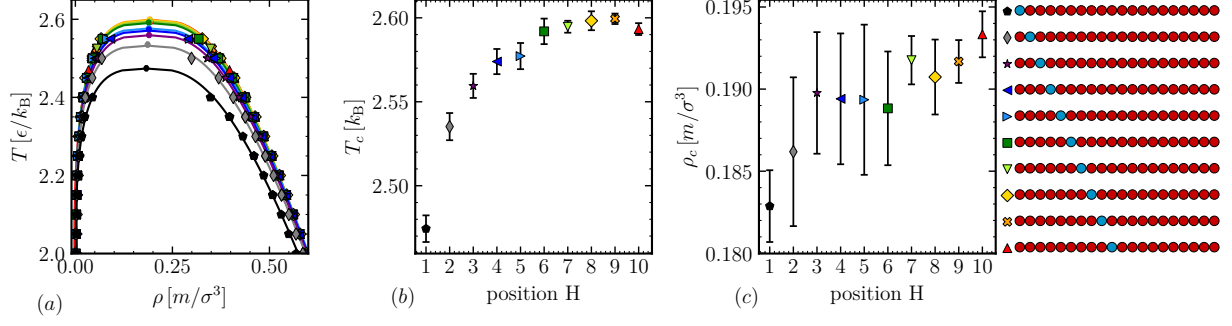


FIG. S2. (a) Phase diagrams for sequences with $f_T = 0.95$ with varying position of the one hydrophilic bead. The locations of the critical points are indicated by circles; the lines show the fit of Eqs. (4)–(5). (b) Scaling of the critical temperature and (c) of the critical density as a function of the position of the repulsive H (●) bead going from the end of the chain (1) to the middle (10).

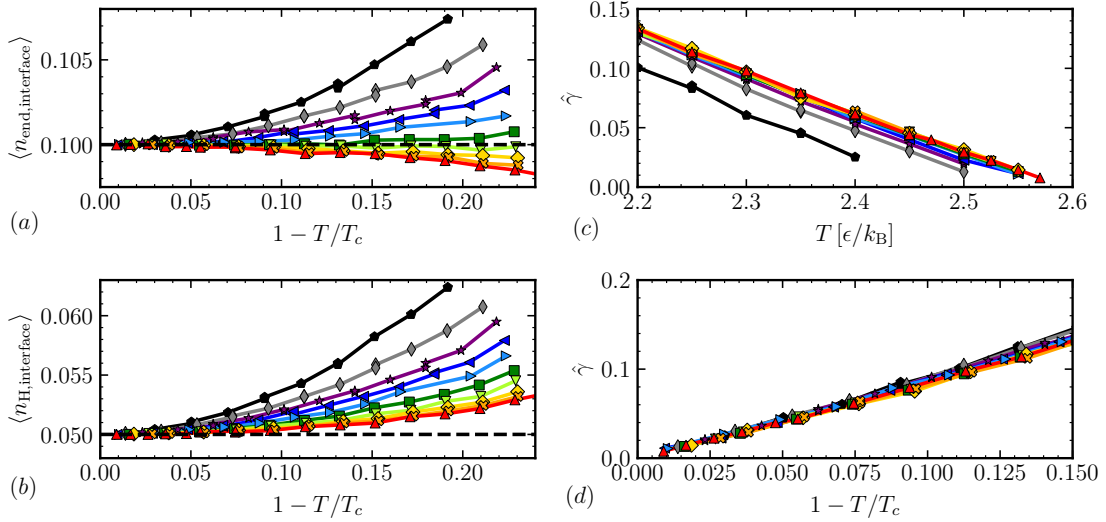


FIG. S3. Average number of end beads (a), and average number of H beads (b) in the interfacial region. The dashed lines indicate the average numbers for completely random distributions. $\hat{\gamma}$ is shown as a function of temperature (c) and as a function of distance to the critical temperature, $1 - T/T_c$ (d). The legend is the same as in Fig. S2.

D. Density histograms

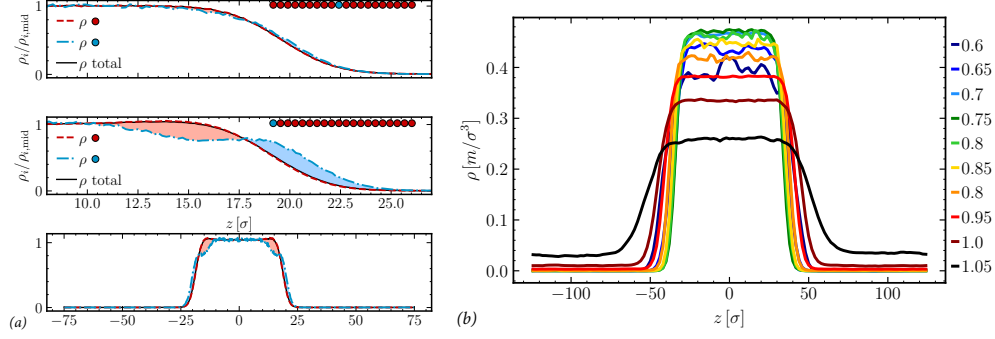


FIG. S4. Density histograms (a) for two $f_T = 0.95$ sequences. The top panel and middle panel show the interfacial region, and the bottom panel shows the full density histogram for HT_{19} . (b) Density histograms of a single simulation run for different temperatures as indicated of the re-entrant sequence $T_3H_3T_3H_2T_3H_3T_3$. The corresponding phase diagram is shown in Fig.4 of the main text.

E. Logarithmic Phase diagram

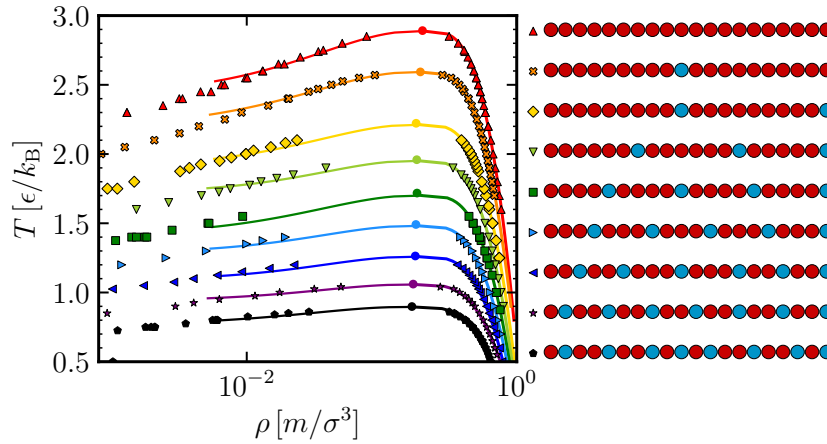


FIG. S5. Same data as Fig.1, plotted on a logarithmic scale to highlight the dilute phase densities. The locations of the critical points are shown by circles and the lines show the fit of Eqs. (4)–(5) to the upper part of the phase envelope. The legend shows the corresponding sequences (sorted from high to low T_c) with filled red circles for T beads (●) and filled blue circles for H beads (●).

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

Fig. S6 shows the phase diagrams for all the sequences with $f_T = 0.6$ that showed conventional phase separation or reentrant behavior.

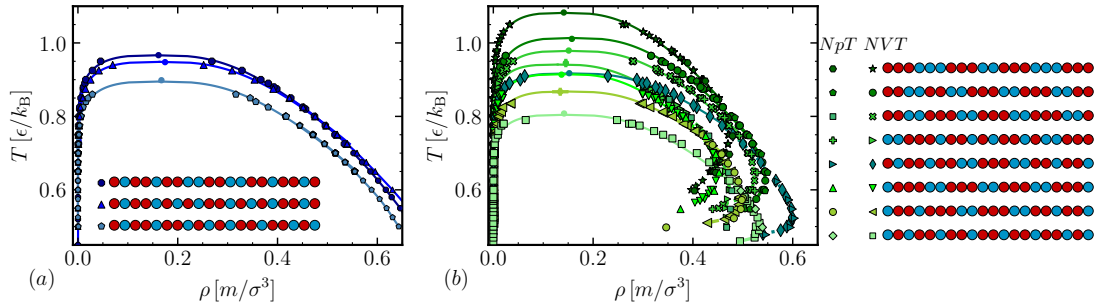


FIG. S6. (a) Coexistence densities for sequences with $f_T = 0.6$ showing conventional liquid-like phase behavior. (b) Coexistence densities for all sequences with $f_T = 0.6$ showing reentrant phase behavior. The densities were measured in NVT and NpT ensembles as indicated. The locations of the critical points are indicated by circles, and the solid lines show the fit of Eqs. (4)–(5) to the upper part of the NVT phase envelope.

G. Structure of the liquid and large-scale aggregates

For ordered proteins, it is common to determine the protein contact map to identify structures like α helices or β sheets within the protein. Even though the subject of this study is disordered proteins, these are not completely unstructured, so we determined the

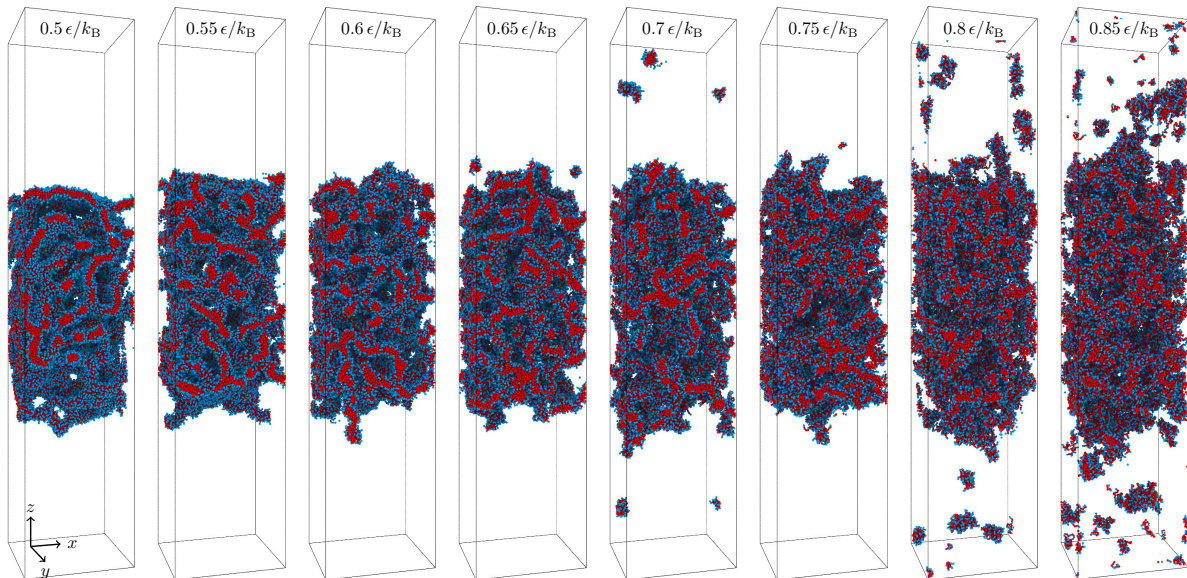


FIG. S7. Typical configurations of the regular sequence $[\text{T}_3\text{H}_2]_4$ () at different temperatures. Snapshots were generated using OVITO⁷.

intra-chain distances for the different sequences to quantify their degree of ordering. Fig. S8 shows the distance $\text{dist}(i, j)$ of monomer i to j in comparison to the expected random coil value for different sequences. Blue indicates a larger than expected distance, red indicates a smaller than expected distance. All sequences except the liquid-like systems show some structure in the distance map at cold temperatures. Aside from the micelle-like systems, it is not possible to unambiguously identify the type of large-scale aggregate from the distance maps.

Other short-ranged chain properties like average bond distances, angles and dihedral angles did not show distinctive features for the different large-scale aggregates, as visible in Figs. S9(a)–(c). Inter-chain structural properties like the pair correlation function, shown in Fig. S10(a), also did not reveal a clear way to identify the aggregates, and neither did the radius of gyration, which we determined for a single chain as a function of temperature, shown in Fig. S10(b). We measured the properties of the clusters of the system for all sequences forming aggregates. Fig. S11(a) shows results for the average largest cluster size in the system. In all cases except for the systems that form isolated spherical micelles (T_{12}H_8 and $\text{H}_4\text{T}_{12}\text{H}_4$) and a membrane-forming system ($\text{H}_3[\text{T}_4\text{H}]_3\text{H}_2$) most chains condensed into one single large cluster at both densities and all investigated temperatures, as indicated by $\langle N_l/N_{\text{tot}} \rangle \approx 1$ in Fig. S11(a).

Fig. 5 and 6 in the main text showed that the different morphologies had a different degree of microphase separation between the hydrophobic and hydrophilic beads. To quantify this effect, we calculated the average number of T neighbors within a distance of 2σ of T beads n_{TT} . This quantity, normalized by the total number of neighbors n_{tot} , is shown in Fig. S11(b). Note that while the absolute value of $\langle n_{\text{TT}}/n_{\text{all}} \rangle$ shifted for different temperatures, the ordering of the different sequences remained the same. While we observed a rough trend where liquids showed the lowest and micelles showed the highest values of microphase separation, we cannot clearly distinguish different types of large-scale aggregates.

An interesting difference between the conventional condensed phases and the large-scale aggregates is the variance of local densities. We divided the condensed phase into smaller subsections of size $5\sigma \times 5\sigma \times 5\sigma$ and determined the density ρ_i in each. The resulting histograms for different temperatures are shown in Fig. S12(a) for three example sequences. These results hold true for all studied sequences. For a conventional condensed phase (top panel), the distribution of densities in the dense phase shifts to higher densities and becomes narrower as temperature decreases, as expected. This can also be seen in Fig. S12(b), where we plot the variance of the measured densities ρ_i as a function of temperature. For reentrant phase behavior (middle panel), the distributions first shift like in the conventional case, but then flatten out. Some sub-regions contained only a small amount of chains or no chains at all, with $\rho_i \approx 0$, while others were dense. This is a reflection of the formation of fibril-like structures, where some regions in space are very dense and others have large voids. Consequently, the variance first decreases and then increases with lowering the temperature. For systems that only formed large-scale aggregates (bottom panel), all histograms look similar and do not have a distinguishable peak at a finite density. The variance is high and increases with decreasing temperature, the opposite of a conventional liquid. This behavior can be explained by the increasing degree of microphase separation and the formation of more and more large-scale voids and fibril-like aggregates, making the system increasingly heterogeneous at lower temperatures.

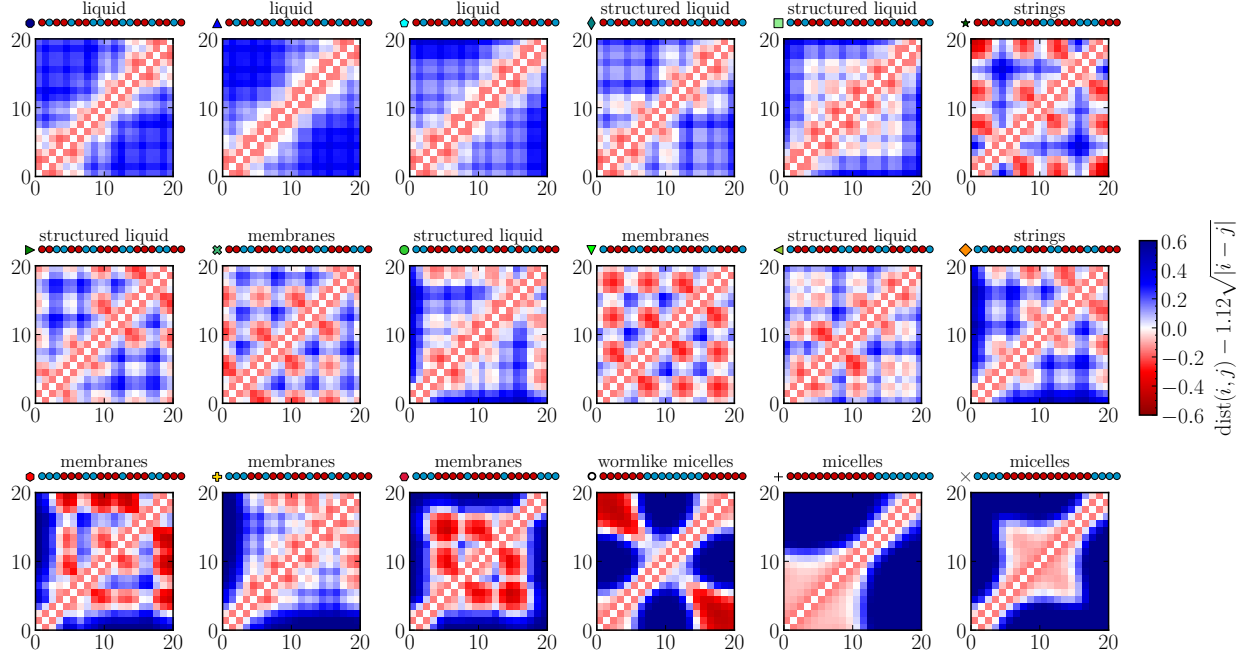


FIG. S8. Intra-chain distance maps of bead i to j along a sequence for all sequences with $f_T = 0.6$. The density was $\rho_{\text{tot}} = 0.1m/\sigma^3$ and the temperature was $T = 0.55\epsilon/k_B$. The color indicates the deviation of the measured distance from the expected random coil value, $1.12\sqrt{|i-j|}$. The labels were assigned according to the classification in Fig. 8.

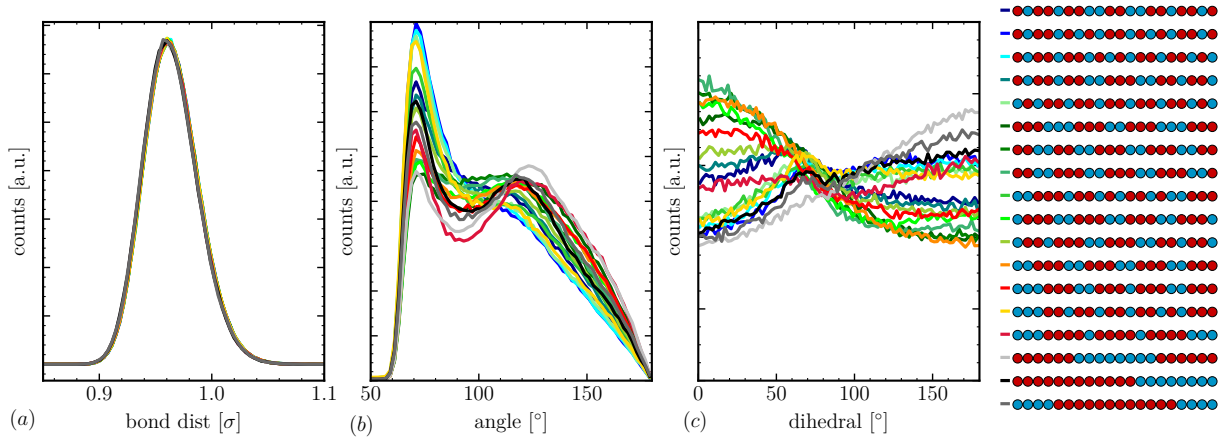


FIG. S9. (a) Average bond distance, (b) average bond angle, and (c) dihedral angle distributions of the different sequences with $f_T = 0.6$ at $T = 0.55\epsilon/k_B$ and $\rho_{\text{tot}} = 0.05m/\sigma^3$.

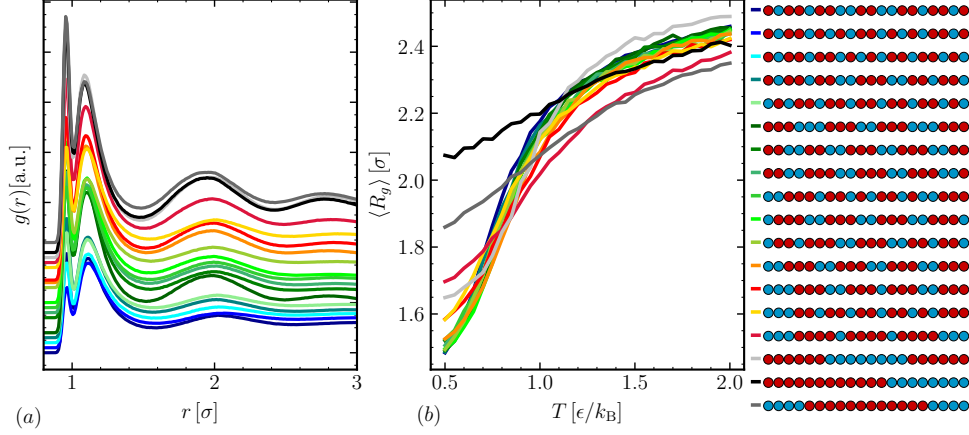


FIG. S10. (a) Pair correlation function $g(r)$ for all the sequences with $f_T = 0.6$ at $T = 0.55\epsilon/k_B$ and $\rho_{\text{tot}} = 0.05m/\sigma^3$. The pair correlation functions are shifted horizontally for clarity. (b) Average radius of gyration $\langle R_g \rangle$ of an isolated chain as function of temperature.

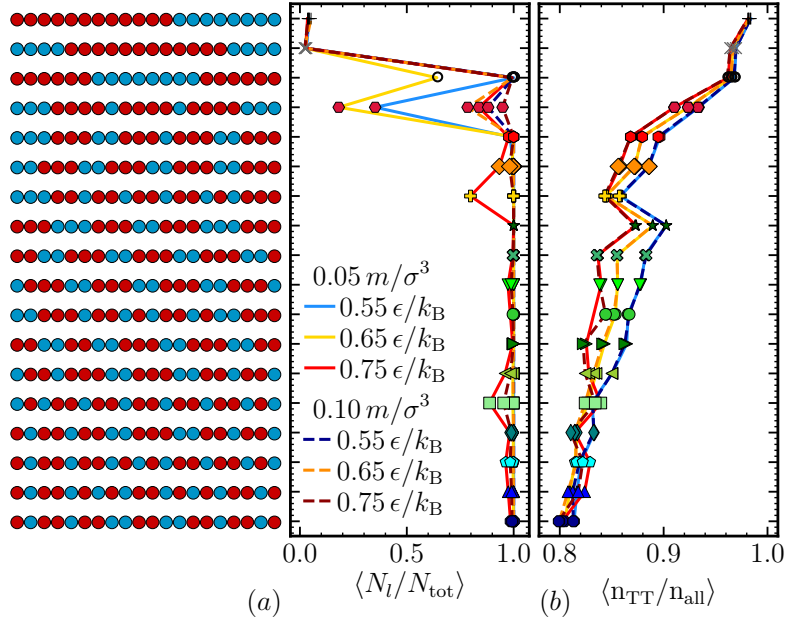


FIG. S11. (a) Average fraction of chains in the largest cluster $\langle N_l/N_{\text{tot}} \rangle$ and (b) average normalized number of T neighbors of a T bead $\langle n_{\text{TT}}/n_{\text{all}} \rangle$. The different lines correspond to three temperatures for $\rho_{\text{tot}} = 0.05 m/\sigma^3$ (solid) and $\rho_{\text{tot}} = 0.1 m/\sigma^3$ (dashed). The y-axis represents all sequences with $f_T = 0.6$, with T beads in red ($\bullet$) and H beads in blue ($\bullet$).

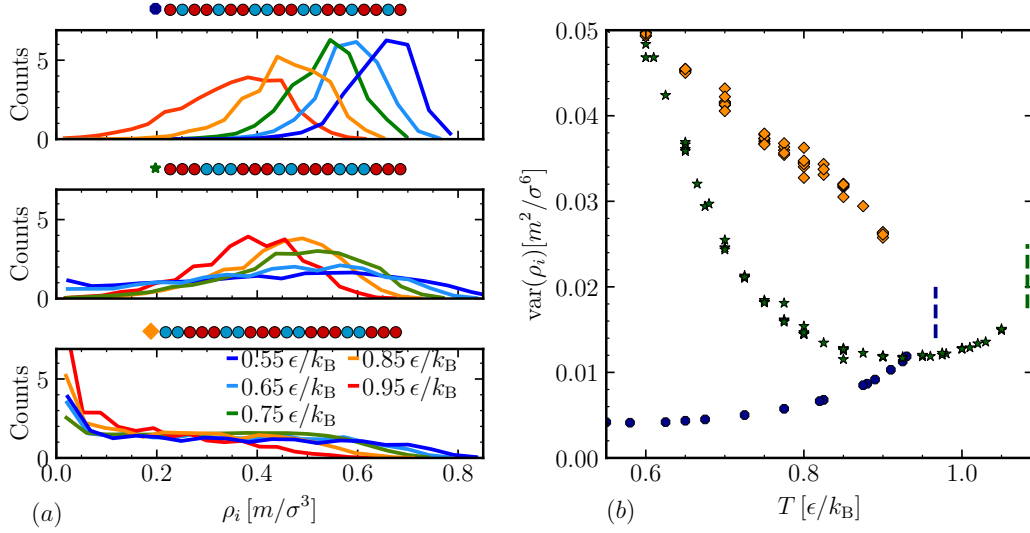


FIG. S12. (a) Histograms of local densities ρ_i in $5\sigma^3$ sub-regions within liquid phase at five different temperatures. The three panels correspond to sequences with: a conventional condensed phase (top), a reentrant phase (middle), and string-like aggregates (bottom). (b) Variance of the local density histograms as a function of temperature for the same three sequences. Dashed lines indicate the location of the critical temperatures of the sequences with conventional and reentrant phase behavior.

H. Order parameters

In this work, we have tested multiple order parameters. All results are presented in Fig. S13. We have chosen to sort the sequences according to their phase behavior (separated by horizontal dashed lines) and then within each group according to their critical temperature, if known. The y -axis in Fig. S13 indicates each sequence in order.

The first and simplest choice of order parameter is the the average length L_T of the hydrophobic segments in the sequence,⁸ shown in Fig. S13. This order parameter roughly sorts the sequences, but does not capture the systematic influence of the terminal beads. We also expect that L_T will not perform well for comparing multiple chain lengths.

We studied the normalized mean-square fluctuation $\Psi^{(s)}$ of a sub-section of length s .^{9,10} We varied the length of the sub-sections/blocks s and achieved best results with $s = 5$. With a few exceptions, $\Psi^{(5)}$ sorts the sequences according to their critical temperature, but cannot be used to distinguish different phase and aggregation behavior, as it only seems to separate out micelle-forming systems.

The order parameter κ ,¹¹ commonly used for proteins, does not perform well in this system. We picked a residue blob size of 5. While κ can distinguish different trends within a set of sequences with the same f_T , it does not perform as well across different values of f_T . We suspect that the reason for this is that κ is normalized to its maximal value, which depends on f_T . To unify our choice for all values of f_T , we normalized against the κ value of a corresponding block of composition f_T . It is possible that a different choice of normalization might improve results.

We also determined the (corrected) probability of finding a T segment after a T segment,

$P_{TT} - f_T$, where values of $P_{TT} - f_T \approx 0$ correspond to random sequences, $P_{TT} - f_T < 0$ to alternating, and $P_{TT} - f_T > 0$ to blocky sequences.^{12,13} We observed both negative and positive values within each type of phase behavior except for the sequences that formed finite-sized aggregates, which were all classified as blocky.

The sequence charge decoration (SCD) is the commonly used order parameter that performed best out of the ones we studied. However, it did not reproduce the systematic influence of the terminal beads. SCD, like the κ parameter, is frequently used for proteins,¹⁴ as opposed to the other studied parameters, which are usually applied to co-polymer systems.

Since the order parameters from the literature generally performed poorly in our system, we defined an effective reweighted f_T^* as

$$f_T^* = \frac{\sum_{i=1}^M t_i}{1 - \sum_{i=1}^M \frac{T_c^{0.95}(i)}{T_c^{1.00}}}, \quad t_i = \begin{cases} 1 - \frac{T_c^{0.95}(i)}{T_c^{1.00}} & \text{for } i = \text{T} \\ 0 & \text{for } i = \text{H} \end{cases} \quad (2)$$

where $T_c^{1.00}$ is the critical temperature of the pure homopolymer and $T_c^{0.95}(i)$ is the critical temperature of the chain with only one H at position i . We have given H beads zero weight because they do not have an energetic contribution to the phase separation, and T beads a weight that depends on their position in the chain.

By defining f_T^* as described above, we were able to account for the fact that hydrophobic beads closer to the ends of the chain appear to be more important in promoting phase separation. We achieved a fairly linear correlation between f_T^* and T_c for all investigated sequences which had a critical point, as visible in Fig. 9 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 terminal beads for the location of the critical point, because we were able to account for their effect with reweighting.

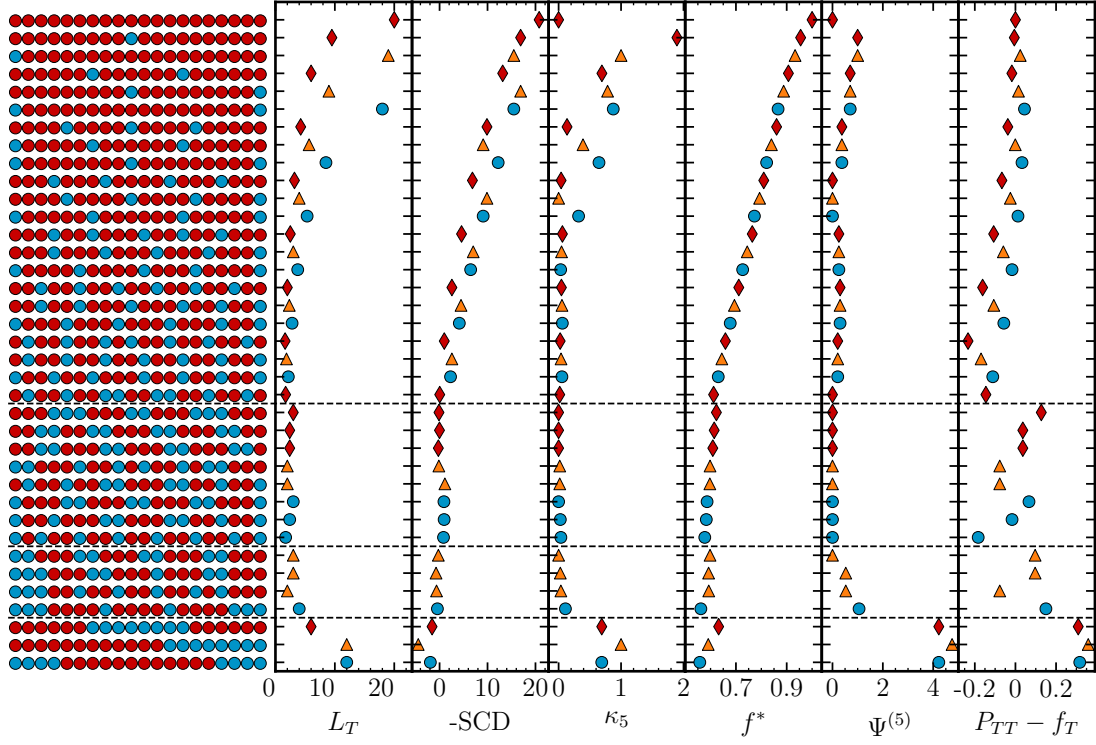


FIG. S13. Red diamonds ($\blacklozenge$) indicate sequences where both terminal beads are T, blue circles ($\bullet$) mark sequences with two H ends, and orange triangles ($\blacktriangle$) indicate the sequences with mixed terminal beads. The dashed vertical lines separate sequences which form finite aggregates (bottom), then infinite-range aggregates, then sequences that show reentrant phase behavior, and finally sequences that show conventional dense-dilute phase separation.

TABLE S1. Sequence names, architectures, critical point location T_c and ρ_c , and fraction of attractive beads f_T . Subscript indicates error on the last digit.

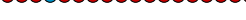
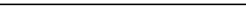
sequence	architecture	T_c	ρ_c	f_T
T ₂₀		2.892 ₅	0.200 ₂	1.0
HT ₁₉		2.475 ₈	0.183 ₂	0.95
THT ₁₈		2.535 ₉	0.186 ₅	0.95
T ₂ HT ₁₇		2.561 ₇	0.189 ₄	0.95
T ₃ HT ₁₆		2.575 ₈	0.189 ₄	0.95
T ₄ HT ₁₅		2.577 ₉	0.189 ₅	0.95
T ₅ HT ₁₄		2.593 ₇	0.189 ₃	0.95
T ₆ HT ₁₃		2.595 ₄	0.192 ₂	0.95
T ₇ HT ₁₂		2.598 ₅	0.191 ₂	0.95
T ₈ HT ₁₁		2.599 ₃	0.192 ₁	0.95
T ₉ HT ₁₀		2.599 ₂	0.1939 ₉	0.95
HT ₁₈ H		2.090 ₇	0.167 ₄	0.9
[T ₆ H] ₂ T ₆		2.319 ₄	0.188 ₂	0.9
[T ₉ H] ₂		2.223 ₈	0.175 ₃	0.9
[T ₆ H] ₂ T ₅ H		1.959 ₂	0.1818 ₈	0.85
HT ₈ HT ₉ H		1.869 ₂	0.1760 ₁	0.85
[T ₄ H] ₃ T ₅		2.049 ₂	0.188 ₁	0.85
HT ₅ HT ₆ HT ₅ H		1.637 ₂	0.1837 ₇	0.8
[T ₃ H] ₂ T[T ₃ H] ₂ T ₃		1.792 ₃	0.187 ₁	0.8
[T ₄ H] ₄		1.716 ₄	0.186 ₁	0.8
[HT ₄] ₂ HT ₃ HT ₄ H		1.425 ₂	0.1818 ₇	0.75
[T ₃ HT ₂ H] ₂ T ₂ HT ₃		1.567 ₂	0.1830 ₉	0.75
[T ₃ H] ₅		1.491 ₃	0.179 ₁	0.75
[HT ₃] ₄ HT ₂ H		1.208 ₁	0.1777 ₇	0.7
[T ₂ H] ₆ T ₂		1.340 ₂	0.178 ₁	0.7
[T ₂ H] ₂ [T ₃ H] ₂ [T ₂ H] ₂		1.264 ₂	0.1773 ₉	0.7
[HT ₂] ₃ T[HT ₂] ₃ H		1.018 ₁	0.1694 ₁	0.65
[THT] ₃ HT[THT] ₃		1.115 ₉	0.175 ₅	0.65
TH[T ₂ H] ₆		1.065 ₄	0.181 ₂	0.65
H ₂ T ₃ HT ₂ HTHT ₂ HHT ₂ T ₃		0.95 ₁	0.145 ₆	0.6
[HT] ₂ TH[T ₂ H] ₄ TH		0.808 ₆	0.142 ₅	0.6
HT ₂ HT ₃ H ₂ T ₃ H ₂ T ₂ HT ₂ H		0.87 ₁	0.13 ₂	0.6
H[T ₃ H ₂] ₃ T ₃ H		0.914 ₄	0.137 ₂	0.6
THT ₂ H[TH] ₂ T ₂ H[TH] ₂ T ₂ HT		0.948 ₂	0.175 ₁	0.6
THT ₂ H[T ₂ H ₂] ₂ [T ₂ H] ₂ T		0.967 ₃	0.161 ₃	0.6
THT ₂ HT ₂ H ₂ T ₃ H ₂ [T ₂ H] ₂		0.918 ₃	0.152 ₄	0.6
HT[T ₂ H] ₂ [TH] ₂ [T ₂ H] ₂ TH		0.898 ₂	0.167 ₁	0.6
[T ₂ H ₂] ₂ [T ₃ H ₂] ₂ T ₂		1.011 ₂	0.157 ₂	0.6
T ₂ H ₂ [T ₃ H ₂] ₃ T		0.979 ₄	0.151 ₃	0.6
T ₃ H ₃ T ₃ H ₂ T ₃ H ₃ T ₃		1.083 ₃	0.141 ₂	0.6

TABLE S2. Names and architectures of sequences without a critical point.

sequence	architecture	f_T
$[H_2T_3]_4$		0.6
$H_3[T_2H]_4THT_3$		0.6
$H_3T_3H_2T_3HT_3H_2T_3$		0.6
$H_3[T_4H]_3H_2$		0.6
$H_4T_{12}H_4$		0.6
$T_6H_8T_6$		0.6
$T_{12}H_8$		0.6

REFERENCES

- ¹G. 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," *Proceedings of the National Academy of Sciences* **115**, 9929–9934 (2018).
- ²Y.-H. Lin and H. S. Chan, "Phase separation and single-chain compactness of charged disordered proteins are strongly correlated," *Biophysical Journal* **112**, 2043 – 2046 (2017).
- ³Y.-H. Lin, J. P. Brady, J. D. Forman-Kay, and H. S. Chan, "Charge pattern matching as a 'fuzzy' mode of molecular recognition for the functional phase separations of intrinsically disordered proteins," *New Journal of Physics* **19**, 115003 (2017).
- ⁴A. Z. Panagiotopoulos, V. Wong, and M. A. Floriano, "Phase equilibria of lattice polymers from histogram reweighting monte carlo simulations," *Macromolecules* **31**, 912–918 (1998).
- ⁵S. Senapati and M. L. Berkowitz, "Computer simulation study of the interface width of the liquid/liquid interface," *Phys. Rev. Lett.* **87**, 176101 (2001).
- ⁶J. S. Rowlinson and B. Widom, *Molecular theory of capillarity* (Courier Corporation, 2013).
- ⁷A. Stukowski, "Visualization and analysis of atomistic simulation data with OVITO—the open visualization tool," *Modelling and Simulation in Materials Science and Engineering* **18**, 015012 (2009).
- ⁸G. Pandav, V. Pryamitsyn, K. C. Gallow, Y.-L. Loo, J. Genzer, and V. Ganesan, "Phase behavior of gradient copolymer solutions: a monte carlo simulation study," *Soft Matter* **8**, 6471–6482 (2012).
- ⁹A. Irbäck and E. Sandelin, "On hydrophobicity correlations in protein chains," *Biophysical Journal* **79**, 2252 – 2258 (2000).
- ¹⁰A. Irbäck, C. Peterson, and F. Potthast, "Evidence for nonrandom hydrophobicity structures in protein chains," *Proceedings of the National Academy of Sciences* **93**, 9533–9538 (1996).
- ¹¹R. K. Das and R. V. Pappu, "Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues," *Proceedings of the National Academy of Sciences* **110**, 13392–13397 (2013).
- ¹²P. J. Flory, "Theory of crystallization in copolymers," *Trans. Faraday Soc.* **51**, 848–857 (1955).
- ¹³C. L. P. Shan and L. G. Hazlitt, "Block index for characterizing olefin block copolymers," *Macromolecular Symposia* **257**, 80–93 (2007).

¹⁴S. Das, A. N. Amin, Y.-H. Lin, and H. S. Chan, “Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters,” *Physical Chemistry Chemical Physics* **20**, 28558–28574 (2018).