Complete phase diagram for liquid-liquid phase separation of intrinsically disordered proteins

James McCarty,[†] Kris T. Delaney,[‡] Scott P. O. Danielsen,[‡] Glenn H. Fredrickson,[‡] and Joan-Emma Shea[†]

Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA., Materials Research Laboratory, University of California, Santa Barbara, California 93106, USA., Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA., Materials Department, University of California, Santa Barbara, California 93106, USA., and Department of Physics, University of California Santa Barbara, Santa Barbara, California 93106, USA.

E-mail:

[*]To whom correspondence should be addressed

[†]Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, USA.

[‡]Materials Research Laboratory, University of California, Santa Barbara, California 93106, USA.

[¶]Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA.

[§]Materials Department, University of California, Santa Barbara, California 93106, USA.

[‖]Department of Physics, University of California Santa Barbara, Santa Barbara, California 93106, USA.

Abstract

A number of intrinsically disordered proteins have been shown to self-assemble via liquid-liquid phase separation into protein-rich and dilute phases. The resulting coacervates can have important biological functions, and the ability to form these assemblies is dictated by the protein's primary amino acid sequence as well as by the solution conditions. We present a complete phase diagram for the simple coacervation of a polyampholyte intrinsically disordered protein using a field theoretic simulation approach. We show that differences in the primary amino acid sequence and in the distribution of charged amino acids along the sequence leads to differences in the phase window for coacervation, with block-charged sequences having a larger coacervation window than sequences with a random patterning of charges. The model also captures how changing solution conditions modifies the phase diagram and can serve to guide experimental studies.

Intrinsically disordered proteins (IDP), unlike globular proteins, do not fold to a well-defined folded three-dimensional structure, but rather populate a variety of partially structured, marginally stable conformations under physiological conditions. IDPs play a variety of roles in the cell, from cell signaling to transduction, functions that take advantage of the IDPs ability to adopt structures that can bind to different partners.[1–3] The lack of a well-folded structure with a buried hydrophobic core heightens the likelihood that these structures will aggregate in a concentrated cellular environment. A well-characterized aggregated form of IDPs is the amyloid fibril, a solid entity with an ordered cross-β-sheet structure, that deposit on organs in the body, and is a hallmark of amyloid diseases.[4,5] A growing number of IDPs have also been observed to undergo a different form of assembly process, a liquid-liquid phase separation (LLPS) known as coacervation, which manifests as the formation of protein-rich liquid droplets of biomolecular aggregates.[6–11] The process of coacervation has been observed in proteins associated with amyloid-diseases (for example the tau protein implicated in tauopathies,[10] and the FUS protein linked to ALS[8]), which has led to speculation that liquid droplet formation might play a role in the IDP fibrillization process.[12–14]

It remains unclear whether the droplets are on-pathway to fibril formation, maturing from a liquid state to a solid state, or whether the droplets rather sequester peptides and offer an alternate to fibrillization. In some instances, the role of the liquid droplets appears to be purely physiological, forming membrane-less organelles such as the Cajal and nuclear bodies, storing peptides, or serving as reactive centers in which biomolecular reactions can proceed at an accelerated rate.[6,11,15,16]

While it is recognized that the unique amino acid sequence for a given IDP ultimately determines the phase behavior, few theories are able to model polyampholyte coacervation for heterogenous polymers such as IDPs.[17] In IDPs, as opposed to synthetic polyelectrolytes, the placement of charges along the chain is not random, and one expects correlations among charges introduced by chain connectivity to play an important role in controlling the phase behavior. However, these intramolecular charge correlations are neglected by mean-field approaches such as the widely used Voorn-Overbeek (VO) model.[18,19] Recent theoretical work has called attention to the fact that the simple VO model neglects this important physics, leading to incorrect scaling predictions over all concentration ranges.[20] If a polymer physics model is to be applied to IDPs, it must properly account for heterogeneous sequences and make predictions based on the specific charge distribution of a given IDP. Explicit charge patterning was introduced in a recently proposed theoretical model[21,22] based on the random phase approximation (RPA),[23] demonstrating that the pattern of charges has a large effect on the phase diagram.[24] The RPA model is a marked improvement over mean-field approaches and can be formally derived by considering Gaussian fluctuations in the fields.[25] Despite its qualitative success, the RPA is expected to break down under conditions at which field fluctuations become large. For example, Monte Carlo simulations of lattice chains have shown that the RPA model over-predicts the dependence of charge sequence on the stability of the homogeneous phase.[26] Recent theoretical efforts have aimed at a more sophisticated treatment of fluctuations and electrostatic correlations either through a renormalization approach[27,28] or a hybrid approach in which theory is informed from

simulation.[29–31]

Complementary to analytical theories are numerical methods, such as molecular dynamics (MD) simulation. All-atom explicit-water MD is useful to study the early stages of aggregation in short chains, but is computationally expensive, which hinders its applicability for studying IDP phase separation. On the other hand, coarse-grained models are useful for understanding how the parameterization of the potential of mean force acting between effective segments of the polyampholyte model can affect the phase diagram.[32] For example, an approach using a slab geometry developed for vapor-liquid transitions,[33,34] has recently been introduced to achieve a converged density profile within a coarse-grained MD simulation, enabling the computation of the phase diagram.[32,35] Still, these methods require long equilibration times to obtain the phase coexistence points and suffer from finite size effects due to the limited number of chains that can be reasonably included in the simulation box.

An alternative to these theoretical and particle-based approaches is to employ field theoretic simulation (FTS),[36,37] which is a numerical technique based on the formal equivalence between a particle-based representation and a statistical field description in which ensemble averages are evaluated as path integrals over spatially varying chemical and electrostatic potential field configurations. A major advantage of FTS over other methods such as self-consistent field theory or the RPA is that FTS stochastically samples the fully fluctuating field theoretic Hamiltonian, making it possible to generate phase diagrams without additional approximations. In the context of IDP phase separation, FTS is appealing because one can efficiently simulate large dense systems without invoking uncontrolled approximations, as done in the RPA, and without the need for long equilibration times, as required by particle-based MD. This method has been applied to study polyelectrolyte coacervation for relatively simple homopolymers and block co-polymers.[20,38–40] We introduce in this letter a discrete-chain implementation to study the effect of sequence variation at the individual monomer level, a critical feature in order to describe IDP coacervation. IDP coacervation can occur in a simple or complex manner (the latter requiring the presence of a second

type of molecule, typically RNA), but we focus here on simple coacervation, noting that the methodology presented can be readily extended to complex coacervation. IDPs come in a variety of flavors, and we study in this letter the archetypal liquid-liquid phase-separating IDP, a polyelectrolyte consisting of the same number of positively charged and negatively charged monomers. Electrostatic effects on the phase diagram emerge not just as a result of the net charge or charge density, but also on how the individual charged residues are arranged along the chain into charged "patches." Clustering of like charges into a local charge patch amplifies their effect, whereas scattering charges diminishes their effect due to screening from proximal oppositely charged residues.[11,24,41,42]

To study the role of charge-charge interactions in driving self-coacervation, we consider different permutations of positively and negatively charged monomers based on the sequences of 25 lysine (K) and 25 glutamic acid (E) residues introduced by Das and Pappu.[41] The five sequences considered exhibiting different patterning of E and K are shown in Fig. 1. Since FTS fully samples compositional fluctuations, we are able to efficiently compute both dilute and concentrated regimes and to present a complete phase diagram for the coacervation of model IDP sequences with different distributions of charged residues. We also introduce particle-based MD simulations to expose the individual-chain structures in the dilute states and to provide an additional consistency check of our FTS results. A detailed comparison of the structural correlations within the solution phase for the various EK sequences is presented in the Supporting Information.

Our system consists of n polymers with N monomers per chain in a volume V, with a resulting monomer number density $\rho = nN/V$. Each monomer has a site-specific charge in units of the electronic charge e such that each chain α has a charge pattern $\{\sigma_{\alpha,j}\} = \{\sigma_{\alpha,1}, \sigma_{\alpha,2}, ..., \sigma_{\alpha,N}\}$ where $\sigma_{\alpha,j}$ is the charge of the jth residue on chain α. Chain connectivity is enforced by a harmonic bond potential $\beta U_{\text{bond}} = \frac{3}{2b^2}\sum_{\alpha=1}^{n}\sum_{j=1}^{N}(|\mathbf{r}_{\alpha,j} - \mathbf{r}_{\alpha,j-1}|)^2$ where $\mathbf{r}_{\alpha,j}$ denotes the position of bead j on chain α, b is the statistical segment length, and $\beta = 1/k_BT$. The solvent is treated implicitly as a uniform dielectric continuum. All

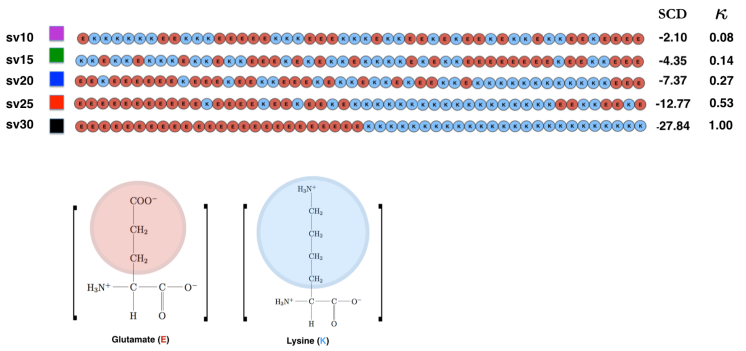

Figure 1: Five model IDP sequences of glutamate and lysine considered in this work. Sequences and nomenclature were originally introduced in Ref 41. Also shown are the sequence charge decoration (SCD) metric of Sawle and Ghosh,[43] and the κ parameter of Das and Pappu[41]

monomers interact through an excluded volume potential $\beta U_{\rm ex} = v\delta(\mathbf{r})$, where v is the excluded volume parameter, taken to be identical for all monomer types.[44,45] Electrostatic interactions are described by a Coulomb potential $\beta U_{\rm el} = \frac{l_B \sigma_i \sigma_j}{r}$ where $l_B = e^2/(4\pi\epsilon_0\epsilon_r k_B T)$ is the Bjerrum length, ϵ_0 is the vacuum permittivity, ϵ_r the relative dielectric strength of the solvent, and e is the unit of electronic charge. The total microscopic bead number density is $\hat{\rho}(\mathbf{r}) = \sum_\alpha^n \sum_j^N \delta(\mathbf{r} - \mathbf{r}_{\alpha,j})$, and similarly the total microscopic charge density is $\hat{\rho}_e(\mathbf{r}) = \sum_\alpha^n \sum_j^N \sigma_{\alpha,j}\delta(\mathbf{r} - \mathbf{r}_{\alpha,j})$. These microscopic densities are smeared so that both masses and charges have a Gaussian distribution.[20,46] The smeared densities corresponding to either $\hat{\rho}(\mathbf{r})$ or $\hat{\rho}_e(\mathbf{r})$ are

$$\bar{\rho}(\mathbf{r}) = \int d\mathbf{r}' \; \Gamma(|\mathbf{r} - \mathbf{r}'|)\hat{\rho}(\mathbf{r}') \tag{1}$$

with

$$\Gamma(r) = \left(\frac{1}{2\pi a^2}\right)^{3/2} \exp\left(-r^2/2a^2\right) \tag{2}$$

a normalized Gaussian with smearing width a. The resulting interaction energy for this model is

$$\beta\bar{U}[\mathbf{r}] = \frac{3}{2b^2}\sum_{\alpha=1}^{n}\sum_{j=1}^{N}(|\mathbf{r}_{\alpha,j} - \mathbf{r}_{\alpha,j-1}|)^2 + \frac{v}{2}\int d\mathbf{r} \;\; \bar{\rho}^2(\mathbf{r}) + \frac{l_B}{2}\int\int d\mathbf{r} \; d\mathbf{r}' \; \frac{\bar{\rho}_e(\mathbf{r})\bar{\rho}_e(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{3}$$

The essential features of the model are chain connectivity (first term), short-ranged excluded volume repulsion (second term), and long-ranged electrostatic interactions (third term). A

major effort of this work is to delineate how differences in the charge pattern $\{\sigma_{\alpha,j}\}$ affect the structural and thermodynamic properties of this model. To accomplish this, we take advantage of an exact transformation of this model into a statistical field theory.

Following the method described in Ref. 25, the canonical partition function for the model Hamiltonian given by Eqn. 3 can be expressed as a complex-valued statistical field theory,

$$\mathcal{Z}_c = \mathcal{Z}_0 \int Dw \int D\varphi \; e^{-H[w,\varphi]} \tag{4}$$

where $\mathcal{Z}_0$ is the partition function for an ideal gas of discrete Gaussian chains and self-interaction correction terms. An extension of the model to include small ions explicitly is described in the Supplemental Information. The field-theoretic Hamiltonian is

$$H[w,\varphi] = \frac{1}{2}\int d\mathbf{r}\left(\frac{[w(\mathbf{r})]^2}{v} + \frac{|\nabla\varphi(\mathbf{r})|^2}{4\pi l_B}\right) - n\ln Q[w,\varphi] \tag{5}$$

where w is a fluctuating chemical potential field and φ is a fluctuating electrostatic potential field, which serve to decouple pairwise interactions. $Q[w,\varphi]$ is the partition function for a single chain in a complex-valued external field, and can be computed using a Gaussian chain propagator

$$Q[w,\varphi] = \frac{1}{V}\int d\mathbf{r} \; q_N(\mathbf{r};[w,\varphi]) \tag{6}$$

The chain propagator $q_N(\mathbf{r};\psi)$ can be constructed from the following Chapman-Kolmogorov -type equation

$$q_{j+1}(\mathbf{r};\psi) = \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp[-\psi_{j+1}(\mathbf{r})]\int d\mathbf{r}' \;\; q_j(\mathbf{r}';\psi)\exp\left(-\frac{3|\mathbf{r}-\mathbf{r}'|^2}{2b^2}\right) \tag{7}$$

with $\psi_j = i\Gamma \star (w + \sigma_j\varphi)$, where $i = \sqrt{-1}$ and $\star$ denotes a spatial convolution. The initial condition is $q_0(\mathbf{r};[\psi]) = \exp[-\psi_0(\mathbf{r})]$.

As described elsewhere[20,46] field theoretic simulations are performed using complex Langevin (CL) sampling of the fields. After promoting the fields to be complex variables, the CL equa-

tions of motion are

$$\begin{aligned} \frac{\partial w(\mathbf{r},t)}{\partial t} &= -\lambda_w \frac{\delta H[w,\varphi]}{\delta w(\mathbf{r},t)} + \eta_w(\mathbf{r},t) \\ \frac{\partial \varphi(\mathbf{r},t)}{\partial t} &= -\lambda_\varphi \frac{\delta H[w,\varphi]}{\delta \varphi(\mathbf{r},t)} + \eta_\varphi(\mathbf{r},t) \end{aligned} \tag{8}$$

where $\eta_i(\mathbf{r},t)$ are real-valued Gaussian white noise random variables with zero mean and variance proportional to the dissipative coefficients λ_w and λ_φ. By propagating the CL equation of motion in time, FTS stochastically samples the configuration space of the conjugate field variables. Details of the CL-FTS procedure are presented in the Supporting Information.

For use in conventional molecular dynamics (MD) simulation of bead-spring chains, we present an equivalent particle-based representation for a system with potential energy given by Eqn 3. This amounts to expressing each of the interactions in Eqn 3 as effective pair potentials. The bond potential between beads separated by distance r is

$$\beta V_{bond}(r) = \frac{3}{2b^2} r^2. \tag{9}$$

The smeared Gaussian densities of Eqn 3 can be equivalently treated as point particles interacting through a soft potential, finite at contact. In this representation, the excluded volume interaction between monomers is equivalent to an effective soft Gaussian repulsive interaction of the form

$$\beta U_{ev}(r) = \frac{v}{8\pi^{3/2}a^3} e^{-r^2/4a^2}, \tag{10}$$

and the electrostatic potential is similarly[47,48]

$$\beta U_{el}(r) = \frac{l_B \sigma_i \sigma_j}{r} \mathrm{erf}\left(\frac{r}{2a}\right) \tag{11}$$

Particle simulations were carried out using the LAMMPS MD code[49] (http://lammps.sandia.gov). Details of the simulation procedure are presented in the Supporting Information. It is im-

portant to note that both MD and FTS are here performed on the same polymer model. While the FTS and MD simulations numerically sample the full Hamiltonian, it also proves instructive to consider the RPA approximation for this model, which provides an analytical approximation that we compare directly to FTS simulation and discuss the limits of its validity. Details of the derivation of the RPA expressions are given in the Supporting Information.

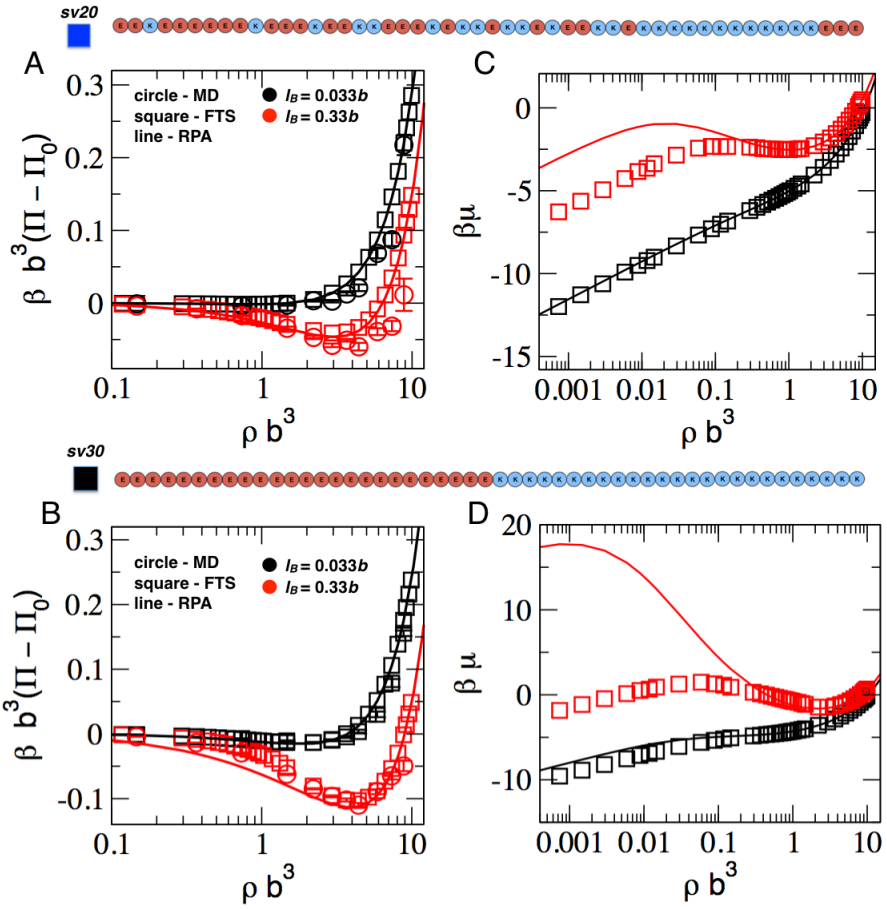

Figure 2: Left column: Excess osmotic pressure evaluated from particle MD (open circle), FTS (open square), and RPA solid line for model sequences $sv20$ (A) and $sv30$ (B) as a function of monomer density. Simulations were performed using a Bjerrum length of $l_B = 0.033b$ (black) and $l_B = 0.33b$ (red). Right column: The chemical potential as a function of monomer density for $sv20$ (C) and $sv30$ (D) with Bjerrum length $l_B = 0.033b$ (black) and $l_B = 0.33b$ (red). FTS results are represented with open squares and the solid line depicts the RPA expression.

The equilibrium conditions for the phase coexistence between a dense coacervate phase

and a dilute solution phase are the equivalence in osmotic pressure Π and chemical potential μ of the two phases. Before investigating the phase diagram, it is instructive to compare these thermodynamic observables computed from FTS, MD, and RPA. A plot of the osmotic pressure Π as a function of the monomer density at fixed temperature gives the numerical equation of state. Fig. 2 (left column) presents a comparison between FTS (square), RPA (line), and particle MD (circle) of the excess osmotic pressure for two representative chains $sv20$ (Fig. 2A) and $sv30$ (Fig. 2B). Results are presented for two values of the Bjerrum length l_B which parameterizes the strength of the electrostatic potential in Eqn. 3. Increasing l_B by an order of magnitude from a relatively weak value $l_B = 0.033b$ (black) to a more modest $l_B = 0.33b$ (red) has the effect of lowering the osmotic pressure curve due to the increased strength of favorable electrostatic attractions between oppositely charged monomers. This effect is more pronounced in the diblock $sv30$ sequence. For these parameter ranges the analytical RPA approximation is in agreement with FTS. Particle MD simulations become increasingly difficult to equilibrate at high polymer densities but agree with both FTS and RPA at low and intermediate densities.

Fig. 2 (right column) compares the chemical potential between the analytical RPA prediction (line) and the exact value from FTS (square). (Note that direct computation of the chemical potential with MD simulations is not shown as the calculation requires either thermodynamic integration or a Widom insertion method.) Importantly, while accurate for large monomer densities, the RPA approximation breaks down at low density. The difference between FTS and RPA becomes more pronounced at higher values of l_B. Fig. 2 demonstrates that particle MD simulations are suitable for low to intermediate densities whereas the RPA is valid at high densities. This observation motivates the use of FTS as a method to study phase behavior as it is capable of accurately describing the thermodynamics over the entire density range of interest. Although it is possible to use particle-based MD simulations to compute phase diagrams using the method described in Refs. 32 and 35, this requires equilibrating an MD simulation at each state point of interest, and is not pursued

here as it is computationally expensive. The purpose of this letter is to demonstrate the applicability of FTS to efficiently compute phase diagrams that accurately capture both the dilute and concentrated branches.

As discussed above, the Bjerrum length l_B controls the strength of the electrostatic potential. Increasing l_B leads to an increasing density fluctuation length scale implying an approaching critical point (See Supporting Information). The magnitute of the Bjerrum length in this work is chosen to expose the critical region of the phase diagram for these model peptides. For water at 300 K the Bjerrum length is typically on the order of b; however, the model peptides examined here have a much higher charge density than typical proteins and thus phase separate at lower values of the Bjerrum length.

Performing FTS in the Gibbs ensemble[50,51] is an efficient way to construct the exact binodal curve for all five of the sequences shown in Fig. 1 as a function of the reduced Bjerrum length l_B/b. Details of the Gibbs ensemble procedure are presented in the Supporting Information. The coexistence points determined from FTS in the Gibbs ensemble provide an efficient enumeration of the phase diagram for an intrinsically disordered protein described by Eqn. 3 without introducing additional approximations. At low density the dilute solution phase is favored. Increasing the protein concentration through ρb^3 leads to a two-phase region in which both the solution and coacervate phase are in coexistence (with concentrations shown by the data points). At sufficiently high protein concentration only a dense coacervate phase is stable. A snapshot of the density profile of each of these regions from FTS is presented in the top panel of Fig. 3A.

The exact phase diagram for this coarse-grained model, fully accounting for compositional fluctuations, for the five model sequences is shown in Fig. 3B and 3C and is compared with the RPA approximation. The Bjerrum length can be used to define a reduced temperature $T^* = b/l_B = 4\pi\epsilon_0\epsilon_r k_B T b/e^2$. Thus, Fig. 3B is analogous to an experimental inverse temperature vs. concentration plot. For comparison Fig. 3C shows the same data plotted as a reduced temperature vs. monomer density. Both RPA and FTS show that by sequestering

charges into larger "patches," the critical Bjerrum length for phase separation is lowered. Although the RPA model qualitatively describes the sequence dependent critical point, the low-density branch of the binodal is incorrect. The reason for the failure of the RPA at low concentration is largely due to the estimation of the chemical potential at low density as seen in Fig. 2. Nonetheless, the RPA prediction is qualitatively correct at modeling the sequence dependence and becomes quantitative at sufficiently high densities. This finding is in agreement with Lin and Chan who used the RPA model to study these same sequences in Ref. 24. We note that the linear density scale in Fig. 3C conceals this difference between FTS and RPA for the dilute branch.

In dilute polymer solutions, electrostatic effects can lead to single chain compaction or swelling.[52–54] At high enough polymer densities, it has been argued that these same electrostatic effects lead to inter-polymer associations, driving phase separation.[24,55–57] To investigate this, we performed particle MD simulations of single chains in implicit solvent. Fig 4A shows that larger charge "patches" of oppositely charged residues along the same chain result in a lower average radius of gyration R_g due to electrostatic attraction. The lower R_g is due to the chain sampling fewer extended configurations. The observed dependence of R_g on the charge patterning shown in Fig. 4A is in qualitative agreement with results from coarse-grained simulations presented by Das and Pappu;[41] however, those simulations used a hard repulsive potential instead of the soft Gaussian potential employed here, which allows the chains to become more compact in the latter case. Fig. 4B shows an example of the compact configuration for the diblock chain ($sv30$), and Fig. 4C shows a similar snapshot of two chains taken from a simulation of 300 chains in the early stages of IDP aggregation. These snapshots suggest that similar electrostatic effects control single chain compactness and multiple chain aggregation.

Fig. 4 implies that under dilute conditions the Gaussian form factors used in the RPA approximation are not physically realistic and may be partially responsible for the break-down of the RPA at dilute concentrations. Using the numerical form factor from short MD

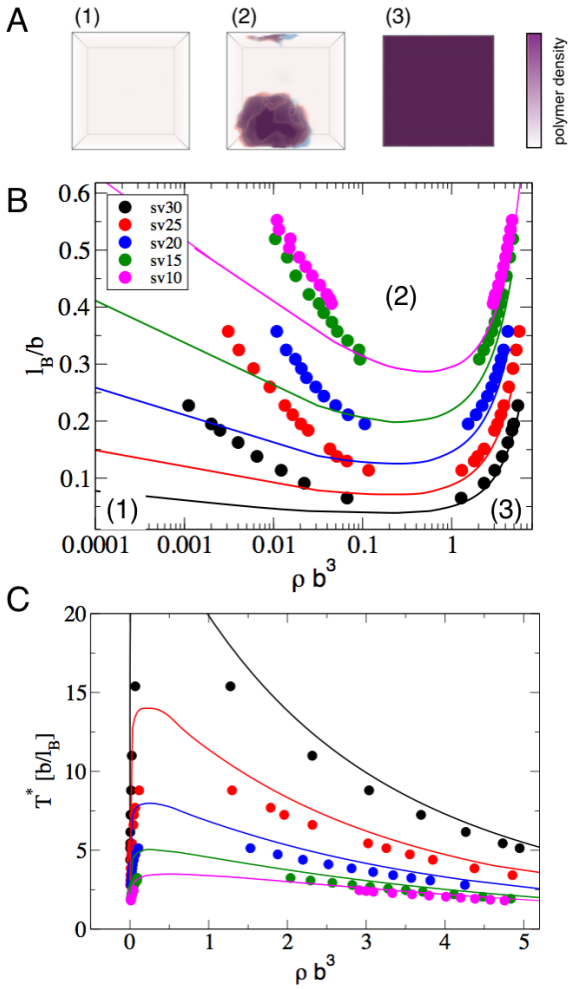

Figure 3: A: Snapshot of the polymer density profiles from FTS for three regions of the phase diagram: (1) a dilute solution phase; (2) a two-phase region in which a dilute supernatant and dense coacervate phase coexist; and (3) a dense protein region. B: Coexistence curves evaluated from FTS using the Gibbs ensemble (points) and computed from RPA (solid lines). The coloring scheme is the same as in Fig. 1 representing the five different chain sequences, showing a strong charge-sequence dependence on the two phase window. C: Coexistence curves showing the reduced temperature vs. monomer density. The linear scale of the horizontal axis highlights the features of the dense branch of the phase diagram but suppresses the dilute branch. All simulations were performed without explicit counter ions present.

simulations could potentially improve the RPA prediction. Such an approach would be similar in spirit to the recent Gaussian-renormalization model, which introduced a self-consistent single-chain structure to improve the RPA prediction for homopolyelectrolytes.[27] Thus far we have focused exclusively on varying the Bjerrum length l_B. In our model, two other factors

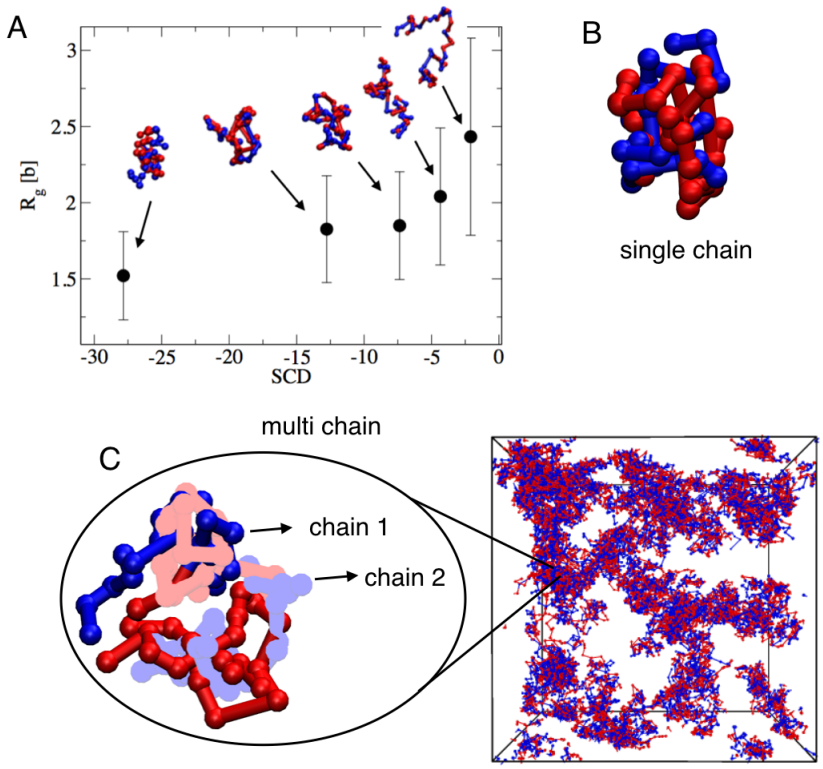

Figure 4: A: The average radius of gyration for each of the five sequences evaluated from a single chain particle MD simulation. The horizontal axis is the sequence charge decoration (SCD) metric of Sawle and Ghosh.[43] A representative chain configuration for each sequence is also shown. All single chain simulations we performed at $l_B = 0.33b$ without explicit counter ions present. B: Representative configuration from a single chain particle MD simulation showing intramolecular association between oppositely charged "patches" for the diblock chain $sv30$. C: Snapshot of two representative chains from a particle MD simulation of 300 chains showing early stages of aggregation due to intermolecular association between oppositely charged groups. Chain 1 is opaque and chain 2 is shaded for contrast.

control the phase behavior, namely the strength of the excluded volume parameterized by v in Eqn. 3 and the concentration of excess salt which is expected to screen electrostatic interactions. The excluded volume is a non-electrostatic interaction parameter that describes the solvent quality with a temperature dependence typically of the form $v = v_0(1 - \theta/T)$, where v_0 controls the strength of the interaction and θ is a protein-specific reference temperature. It should be noted that for largely hydrophobic sequences instead of the charged sequences considered in this work, the temperature dependence of v may be more complex. Fig. 5A shows the binodal coexistence points as a function of v for four of the five sequences at a fixed

value of $l_B = 0.33b$. Note that at this value of l_B the most charge-scrambled sequence $sv10$ is in the solution phase for all values of v and is not shown. It can be seen that increasing the excluded volume strength v, i.e. increasing the solvent quality, tends to stabilize the mixed phase and prevent phase separation, in agreement with the theoretical work of Perry and Sing[58] on model synthetic polyelectrolytes. This is due to an increase in the repulsive interaction between monomer segments that counteracts the electrostatic attraction between oppositely charged segments.

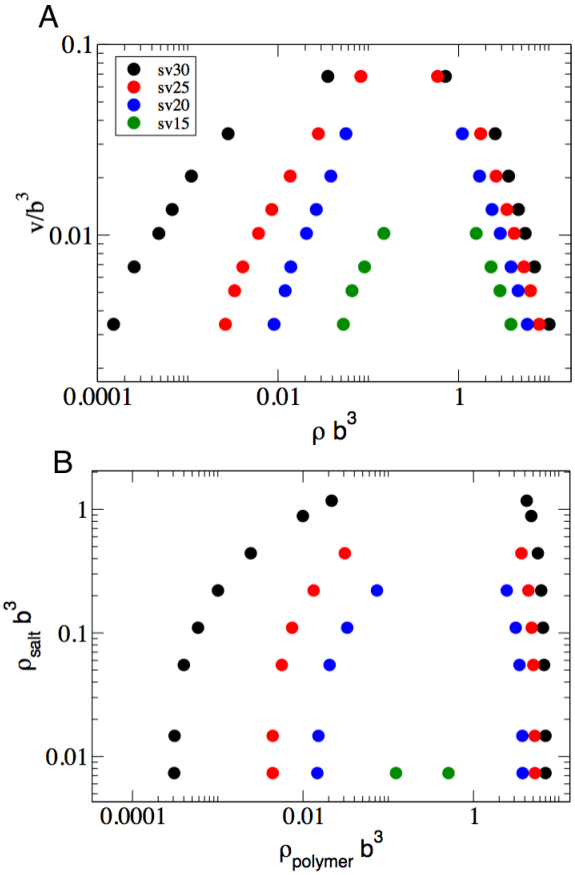

Figure 5: A: Coexistence points evaluated from FTS using the Gibbs ensemble at fixed $l_B = 0.33b$ with varying excluded volume strength. Simulations were performed without explicit counter ions present. The coloring scheme is the same as in Fig. 1 representing four different chain sequences. B: Coexistence points evaluated at fixed $v = 0.0069b^3$ and $l_B = 0.33b$ at varying salt concentration by introducing explicit salt ions as point particles into the polymer model.

Fig. 5B shows a similar suppression of the two-phase window by added salt. Explicit salt

ions are introduced as point charges with microscopic density smeared with a normalized Gaussian profile. The field theoretic Hamiltonian with explicit salts is presented in the Supporting Information. For simplicity the smearing length is the same for both salt and monomers; however, one could adopt different smearing lengths to model different salt ion sizes. The observed trend in Fig. 5B with salt is in qualitative agreement with experimental observations of simplified systems and real IDPs. A more detailed discussion of the effect of salt is the subject of a forthcoming paper.[59]

In this letter we have introduced a polyampholyte model to describe intrinsically disordered proteins and have demonstrated the applicability of FTS to study how the spatial organization of amino acids can mediate IDP phase behavior. The polyampholyte model we have considered represents the simplest coarse-grained model that still captures the relevant physical features of the problem. While in this work we have selectively focused on the clustering of charged amino acids, and so have considered a low-complexity sequence of just two amino acids (E and K), the model can be generalized to introduce additional non-electrostatic interactions such as hydrophobic or π-stacking interactions. Using FTS we construct an exact phase diagram for this model in terms of the dielectric strength, polymer excluded volume (solvent quality), and salt concentration. Our results show that clustering of charged amino acids into charged patches significantly modulates the phase diagram, in agreement with experiments performed on sequence variants of the germ-granule protein Ddx4.[60]

The direct comparison between particle MD simulation, FTS, and RPA theory for the same polyelectrolyte model allows us to comment on the limitations and merits of these different approaches. Because particle MD simulation provides structural information about single chain configurations, it can be useful to test predictions of intramolecular form factors and chain configurational distributions. It also offers site-specific resolution of chain associations as well as information about dynamics. However, MD simulation requires exceedingly long equilibration times for dense phase-separating systems and is thus limited to the regime

of relatively low polymer density. On the other hand, the analytical RPA theory is valid only at sufficiently high polymer densities, in agreement with previous observations.[20,38,39] This implies that one should be cautious about fitting the RPA model to the low-density branch of an experimental phase diagram. Despite its quantitative limitations compared to FTS, the RPA qualitatively captures the effect of charge sequence variation on the phase diagram, validating the recent application of the RPA model for these sequences.[21,24] By providing an exact numerical result, FTS is a useful technique against which RPA or other more advanced theories may be tested.

In summary, this letter presents a general framework from which to study liquid-liquid phase separation in biology. By considering a model system of low sequence-complexity peptides, we have demonstrated the ability of FTS to generate complete phase diagrams that depend on the primary amino acid sequence. The ability to model these effects from theory can be exploited to interpret experimental observation of the phase behavior of real IDPs or to design further experiments to understand the thermodynamic driving forces of coacervation in biological polymers.

Acknowledgement

This research is supported by the National Institutes of Heath (NIH) under grant number R01AG05605 and by the MRSEC Program of the National Science Foundation under Award No. DMR 1720256. The authors also acknowledge support from the National Science Foundation NSF under Award No. MCB-1716956. This research used resources of the Extreme Science and Engineering Discovery Environment (XSEDE, supported by the NSF Project TG-MCA05S027) and the Center for Scientific Computing from the California NanoSystems Institute UC Santa Barbara (CNSI) available through the Materials Research Laboratory (MRL): an NSF MRSEC (DMR-1720256) and NSF CNS-1725797.

References

(1) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology* **2005**, *6*, 197.

(2) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: structure, function, and bioinformatics* **2000**, *41*, 415–427.

(3) Forman-Kay, J. D.; Mittag, T. From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. *Structure* **2013**, *21*, 1492–1499.

(4) Riek, R.; Eisenberg, D. S. The activities of amyloids from a structural perspective. *Nature* **2016**, *539*, 227.

(5) Chiti, F.; Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.

(6) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **2009**, *324*, 1729–1732.

(7) Uversky, V. N.; Kuznetsova, I. M.; Turoverov, K. K.; Zaslavsky, B. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS letters* **2015**, *589*, 15–22.

(8) Patel, A.; Lee, H. O.; Jawerth, L.; Maharana, S.; Jahnel, M.; Hein, M. Y.; Stoynov, S.; Mahamid, J.; Saha, S.; Franzmann, T. M. et al. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **2015**, *162*, 1066–1077.

(9) Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **2015**, *163*, 123–133.

(10) Zhang, X.; Lin, Y.; Eschmann, N. A.; Zhou, H.; Rauch, J. N.; Hernandez, I.; Guzman, E.; Kosik, K. S.; Han, S. RNA stores tau reversibly in complex coacervates. *PLoS biology* **2017**, *15*, e2002183.

(11) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D. et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* **2015**, *57*, 936–947.

(12) Ambadipudi, S.; Biernat, J.; Riedel, D.; Mandelkow, E.; Zweckstetter, M. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature communications* **2017**, *8*, 275.

(13) Wegmann, S.; Eftekharzadeh, B.; Tepper, K.; Zoltowska, K. M.; Bennett, R. E.; Dujardin, S.; Laskowski, P. R.; MacKenzie, D.; Kamath, T.; Commins, C. et al. Tau protein liquid–liquid phase separation can initiate tau aggregation. *The EMBO journal* **2018**, e98049.

(14) Ferreon, J. C.; Jain, A.; Choi, K.-J.; Tsoi, P. S.; MacKenzie, K. R.; Jung, S. Y.; Ferreon, A. C. Acetylation Disfavors Tau Phase Separation. *International journal of molecular sciences* **2018**, *19*, 1360.

(15) Brangwynne, C. P.; Mitchison, T. J.; Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences* **2011**, *108*, 4334–4339.

(16) Itakura, A. K.; Futia, R. A.; Jarosz, D. F. It pays to be in phase. *Biochemistry* **2018**, *57*, 2520–2529.

(17) Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates. *Biochemistry* **2018**,

(18) Overbeek, J. T. G.; Voorn, M. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *Journal of Cellular Physiology* **1957**, *49*, 7–26.

(19) Veis, A.; Aranyi, C. Phase separation in polyelectrolyte systems. I. Complex coacervates of gelatin. *The Journal of Physical Chemistry* **1960**, *64*, 1203–1210.

(20) Delaney, K. T.; Fredrickson, G. H. Theory of polyelectrolyte complexation: Complex coacervates are self-coacervates. *The Journal of chemical physics* **2017**, *146*, 224902.

(21) Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Sequence-specific polyampholyte phase separation in membraneless organelles. *Physical review letters* **2016**, *117*, 178101.

(22) Lin, Y.-H.; Song, J.; Forman-Kay, J. D.; Chan, H. S. Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *Journal of Molecular Liquids* **2017**, *228*, 176–193.

(23) González-Mozuelos, P.; De La Cruz, M. O. Random phase approximation for complex charged systems: Application to copolyelectrolytes (polyampholytes). *The Journal of chemical physics* **1994**, *100*, 507–517.

(24) Lin, Y.-H.; Chan, H. S. Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophysical journal* **2017**, *112*, 2043–2046.

(25) Fredrickson, G. *The equilibrium theory of inhomogeneous polymers*; Oxford University Press on Demand, 2006; Vol. 134.

(26) Das, S.; Eisen, A.; Lin, Y.-H.; Chan, H. S. A lattice model of charge-pattern-dependent polyampholyte phase separation. *The Journal of Physical Chemistry B* **2018**, *122*.

(27) Shen, K.; Wang, Z.-G. Electrostatic correlations and the polyelectrolyte self energy. *The Journal of Chemical Physics* **2017**, *146*, 084901.

(28) Shen, K.; Wang, Z.-G. Polyelectrolyte Chain Structure and Solution Phase Behavior. *Macromolecules* **2018**, *51*, 1706–1717.

(29) Lytle, T. K.; Sing, C. E. Transfer matrix theory of polymer complex coacervation. *Soft matter* **2017**, *13*, 7001–7012.

(30) Radhakrishna, M.; Basu, K.; Liu, Y.; Shamsi, R.; Perry, S. L.; Sing, C. E. Molecular connectivity and correlation effects on polymer coacervation. *Macromolecules* **2017**, *50*, 3030–3037.

(31) Chang, L.-W.; Lytle, T. K.; Radhakrishna, M.; Madinya, J. J.; Vélez, J.; Sing, C. E.; Perry, S. L. Sequence and entropy-based control of complex coacervates. *Nature communications* **2017**, *8*, 1273.

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

(33) Blas, F. J.; MacDowell, L. G.; de Miguel, E.; Jackson, G. Vapor-liquid interfacial properties of fully flexible Lennard-Jones chains. *The Journal of chemical physics* **2008**, *129*, 144703.

(34) Silmore, K. S.; Howard, M. P.; Panagiotopoulos, A. Z. Vapour–liquid phase equilibrium and surface tension of fully flexible Lennard–Jones chains. *Molecular Physics* **2017**, *115*, 320–327.

(35) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS computational biology* **2018**, *14*, e1005941.

(36) Fredrickson, G. H.; Ganesan, V.; Drolet, F. Field-theoretic computer simulation methods for polymers and complex fluids. *Macromolecules* **2002**, *35*, 16–39.

(37) Ganesan, V.; Fredrickson, G. Field-theoretic polymer simulations. *EPL (Europhysics Letters)* **2001**, *55*, 814.

(38) Lee, J.; Popov, Y. O.; Fredrickson, G. H. Complex coacervation: A field theoretic simulation study of polyelectrolyte complexation. *The Journal of chemical physics* **2008**, *128*, 224908.

(39) Popov, Y. O.; Lee, J.; Fredrickson, G. H. Field-theoretic simulations of polyelectrolyte complexation. *Journal of Polymer Science Part B: Polymer Physics* **2007**, *45*, 3223–3230.

(40) Riggleman, R. A.; Kumar, R.; Fredrickson, G. H. Investigation of the interfacial tension of complex coacervates using field-theoretic simulations. *The Journal of chemical physics* **2012**, *136*, 024903.

(41) Das, R. K.; Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences* **2013**, *110*, 13392–13397.

(42) Danielsen P. O., S.; McCarty, J.; Shea, J.-E.; Delaney, K.; Fredrickson, G. H. Molecular Design of Sef-Coacervation Phenomena in Block Polyampholytes. *manuscript in preparation* **2018**,

(43) Sawle, L.; Ghosh, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *The Journal of chemical physics* **2015**, *143*, 08B615_1.

(44) Doi, M.; Edwards, S. F. *The theory of polymer dynamics*; oxford university press, 1988; Vol. 73.

(45) Firman, T.; Ghosh, K. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *The Journal of Chemical Physics* **2018**, *148*, 123305.

(46) Delaney, K. T.; Fredrickson, G. H. Recent developments in fully fluctuating field-

theoretic simulations of polymer melts and solutions. *The Journal of Physical Chemistry B* **2016**, *120*, 7615–7634.

(47) Coslovich, D.; Hansen, J.-P.; Kahl, G. Ultrasoft primitive model of polyionic solutions: Structure, aggregation, and dynamics. *The Journal of chemical physics* **2011**, *134*, 244514.

(48) Warren, P. B.; Vlasov, A.; Anton, L.; Masters, A. J. Screening properties of Gaussian electrolyte models, with application to dissipative particle dynamics. *The Journal of chemical physics* **2013**, *138*, 204907.

(49) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics* **1995**, *117*, 1–19.

(50) Riggleman, R. A.; Fredrickson, G. H. Field-theoretic simulations in the Gibbs ensemble. *The Journal of chemical physics* **2010**, *132*, 024104.

(51) Mester, Z.; Lynd, N. A.; Fredrickson, G. H. Numerical self-consistent field theory of multicomponent polymer blends in the Gibbs ensemble. *Soft Matter* **2013**, *9*, 11288–11294.

(52) Higgs, P. G.; Joanny, J.-F. Theory of polyampholyte solutions. *The Journal of chemical physics* **1991**, *94*, 1543–1554.

(53) Ha, B.-Y.; Thirumalai, D. Conformations of a polyelectrolyte chain. *Physical Review A* **1992**, *46*, R3012.

(54) Samanta, H. S.; Chakraborty, D.; Thirumalai, D. Charge fluctuation effects on the shape of flexible polyampholytes with applications to intrinsically disordered proteins. *The Journal of Chemical Physics* **2018**, *149*, 163323.

(55) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. A polymer physics perspective on

driving forces and mechanisms for protein aggregation. *Archives of biochemistry and biophysics* **2008**, *469*, 132–141.

(56) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2010**, *107*, 14609–14614.

(57) Dignon, G. L.; Zheng, W.; Best, R. B.; Kim, Y. C.; Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2018**, *115*, 9929–9934.

(58) Perry, S. L.; Sing, C. E. Prism-based theory of complex coacervation: Excluded volume versus chain correlation. *Macromolecules* **2015**, *48*, 5040–5053.

(59) Danielsen P. O, S.; McCarty, J.; Shea, J.-E.; Delaney, K.; Fredrickson, G. H. Small Ion Effects on Self-Coacervation Phenomena in Block Polyampholytes. *manuscript in preparation* **2018**,

(60) Brady, J. P.; Farber, P. J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T. J.; Chan, H. S.; Baldwin, A. J.; Forman-Kay, J. D. et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proceedings of the National Academy of Sciences* **2017**, *114*, E8194–E8203.

Graphical TOC Entry