A unified analytical theory of heteropolymers for sequence-specific phase behaviors of polyelectrolytes and polyampholytes

View Online   Export Citation   CrossMark

Yi-Hsuan Lin,[1,2] Jacob P. Brady,[1,3,4] Hue Sun Chan,[1,a] (iD) and Kingshuk Ghosh[5,6,b] (iD)

AFFILIATIONS

[1] Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[2] Molecular Medicine, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada
[3] Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[4] Department of Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada
[5] Department of Physics and Astronomy, University of Denver, Colorado, Colorado 80208, USA
[6] Molecular and Cellular Biophysics, University of Denver, Colorado, Colorado 80208, USA

[a] Electronic mail: chan@arrhenius.med.utoronto.ca
[b] Electronic mail: kingshuk.ghosh@du.edu

ABSTRACT

The physical chemistry of liquid-liquid phase separation (LLPS) of polymer solutions bears directly on the assembly of biologically functional dropletlike bodies from proteins and nucleic acids. These biomolecular condensates include certain extracellular materials and intracellular compartments that are characterized as "membraneless organelles." Analytical theories are a valuable, computationally efficient tool for addressing general principles. LLPS of neutral homopolymers is quite well described by theory, but it has been a challenge to develop general theories for the LLPS of heteropolymers involving charge-charge interactions. Here, we present a theory that combines a random-phase-approximation treatment of polymer density fluctuations and an account of intrachain conformational heterogeneity based on renormalized Kuhn lengths to provide predictions of LLPS properties as a function of pH, salt, and charge patterning along the chain sequence. Advancing beyond more limited analytical approaches, our LLPS theory is applicable to a wide variety of charged sequences ranging from highly charged polyelectrolytes to neutral or nearly neutral polyampholytes. This theory should be useful in high-throughput screening of protein and other sequences for their LLPS propensities and can serve as a basis for more comprehensive theories that incorporate nonelectrostatic interactions. Experimental ramifications of our theory are discussed.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5139661

I. INTRODUCTION

Mesoscopic compartmentalization undergirded by liquid-liquid phase separation (LLPS) of intrinsically disordered proteins or regions (IDPs or IDRs) and nucleic acids is now recognized as a versatile means for biomolecular organization and regulation.[1–6] Some of these phase-separated dropletlike compartments are intracellular bodies—such as stress granules, P-granules, and nucleoli—that may be characterized as "membraneless organelles." Outside the cell, biomolecular LLPS can be biologically useful as well, as in the formation of certain extracellular materials. Collectively referred to as biomolecular condensates, these phase-separated bodies participate in many vital functions as highlighted by their recently elucidated roles in endocytosis,[7] silencing chromatin,[8] transcription,[9–11] and translation.[12] The repertoire of relevant discoveries is rapidly expanding.[13–15] LLPS of globular proteins, for example, lens protein solutions, has also been observed and is of biological importance.[16–21]

Recent bioinformatics analyses suggest that IDPs and IDRs comprise a significant fraction of the proteomes of higher organisms and that functional LLPS is likely ubiquitous.[22] The propensity for an IDP or IDR to phase separate is governed by its amino acid sequence

and modulated by solution/environmental conditions (temperature, hydrostatic pressure,[23] pH, ionic strength,[24,25] etc) as well as their interactions with other biopolymers such as RNA. Thus, any "big-picture" survey of the physical basis of biomolecular condensates requires not only consideration of many different sequences but also a large variety of environmental conditions. Adding to this combinatorial complexity is that even for a given wildtype sequence, posttranslational modifications, mutations, and splicing[4,26] can lead to diverse LLPS propensities. In this context, analytical theories are the most computationally efficient tool for large-scale exploration of sequence-dependent biomolecular LLPS. Although explicit-chain simulations provide more energetic and structural details[27–29] and field-theory simulations afford more numerical accuracy,[30–32] currently the number of sequences that can be simulated by these approaches is limited because of their high computational cost. Moreover, analytical theories are valuable for insights into physical principles that are less manifest in simulation studies. With this in mind, we build on recent success in using analytical theories to account for sequence-dependent biomolecular condensates under certain limited conditions[33,34] so as to develop improved theories that are more generally applicable.

Building sequence-specific theories of LLPS will also have implications in phase separation of block polyampholytes and its comparison with complex coacervation between oppositely charged homopolyelectrolytes, a topic of intense research in polymer physics.[35–49] Diblock polyampholytes with repeat units of a polycation segment followed by a polyanion segment can be envisioned to be equivalent to two oppositely charged homopolyelectrolytes. For this reason, LLPS of block polyampholytes—a limiting case of our theory—is often termed self-coacervation[31,32] and shares features similar to complex coacervation of a polycation and polyanion.[49] Experiments and simulation have also reported differences between the phase diagrams of block polyampholytes and homopolyelectrolyte coacervation. The observed differences can be explained by the presence of "charge pattern interfaces" where two segments of oppositely charged blocks merge in polyampholytes. Homopolyelectrolytes, on the other hand, lack such connectivities, thus leading to different types of salt localization in comparison with block polyampholytes.[49] Application of a general sequence-based analytical theory of polyampholyte LLPS will further advance these comparisons between complex coacervation and self-coacervation. Future effort in theory development is needed in this direction. Thus, our framework should be useful not only for high-throughput analyses of the LLPS propensities of naturally occurring biological sequences but also for the design of artificial biological and nonbiological heteropolymers with desired LLPS properties.[50–52]

Inasmuch as sequence-specific analytical theories for biomolecular condensates are concerned, a recent multiple-chain formulation based on the traditional random phase approximation (RPA)[53,54] has been applied to study the dependence of LLPS of IDPs on the charge patterns along their chain sequences.[55] This approach accounts for the experimental difference in LLPS propensity between the Ddx4 helicase IDR and its charge-scrambled mutant.[55,56] It also provides insight into a possible anticorrelation between multiple-chain LLPS propensity and single-chain conformational dimensions[57] as well as the degree of demixing of different charge sequences under LLPS conditions.[58] As an initial step, these advances are useful. As a heteropolymer theory, however, traditional RPA[53,54] is known to

have two main shortcomings. First, the density of monomers of the polymer chains in solution is assumed to be roughly homogeneous as density fluctuations are neglected beyond second order in RPA. A rigorous treatment proposed by Edwards and Muthukumar has shown the importance of including density fluctuations to higher orders.[59–61] Nonetheless, a recent comparison of field-theory simulation and RPA indicates that RPA is reasonably accurate for intermediate to high monomer densities for the cases considered and that significant deviations between RPA and field theory simulation occur only for volume fraction <0.02 that of the highest condensed-phase simulated.[30] Second, traditional RPA neglects the fact that monomer-monomer interactions can cause conformational variation of individual chains by computing the single-chain structure factor using a Gaussian chain with no intrachain interaction. This limitation, which applies to homopolymers as well as heteropolymers, is particularly acute for the latter. Indeed, experimental and computational studies have shown that single-chain conformational heterogeneities and dimensions are sensitive to sequence specific interactions.[62–67] Regarding this shortcoming, recently an improved analytical approach was developed at the single-chain level by replacing the Kuhn length l (termed "bare" Kuhn length) of the Gaussian chain by a set of renormalized Kuhn lengths, l_1, that embodies the sequence-specific interactions approximately.[68–70] Renormalized structure factors have also been exploited to improve homopolymer LLPS theories for polyelectrolytes.[71,72]

Noting that the first shortcoming described above is likely limited only to regimes of extremely low polymer concentrations, here we first focus on rectifying the second shortcoming by combining the earlier, traditional sequence-dependent RPA theory[55,56] with the sequence-dependent single-chain theory that utilizes a renormalized Gaussian (rG) chain formulation[68–70] for a better account of conformational heterogeneity. We refer to this theory as rG-RPA. As a control, we also study a simpler theory, analogous to our earlier formulation,[55,56] that invokes a Gaussian chain with a fixed Kuhn length. Following Shen and Wang,[72] we refer to this $l_1 = l$ theory as fG-RPA. Extensive comparisons of rG-RPA and fG-RPA predictions on various systems indicate that rG-RPA represents a significant improvement over fG-RPA. As will be detailed below, the superiority of rG-RPA is most notable in its ability to account for the LLPSs of both polyampholytes and polyelectrolytes, whereas fG-RPA is inadequate for polyelectrolytic polymers.

II. THEORY

We consider an overall neutral solution of n_p charged polymers, each consisting of N monomers (residues), and small ions including n_s salt ions and n_c counterions with charge numbers z_s and z_c, respectively. The charge pattern of a polymer is given by an N-dimensional vector $|\sigma\rangle = [\sigma_1, \sigma_2, \ldots, \sigma_N]^T$, where σ_τ is the charge on the τth monomer, and $q_c \equiv (\sum_\tau \sigma_\tau)/N$ is the net charge per monomer. For simplicity, we consider the case with only one species of positive and one species of negative ions; their numbers are denoted as n_+ and n_-, respectively. Moreover, "salt" is identified as the small ions that carry charges of the same sign as the polymers, whereas "counterions" are the small ions carrying charges opposite to that of the polymers. Thus, $n_s = n_+$ if $q_c > 0$ and $n_s = n_-$ if $q_c < 0$, and $|q_c|n_pN + z_sn_s = z_cn_c$ for solution neutrality. The densities (ρ) of monomers, salt ions, and counterions are, respectively, $\rho_m = n_pN/\Omega$, $\rho_s = n_s/\Omega$,

and $\rho_c = n_c/\Omega$, where Ω is the solution volume. Although only a simple system with at most two species of small ions is analyzed here for conceptual clarity, our theory can be readily expanded to account for multiple species of small ions.

Details of our formulation are given in Appendixes A and B. Here, we provide the key steps in the derivation. Let F be the total free energy of the system. Then, $f \equiv Fl^3/(k_B T\Omega)$ is free energy in units of $k_B T$ per volume l^3, where l is the bare Kuhn length, k_B is the Boltzmann constant, and T is the absolute temperature. In our theory,

$$f = -s + f_{ion} + f_p + f_0, \tag{1}$$

where s is the mixing entropy, f_{ion} and f_P are the interactions among the small ions and involving the polymers, respectively, that arise from density fluctuations, and f_0 is the mean-field excluded volume interaction, all expressed in the same unit as f. The mixing entropy, which accounts for the configurational freedom of the solutes, takes the Flory-Huggins form, viz.,

$$-s = \frac{\phi_m}{N}\ln\phi_m + \phi_s\ln\phi_s + \phi_c\ln\phi_c + \phi_w\ln\phi_w, \tag{2}$$

where ϕ_m, ϕ_s, ϕ_c, and $\phi_w = 1 - \phi_m - \phi_s - \phi_c$ are volume fractions ($\phi = \rho l^3$), respectively, of polymers, salt ions, counterions, and solvent (water for IDP systems). Following Muthukumar, the charge of each small ion is taken to be distributed over a finite volume comparable to that of a monomer. The corresponding interaction free energy among the small ions is[73]

$$f_{ion} = -\frac{1}{4\pi}\left[\ln(1+\kappa l) - \kappa l + \frac{1}{2}(\kappa l)^2\right], \tag{3}$$

where $1/\kappa = 1/\sqrt{4\pi l_B(z_s^2\rho_s + z_c^2\rho_c)}$ is the Debye screening length and l_B is the Bejurrm length. Polymers interact via a κ-dependent screened Coulomb potential and a uniform excluded-volume repulsion with strength v_2. The origin of this repulsive term is to be understood as an effective interaction involving both the polymer and solvent. By setting v_2 repulsive, we imply that the polymer is in a good solvent. These interactions are contained in the expression

$$\mathscr{U}_p[\mathbf{R}] = \frac{1}{2}\sum_{\alpha,\beta=1}^{n_p}\sum_{\tau,\mu=1}^{N}\left[\frac{\sigma_\tau\sigma_\mu e^{-\kappa|\mathbf{R}_{\alpha,\tau}-\mathbf{R}_{\beta,\mu}|}}{|\mathbf{R}_{\alpha,\tau}-\mathbf{R}_{\beta,\mu}|} + v_2\delta^3(\mathbf{R}_{\alpha,\tau}-\mathbf{R}_{\beta,\mu})\right], \tag{4}$$

where $\mathbf{R}_{\alpha,\tau}$ is the position of the τth monomer in the αth polymer. The $\mathscr{U}_p$ form facilitates the formulation in terms of density fields below. For this purpose, the divergent self-interaction terms in $\mathscr{U}_p$ are either regularized subsequently or inconsequential because they do not contribute to phase-separation properties. Chain connectivity of the polymers is enforced by the potential

$$\mathscr{T}[\mathbf{R}] = \frac{3}{2l^2}\sum_{\alpha=1}^{n_p}\sum_{\tau=1}^{N-1}(\mathbf{R}_{\alpha,\tau+1}-\mathbf{R}_{\alpha,\tau})^2. \tag{5}$$

Thus, aside from a combinatorial factor that has already been included in Eq. (2), the partition function involving the polymers is given by

$$\mathscr{Z}_p = \int\prod_{\alpha=1}^{n_p}\prod_{\tau=1}^{N}d\mathbf{R}_{\alpha,\tau}e^{-\mathscr{T}[\mathbf{R}]-\mathscr{U}_p[\mathbf{R}]}. \tag{6}$$

Now, by applying the Hubbard-Stratonovich transformation and converting real-space to $\mathbf{k}$-space variables, we convert the coordinate-space partition function in Eq. (6) to a $\mathbf{k}$-space partition function[30,31] involving a charge-density field ψ and a matter-density field w, viz.,

$$\mathscr{Z}_p = \mathscr{Z}_0\mathscr{Z}'_p, \quad \mathscr{Z}'_p = \int\prod_{\mathbf{k}\neq\mathbf{0}}\sqrt{\frac{v_k}{v_2}}\frac{d\psi_\mathbf{k}dw_\mathbf{k}}{2\pi\Omega}e^{-\mathscr{H}[\psi,w]}, \tag{7}$$

where $\mathscr{Z}_0 = \exp[-v_2(Nn_p)^2/2\Omega]$ is the factor for $\mathbf{k} = \mathbf{0}$,

$$\mathscr{H}[\psi,w] = \frac{1}{2\Omega}\sum_{\mathbf{k}\neq\mathbf{0}}\left[v_k\psi_{-\mathbf{k}}\psi_\mathbf{k} + \frac{w_{-\mathbf{k}}w_\mathbf{k}}{v_2}\right] - n_p\ln\mathcal{Q}_p[\psi,w], \tag{8}$$

with $v_k \equiv k^2/(4\pi l_B) + (z_s^2\rho_s + z_c^2\rho_c)$, the scalar $k \equiv |\mathbf{k}|$, $\mathcal{Q}_p[\psi,w] = \int\mathscr{D}[\mathbf{R}]\exp(-\mathcal{H}_p[\psi,w])$ is the single-polymer partition function with $\mathscr{D}[\mathbf{R}] \equiv \prod_{\tau=1}^{N}d\mathbf{R}_\tau$ (the chain label α in $\mathbf{R}$ is dropped since the integration here is only over one chain), and

$$\mathcal{H}_p[\psi,w] = \frac{3}{2l^2}\sum_{\tau=1}^{N-1}(\mathbf{R}_{\tau+1}-\mathbf{R}_\tau)^2 + \frac{i}{\Omega}\sum_{\mathbf{k}\neq\mathbf{0}}\sum_{\tau=1}^{N}(\sigma_\tau\psi_\mathbf{k}+w_\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{R}_\tau}. \tag{9}$$

The total interaction free energy involving the polymers in the unit of Eq. (1) is $-(l^3/\Omega)\ln\mathscr{Z}_p$, which we express as the sum of a density-fluctuation contribution $f_P = -(l^3/\Omega)\ln\mathscr{Z}'_p$ and a mean-field contribution $f_0 = -(l^3/\Omega)\ln\mathscr{Z}_0 = \frac{1}{2}v_2\rho_m^2$. The f_0 term involves neither small ions nor electrostatic interactions because the excluded volumes of the small ions are not considered beyond the incompressibility condition in Eq. (2) and the solution system as a whole is neutral.

We evaluate $\mathscr{Z}'_p$ in Eq. (7) perturbatively by expanding $\mathscr{H}[\psi,w]$ to the second order in density,

$$\mathscr{H}[\psi,w] \approx \frac{1}{2\Omega}\sum_{\mathbf{k}\neq\mathbf{0}}\langle\psi_{-\mathbf{k}}\ w_{-\mathbf{k}}|\begin{pmatrix} v_k + \rho_m\xi_\mathbf{k} & \rho_m\zeta_\mathbf{k} \\ \rho_m\zeta_\mathbf{k} & v_2^{-1} + \rho_m g_\mathbf{k} \end{pmatrix}|\begin{matrix}\psi_\mathbf{k}\\ w_\mathbf{k}\end{matrix}\rangle, \tag{10}$$

where $g_\mathbf{k}$, $\xi_\mathbf{k}$, and $\zeta_\mathbf{k}$ are monomer density-monomer density, charge-charge, and monomer density-charge correlation functions in $\mathbf{k}$-space, and $\langle\cdots|$ and $|\cdots\rangle$ are, respectively, row and column vectors. $\mathscr{Z}'_p$ can then be calculated as a Gaussian integral to yield

$$f_P = -\frac{l^3\ln\mathscr{Z}'_p}{\Omega}$$
$$= \frac{l^3}{2}\int\frac{d^3k}{(2\pi)^3}\ln\left[1 + \rho_m\left(\frac{\xi_\mathbf{k}}{v_k} + v_2 g_\mathbf{k}\right) + \frac{v_2}{v_k}\rho_m^2(\xi_\mathbf{k}g_\mathbf{k} - \zeta_\mathbf{k}^2)\right]. \tag{11}$$

Evaluation of $g_\mathbf{k}$, $\xi_\mathbf{k}$, and $\zeta_\mathbf{k}$ requires knowledge of the single-polymer $\mathcal{Q}_p$ [Eq. (8)], which, in general, depends on the sequence charge pattern. fG-RPA makes the simplifying assumption that $\mathcal{Q}_p$ is that of Gaussian chains with a fixed l, i.e., assume that the second term in Eq. (9) vanishes. As introduced above, here we use a renormalized Kuhn length $l_1 = xl$ to better account for the effects of interactions on $\mathcal{Q}_p$ by making the improved approximation

$$\mathcal{Q}_p \approx \int \mathscr{D}[\mathbf{R}] e^{-\mathcal{H}_p^0} \ ; \ \mathcal{H}_p^0 = \frac{3}{2l^2 x} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\tau+1} - \mathbf{R}_\tau)^2. \qquad (12)$$

Accordingly, the correlation functions in Eq. (11) are computed using l_1 instead of l,

$$g_{\mathbf{k}} \to g_k^x = \frac{1}{N}\langle 1|\hat{G}_k^x|1\rangle, \quad \xi_{\mathbf{k}} \to \xi_k^x = \frac{1}{N}\langle \sigma|\hat{G}_k^x|\sigma\rangle,$$
$$\zeta_{\mathbf{k}} \to \zeta_k^x = \frac{1}{N}\langle \sigma|\hat{G}_k^x|1\rangle, \qquad (13)$$

where $\hat{G}_k^x$ is the $N \times N$ correlation matrix of the renormalized Gaussian chain with $[\hat{G}_k^x]_{\tau\mu} = \exp[-(kl)^2 x|\tau - \mu|/6]$ and $\langle 1|$ and $|1\rangle$ are N-dimensional vectors with all elements equal to 1.

The single x variable here for the end-to-end distance serves to provide an approximate account of sequence specific effects in single-chain conformations. A more accurate formalism that may be pursued in the future is to consider x as a function of specific residue pairs, i.e., $x \to x(\tau, \mu)$, so as to provide a structure factor that applies to all length scales as in the approach of Shen and Wang.[71]

A variational approach similar to that in the work of Sawle and Ghosh[68] is applied to obtain a sequence-specific x by first expressing $\mathcal{H}_p$ in Eq. (9) as $\mathcal{H}_p = \mathcal{H}_p^0 + \mathcal{H}_p^1$, where $\mathcal{H}_p^0$ is given by Eq. (12) and $\mathcal{H}_p^1$ is the discrepancy in using the renormalized $\mathcal{H}_p^0$ to approximate $\mathcal{H}_p$. In general, a partially optimized solution for x may be obtained by minimizing the differences in averaged physical quantities computed using $\mathcal{H}_p$ vs those computed using $\mathcal{H}_p^0$, i.e., minimizing contributions from $\mathcal{H}_p^1$. As emphasized above, to simplify this calculation, we use, as in Ref. 68, the polymer squared end-to-end distance $|\mathbf{R}_N - \mathbf{R}_1|^2$ as the physical quantity for the partial optimization of x. The derivation proceeds largely as before,[68] except the monomer-monomer interaction potential in Ref. 68 is now replaced by the effective field-field correlation functions[59]

$$U_{\text{eff}}(\mathbf{k}) \equiv \sum_{\tau,\mu=1}^{N} \left[\sigma_\tau \sigma_\mu \langle \psi_{-\mathbf{k}}\psi_{\mathbf{k}}\rangle + \langle w_{-\mathbf{k}}w_{\mathbf{k}}\rangle + (\sigma_\tau + \sigma_\mu)\langle \psi_{-\mathbf{k}}w_{\mathbf{k}}\rangle \right], \qquad (14)$$

where $\langle \cdots \rangle$ represents averaging over field configurations. This analysis, the details of which are given in Appendixes A and B, leads to an equation that allows us to determine x,

$$1 - \frac{1}{x} - \frac{Nl^2}{18(N-1)} \int \frac{d^3k}{(2\pi)^3} \frac{k^2 \Xi_k^x}{\det \Delta_k^x} = 0, \qquad (15)$$

where Δ_k^x is the 2×2 matrix in Eq. (10) with $g_{\mathbf{k}}$, $\xi_{\mathbf{k}}$, and $\zeta_{\mathbf{k}}$ replaced by their renormalized g_k^x, ξ_k^x, and ζ_k^x in Eq. (13). In the numerator of the integrand in Eq. (15),

$$\Xi_k^x = \frac{\bar{\xi}_k^x}{v_2} + v_k \bar{g}_k^x + \rho_m \left(\bar{\xi}_k^x g_k^x + \xi_k^x \bar{g}_k^x - 2\zeta_k^x \bar{\zeta}_k^x \right), \qquad (16)$$

where

$$\bar{\xi}_k^x = \frac{1}{N}\langle \sigma|\hat{L}_2 \hat{G}_k^x|\sigma\rangle, \quad \bar{g}_k^x = \frac{1}{N}\langle 1|\hat{L}_2 \hat{G}_k^x|1\rangle, \quad \bar{\zeta}_k^x = \frac{1}{N}\langle \sigma|\hat{L}_2 \hat{G}_k^x|1\rangle, \quad (17)$$

with $\hat{L}_2$ being an $N \times N$ matrix with $[\hat{L}_2]_{\tau\mu} = |\tau - \mu|^2$. Now, for any chosen excluded-volume parameter v_2, x can be solved as the only

unknown in Eq. (15). With x determined, f_p can be computed via Eq. (11) and combined with the above expressions for s, f_{ion}, and f_0 to complete the free energy function in Eq. (1) for our rG-RPA theory. Here, we use $v_2 = 4\pi l^3/3$, which is about the $\sim l^3$ size of a monomer, in the applications below.

We note that while $v_2 > 0$ (which disfavors collapsed conformations) is required in the present formulation to solve for an effective Kuhn length, the general trend predicted by our theory is not affected by reasonable variation around the $v_2 = 4\pi l^3/3$ value.

III. RESULTS

A. Salt-free rG-RPA unifies established LLPS trends of both uniformly charged polyelectrolytes and neutral polyampholytes

We first illustrate the more general applicability of rG-RPA by comparing rG-RPA and fG-RPA predictions for salt-free solutions of uniformly charged polyelectrolytes (fully charged homopolymers) and 4-block overall neutral polyampholytes of several different chain lengths (Fig. 1). As stated above, fG-RPA corresponds to setting $x = l_1/l = 1$ and $v_2 = 0$ in rG-RPA. While fG-RPA is not identical to our earlier RPA[55] because fG-RPA subsumes the effects of small ions in a screening potential for the polymers whereas our earlier RPA theory treats the small ions and polymers on the same footing, both theories share the Gaussian-chain approximation and their predicted trends are very similar, as will be illustrated by examples below.

The rG-RPA-predicted critical point $((\phi_m)_{\text{cr}}, 1/(l_B)_{\text{cr}})$ in Fig. 1(a) for polyelectrolytes is insensitive to chain length $[(l_B)_{\text{cr}}$ is the critical Bjerrum length; $1/(l_B)_{\text{cr}}$ is proportional to the
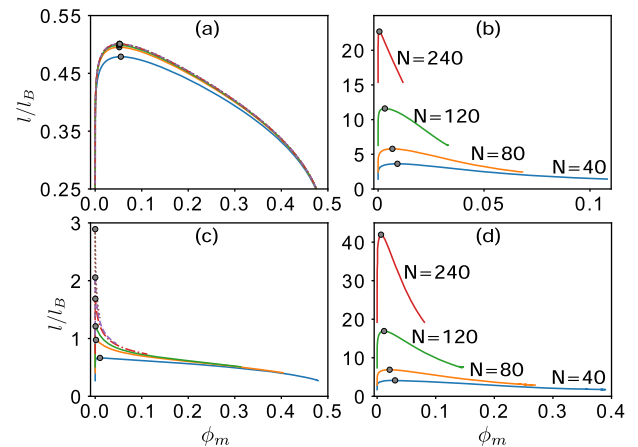

FIG. 1. Salt-free LLPS of polyelectrolytes and polyampholytes. rG-RPA [(a) and (b), top panels] and fG-RPA [(c) and (d), bottom panels] phase diagrams for N = 10, 25, 40, 80, 120, and 240 polyelectrolytes with charge sequences $\sigma_\tau = -1$ for $\tau = 1, 2, \ldots, N$ [(a) and (c), left panels] and N = 40, 80, 120, and 240 4-block polyampholytes with charge sequences $\sigma_\tau = +1$ for $\tau = 1, 2, \ldots, N/4$ and $\tau = N/2 + 1, N/2 + 2, \ldots, 3N/4$, and $\sigma_\tau = -1$ for $\tau = N/4 + 1, N/4 + 2, \ldots, N/2$ and $\tau = 3N/4 + 1, 3N/4 + 2, \ldots, N$ [(b) and (d), right panels]. Gray circles are critical points. For the coexistence curves in [(a) and (c)], N decreases from top to bottom, with the N = 80, 120, and 240 curves in (a) being nearly identical.

critical temperature T_{cr}]. As N increases, $\lim_{N\to\infty} 1/(l_B)_{\text{cr}} \approx 0.5$ and $\lim_{N\to\infty}(\phi_m)_{\text{cr}} \approx 0.05$. These predictions are consistent with lattice-chain simulations[74] and other theories.[71,73,75,76] The fG-RPA predictions are drastically different, viz., $\lim_{N\to\infty} 1/(l_B)_{\text{cr}} \to \infty$ and $\lim_{N\to\infty}(\phi_m)_{\text{cr}} \to 0$ [Fig. 1(c)]. Thus, fG-RPA is limited as earlier RPA theories[53,54] and its predictions for polyelectrolytes are inconsistent with the aforementioned established results.[71,73–76] This comparison between rG-RPA and fG-RPA underscores the importance of appropriately accounting for conformational heterogeneity in understanding polyelectrolyte LLPS and the effectiveness of using renormalized Kuhn lengths for the purpose.

Both rG-RPA and fG-RPA predict $1/(l_B)_{\text{cr}} \to \infty$ and $(\phi_m)_{\text{cr}} \to 0$ as $N \to \infty$ for the polyampholytes [Figs. 1(b) and 1(d)]. These results are consistent with simple RPA theory,[55,56] a charged hard-sphere chain model,[77] and lattice-chain simulations.[78] Not surprisingly, both rG-RPA and fG-RPA posit that T_{cr}'s of polyelectrolytes are much lower than those of neutral polyampholytes because direct electrostatic attractions exist for polyampholytes but effective attractions among polyelectrolytes can only be mediated by counterions.

For the polyampholytes, rG-RPA [Fig. 1(b)] predicts lower T_{cr}'s than fG-RPA [Fig. 1(d)]. With a more accurate treatment of single-chain conformational dimensions, rG-RPA should entail more compact isolated single-chain conformations for block polyampholytes, resulting in less accessibility of the charges for interchain cohesive interactions and therefore a weaker—but physically more accurate—LLPS propensity.

Notably, the fG-RPA-predicted phase boundaries of both polyelectrolytes and polyampholytes exhibit an inverse S-shape phase boundaries [the condensed-phase part of the coexistence curves concave upward; see Figs. 1(c) and 1(d)]. In contrast, rG-RPA predicts that only polyampholytes have inverse S-shape phase boundaries [Fig. 1(b)], whereas polyelectrolytes phase boundaries convex upward with a relatively flat ϕ_m dependence around the critical points [Fig. 1(a)]. This conspicuous difference between the rG-RPA-predicted phase boundaries of polyampholytes and polyelectrolytes is consistent with explicit-chain simulations.[28,74]

B. Salt-free rG-RPA account of pH-dependent LLPS

To address pH dependence under salt-free conditions, we apply rG-RPA to an example of a near-neutral polyampholyte under neutral pH, namely, the N-terminal IDR of the DEAD-box helicase Ddx4 (IDR denoted as Ddx4$^{\text{N1}}$) and its charge-scrambled variant Ddx4$^{\text{N1}}$CS which has the same amino acid composition as Ddx4$^{\text{N1}}$ by a different sequence charge pattern.[4] The sequences are studied at neutral and acidic pH. We refer to the resulting charge patterns as (in obvious notation) Ddx4$^{\text{N1}}_{\text{pH7}}$, Ddx4$^{\text{N1}}CS_{\text{pH7}}$, Ddx4$^{\text{N1}}_{\text{pH1}}$, and Ddx4$^{\text{N1}}CS_{\text{pH1}}$, where pH7 and pH1 are approximate pH values symbolizing neutral and acidic conditions. For the pH7 sequences, each of the 24 arginines (R) and 8 lysines (K) of Ddx4$^{\text{N1}}$ and Ddx4$^{\text{N1}}$CS is assigned a +1 charge, each of the 18 aspartic acids (D) and 18 glutamic acids (E) is assigned a −1 charge, and the 2 histidines (H) carry zero charge. For the pH1 sequences, because the pH is lower than the pKa of the acidic amino acids (3.71 for D and 4.15 for E), they are not ionized and thus carry zero charge but each K or R or H ($pK_{\text{H}} = 6.04$) carries a +1 charge [Fig. 2(a), K, R in blue; H in cyan].
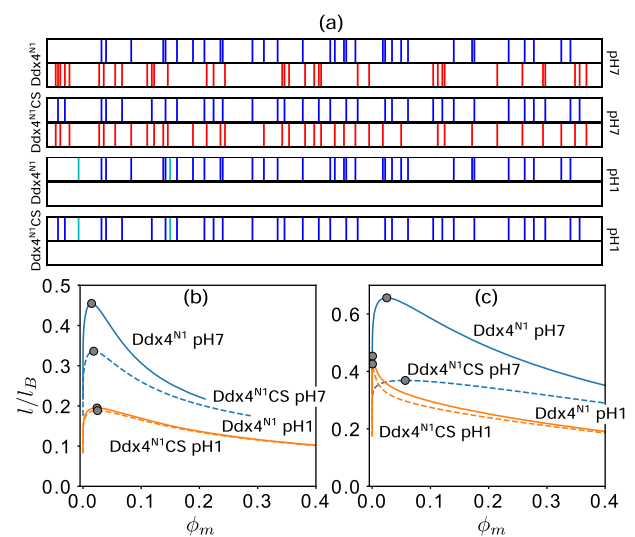

FIG. 2. LLPS at neutral and acidic pH. (a) Charge sequences of Ddx4$^{\text{N1}}$ and Ddx4$^{\text{N1}}$CS (blue/cyan: +1, red: −1, white: 0) and their (b) rG-RPA and (c) fG-RPA phase diagrams.

Thus, Ddx4$^{\text{N1}}_{\text{pH7}}$ and Ddx4$^{\text{N1}}$CS$_{\text{pH7}}$ are near-neutral polyampholytes, whereas Ddx4$^{\text{N1}}_{\text{pH1}}$ and Ddx4$^{\text{N1}}$CS$_{\text{pH1}}$ are polyelectrolytes although these four sequences—unlike those in Fig. 1—contains also many uncharged monomers.

Figure 2(b) indicates that the rG-RPA-predicted T_{cr} is much lower under acidic than under neutral conditions and that T_{cr} of Ddx4$^{\text{N1}}$ is always higher than that of Ddx4$^{\text{N1}}$CS under both pH conditions, underscoring that sequence-specific effects influence the LLPS of not only neutral and nearly neutral polyampholytes[28,55–57,79] but also polyelectrolytes. Intriguingly, inverse S-shaped coexistence curves are seen in Fig. 2(b) not only for neutral pH (blue curves) but also for acidic pH (orange curves). This feature is characteristic of polyampholytes [Fig. 1(b)] but not uniformly charged polyelectrolytes [Fig. 1(a)]. This result suggests that inverse S-shaped phase boundaries can arise, in general, from a heterogeneous sequence charge pattern because it leads to the simultaneous presence of both attractive and repulsive interchain interactions (which can be counterion-mediated in the case of polyelectrolytes) and therefore allows for condensed-phase configurations with lower densities.[28]

As a control, fG-RPA results are shown in Fig. 2(c). In contrast to rG-RPA, fG-RPA predicts that the $l/(l_B)_{\text{cr}}$ value (proportional to T_{cr}) of both Ddx4$^{\text{N1}}$ and Ddx4$^{\text{N1}}$CS at low pH is higher than that of Ddx4$^{\text{N1}}$CS at neutral pH and that the critical volume fractions at low pH are significantly lower than those at neutral pH. Although these differences between fG-RPA and rG-RPA predictions for the Ddx4 IDR remain to be conclusively tested by experiment, the low-pH fG-RPA phase diagrams here [orange curves in Fig. 2(c)] share similar features with the fG-RPA phase diagrams for polyelectrolytes in Fig. 1(c) which, as discussed above, are at odd with trends observed in prior theories and experiments. The fG-RPA results and those obtained using our earlier, simple formulation of RPA[55] are very similar (Fig. 3).
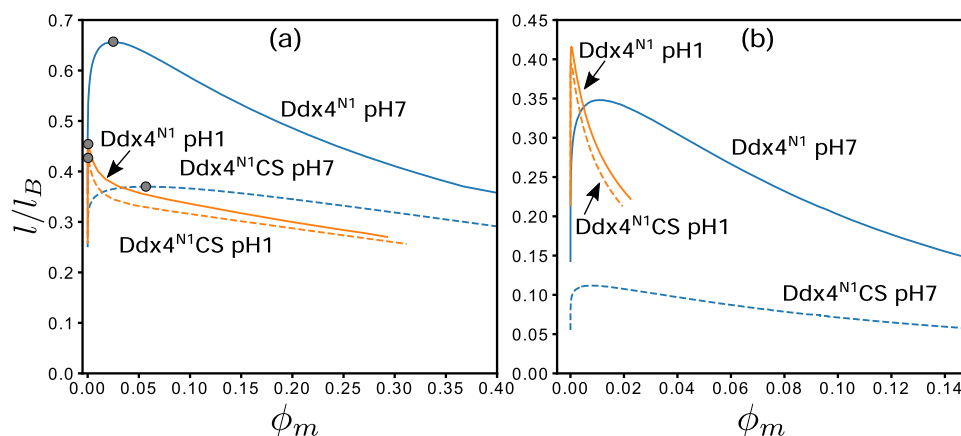
FIG. 3. Simple RPA[55,56] salt-free phase diagrams for the four Ddx4 sequences in Fig. 2(a). (a) Phase diagrams computed using the Coulomb potential in Fourier space, $U_k = 4\pi l_B/k^2$, are very similar to the fG-RPA phase diagrams in Fig. 2(c). (b) Phase diagrams computed using a Coulomb potential with a short-range cutoff, $U_k = 4\pi l_B/[k^2(1 + (kl)^2)]$; the same potential used in our previous simple-RPA studies.[24,55–58] This Coulomb potential with a short-range cutoff predicts that the two pH1 sequences have critical temperatures even higher than that of wildtype Ddx4 at pH7. This prediction, however, contradicts the physical intuition that polyelectrolytes should have lower phase separation propensities than neutral or near-neutral polyampholytes of the same chain length.

C. Salt-free rG-RPA rationalizes pH-dependent LLPS of IP5

We now utilize our theory to rationalize part of the experimental pH-dependent LLPS trend of the lyophilized 39-residue peptide IP5,[80] the isoelectric point of which is pH = 4.4 [Figs. 4(a) and 4(b)].[81] The pH-dependent charge σ of a basic or acidic residue is computed[82] here by

$$\sigma = \pm \frac{10^{\pm(pK_a - pH)}}{1 + 10^{\pm(pK_a - pH)}}, \tag{18}$$

where the + and − signs in the ± signs above apply to the basic (R, K, H) and acidic (D, E) residues, respectively. Standard pK_a values,[81] viz., R: 12.10, K: 10.67, H: 6.04, D: 3.71, and E: 4.15, are used in Eq. (18) to construct pH-dependent charge sequences of IP5 [Fig. 4(c)].

The rG-RPA-predicted and fG-RPA-predicted IP5 phase boundaries for the experimental studied pH values are shown in Fig. 4(d). Both theories predict a lower $l/(l_B)_{cr} \approx 0.2$–0.3 than the experiment $l/(l_B)_{cr} \approx 0.5$. Physically, this is not surprising, as has been addressed in previous RPA studies,[55] because nonelectrostatic cohesive interactions are neglected here. Nonetheless, consistent with the experiment, both theories posit that LLPS propensity decreases with increasing pH. Moreover, the rG-RPA-predicted critical volume fraction $(\phi_m)_{cr} \approx 0.020$–$0.024$ is reasonable in view of the experimental value of $\approx$0.036 (Ref. 80), indicating once again that rG-RPA is superior to fG-RPA as the latter predicts much higher $(\phi_m)_{cr}$'s.

D. Salt-dependent rG-RPA for heteropolymeric charge sequences

In view of the superiority of rG-RPA over fG-RPA, only rG-RPA is used below. We consider the four charge sequences in Fig. 2(a) as examples and restrict attention to monovalent salt and

counterions ($z_s = z_c = 1$). In experiments we conducted for this study using described methods,[24] no Ddx4^{N1} LLPS was observed in salt-free solution at room temperature; yet Ddx4^{N1} at room temperature is known[4,24] to phase separate with 100 mM NaCl and that LLPS propensity decreases when [NaCl] is increased to 300 mM. These findings suggest that, similar to LLPS of uniformly charged polyelectrolytes,[83–85] salt dependence of heteropolymer LLPS is nonmonotonic at temperatures slightly higher than the salt-free T_{cr} and therefore such temperatures are of particular interest. For this reason, we apply rG-RPA to compute IDR-salt binary phase diagrams of Ddx4$^{N1}_{pH7}$, Ddx4^{N1}CS$_{pH7}$, Ddx4$^{N1}_{pH1}$, and Ddx4^{N1}CS$_{pH1}$ (Fig. 5), each at an l/l_B value slightly higher than the sequence's salt-free $l/(l_B)_{cr}$ in Fig. 2(b).

As expected, all binary phase diagrams in Fig. 5 exhibit nonmonotonic salt dependence. In general, at temperatures above the salt free critical temperature, i.e., $l/l_B \gtrsim$ salt-free $l/(l_B)_{cr}$, when sufficient salt is added to the salt-free homogeneous solution, LLPS is triggered at $\phi_s = (\phi_s)^L_{cr}$. Adding more salt beyond $(\phi_s)^L_{cr}$ enhances LLPS in that a wider range of overall ϕ_m falls within the LLPS regime, until a turning point $(\phi_s)^T$ is reached. Beyond that, adding more salt [increasing ϕ_s above $(\phi_s)^T$] reduces LLPS (the phase-separated range of ϕ_m narrows). LLPS is impossible for the given temperature when salt concentration is increased above an upper critical point $(\phi_s)^U_{cr}$.

Despite these qualitative commonalities, there are significant sequence-dependent differences. Notably, at neutral pH, the range of salt concentrations that can induce LLPS is much narrower for Ddx4$^{N1}_{pH7}$ [$\phi_s \lesssim 0.000\,85$, Fig. 5(a)] than for Ddx4^{N1}CS$_{pH7}$ [$\phi_s \lesssim 0.005$, Fig. 5(b)]. However, the ranges of LLPS-inducing salt concentrations at low pH for Ddx4$^{N1}_{pH1}$ and Ddx4^{N1}CS$_{pH1}$ are similar [$\phi_s \lesssim 0.01$, Figs. 5(c) and 5(d)], and their $(\phi_s)^L_{cr}$ and $(\phi_s)^U_{cr}$ are significantly larger than those at neutral pH.

Next, we explore these trends at temperatures below salt-free T_{cr}. Figures 6–9 present salt-polymer phase diagrams for four Ddx4 sequences (both wild type and charge scrambled sequences at neutral and acidic pH) at three different temperatures. Panels (a) and

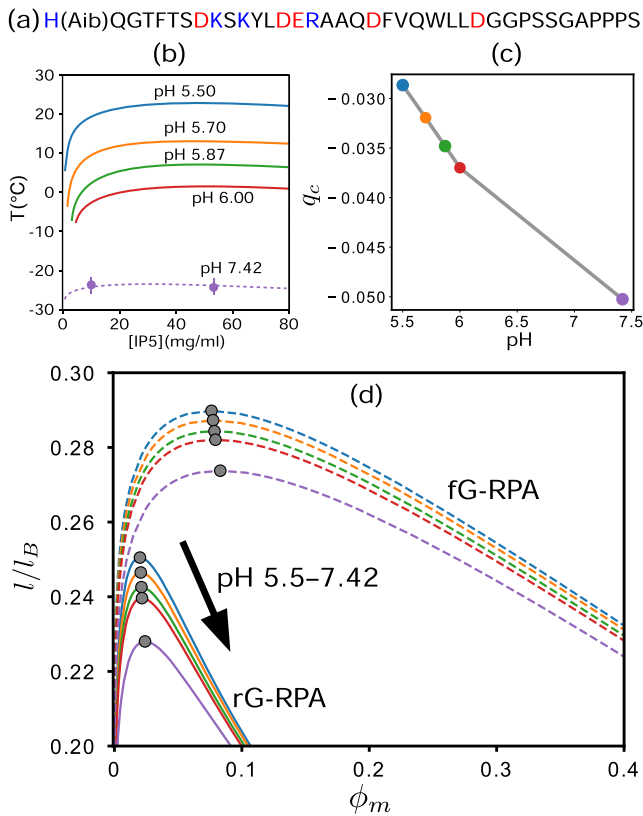(a) H(Aib)QGTFTSDKSKYLDERAAQDFVQWLLDGGPSSGAPPPS

FIG. 4. LLPS of IP5. (a) The IP5 sequence, where basic and acidic residues are in blue and red, respectively; (Aib) is the nonproteinogenic amino acid α-methylalanine.[80] (b) Experimental pH-dependent phase diagrams of IP5 based on the data in Fig. 4 of Ref. [80]; antifreeze was used to obtain some of the low-T results.[80] (c) Net charge per residue, q_c, of IP5. (d) Phase diagrams predicted by rG-RPA (solid curves) and fg-RPA (dashed curves).
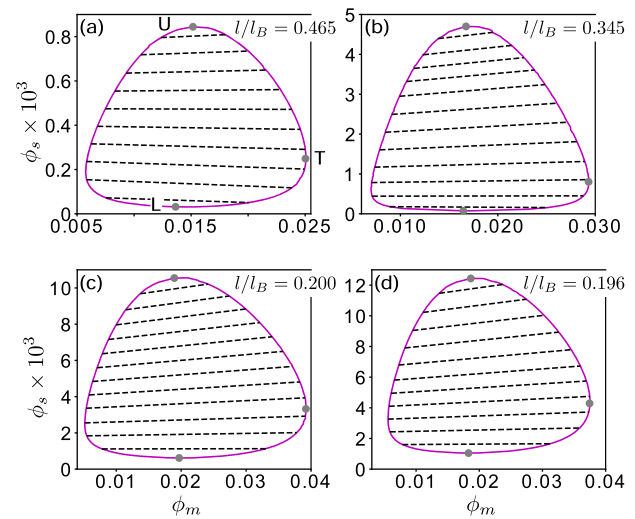

FIG. 5. IDR-salt binary phase diagrams of two Ddx4 variants at low and high pH. Results are for $l/l_B \gtrsim l/(l_B)_{\mathrm{cr}}$, where the salt-free $1/(l_B)_{\mathrm{cr}}$ equals 0.455 for $\mathrm{Ddx4}^{\mathrm{N1}}_{\mathrm{pH7}}$ (a), 0.336 for $\mathrm{Ddx4}^{\mathrm{N1}}\mathrm{CS}_{\mathrm{pH7}}$ (b), 0.195 for $\mathrm{Ddx4}^{\mathrm{N1}}_{\mathrm{pH1}}$ (c), and 0.188 for $\mathrm{Ddx4}^{\mathrm{N1}}\mathrm{CS}_{\mathrm{pH1}}$ (d). The ϕ_s values of the gray circles in (a)–(d) are $(\phi_s)^{\mathrm{U}}_{\mathrm{cr}}$, $(\phi_s)^{\mathrm{T}}$, or $(\phi_s)^{\mathrm{L}}_{\mathrm{cr}}$, as indicated by U, T, and L in (a).

(b) in these figures show phase diagrams at temperatures below the respective salt free T_{cr} for the given sequence, while panel (c) is at a temperature above salt free T_{cr}. The three phase diagrams are compared in panel (d) for a given sequence. These figures reveal that trends for $l/l_B \gtrsim$ salt-free $l/(l_B)_{\mathrm{cr}}$ (above the salt free critical temperature) are largely in line with behaviors at temperatures below salt-free T_{cr}. The only difference is for $l/l_B <$ salt-free $l/(l_B)_{\mathrm{cr}}$, $(\phi_s)^{\mathrm{L}}_{\mathrm{cr}} = 0$. For $l/l_B <$ salt-free $l/(l_B)_{\mathrm{cr}}$, temperatures for different sequences were chosen such that the maximum ϕ_m range of LLPS are similar among the sequences (as in Fig. 5). With this choice of temperature constraint, when the IDR-salt phase diagrams for different sequences (Figs. 6–9) are compared, we note that $(\phi_s)^{\mathrm{U}}_{\mathrm{cr}}$ and $(\phi_s)^{\mathrm{T}}$ of $\mathrm{Ddx4}^{\mathrm{N1}}_{\mathrm{pH7}}$ are much smaller than those of $\mathrm{Ddx4}^{\mathrm{N1}}\mathrm{CS}_{\mathrm{pH7}}$. Furthermore, $(\phi_s)^{\mathrm{U}}_{\mathrm{cr}}$ and $(\phi_s)^{\mathrm{T}}$ of these two pH7 sequences are much smaller than those of the two pH1 sequences. Thus, we conclude that $\mathrm{Ddx4}^{\mathrm{N1}}_{\mathrm{pH7}}$ is more sensitive to salt than $\mathrm{Ddx4}^{\mathrm{N1}}\mathrm{CS}_{\mathrm{pH7}}$, and both are more salt-sensitive than $\mathrm{Ddx4}^{\mathrm{N1}}_{\mathrm{pH1}}$ and $\mathrm{Ddx4}^{\mathrm{N1}}\mathrm{CS}_{\mathrm{pH1}}$. Metrics other than $(\phi_s)^{\mathrm{T}}$ can also be used to determine salt sensitivity. For example, the low-ϕ_m turning point [e.g., at $\phi_m \approx 0.006$, $\phi_s \approx 0.16$ in Fig. 5(a), unlabeled] with a ϕ_s value similar to that of $(\phi_s)^{\mathrm{T}}$ may be used to characterize salt

sensitivity. The resulting trend is similar to the one gleaned from the turning point $(\phi_s)^{\mathrm{T}}$.

The existence of $(\phi_s)^{\mathrm{L}}_{\mathrm{cr}} > 0$ in Fig. 5(a) is consistent with our experimental observation that $\mathrm{Ddx4}^{\mathrm{N1}}$ does not phase separate with [NaCl] < 15–20 mM at pH 6.5, 25 °C ($l/l_B = 0.529$), and 5 mM Tris. Other predictions of our theory remain to be tested. Of particular interest is the slopes of the tie lines in Figs. 5(a) and 5(b) that change from negative to positive as ϕ_s increases, indicating that salt ions and the heteropolymeric IDRs partially exclude each other in low-salt but partially coalesce in high-salt solutions at neutral pH. This intriguing feature was not encountered in solutions of either a single species of uniformly charged or two species of oppositely charged homopolymers[42–44,48,72,86,87] until very recently [Ref. 46]. In contrast, the tie-line slopes in Figs. 5(c) and 5(d) are all positive, indicating that salt ions and the heteropolymeric IDRs always partially coalesce under acidic conditions.

E. Salt-dependent rG-RPA is consistent with established trends in LLPS of homopolymeric, uniformly charged polyelectrolytes

Our model predicts salt and polymers coalesce for $\mathrm{Ddx4}^{\mathrm{N1}}_{\mathrm{pH1}}$ and $\mathrm{Ddx4}^{\mathrm{N1}}\mathrm{CS}_{\mathrm{pH1}}$ [Figs. 5(c) and 5(d)]. These sequences are examples of nonuniformly charged polyelectrolytes. However, these results are in contrast to the experiment and theory on uniformly charged polyelectrolytes that suggest salt ions and polymers tend to exclude each other, leading to tie lines with negative slopes in the polymer-salt phase diagrams.[72,83,86,87] We test the ability of our model to reproduce this established trend by computing salt-polymer phase diagrams for uniformly charged polymers [Fig. 10(a)]. The established feature is captured by our new theory as the slopes of all tie
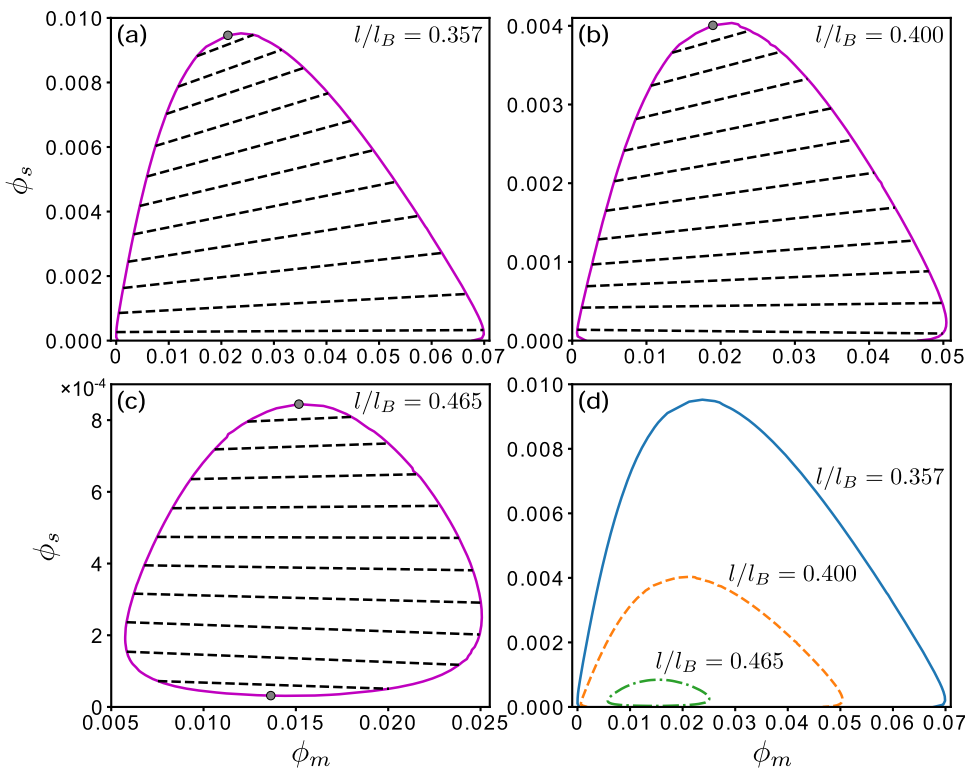
FIG. 6. Polymer-salt coexistence phase diagrams of $\text{Ddx4}^{\text{N1}}_{\text{pH7}}$ at the l/l_B values indicated. The salt-free critical value of l/l_B is $l/(l_B)_{\text{cr}} = 0.455$. Top gray circles in (a)–(c) provide the upper critical salt concentrations $(\phi_s)^{\text{U}}_{\text{cr}}$, whereas the bottom gray circle in (c) provides the lower critical concentration $(\phi_s)^{\text{L}}_{\text{cr}}$ (see discussion in main text). Each dashed line in (a)–(c) is a tie line connecting a pair of coexistent phases. The three phase boundaries in (a)–(c) are compared in (d).

FIG. 7. Polymer-salt coexistence phase diagrams of $\text{Ddx4}^{\text{N1}}\text{CS}_{\text{pH7}}$ at the l/l_B values indicated. The salt-free critical value of l/l_B is $l/(l_B)_{\text{cr}} = 0.336$. Top gray circles in (a)–(c) provide the upper critical salt concentrations $(\phi_s)^{\text{U}}_{\text{cr}}$, whereas the bottom gray circle in (c) provides the lower critical concentration $(\phi_s)^{\text{L}}_{\text{cr}}$. Each dashed line in (a)–(c) is a tie line connecting a pair of coexistent phases. The three phase boundaries in (a)–(c) are compared in (d).
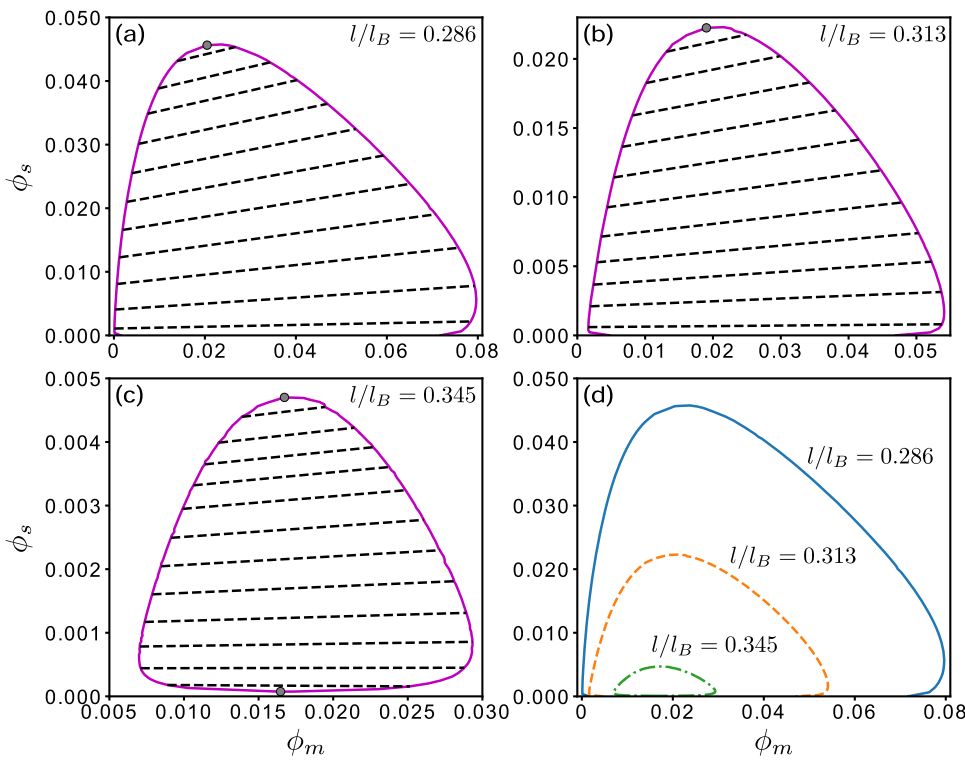
FIG. 8. Polymer-salt coexistence phase diagrams of $Ddx4_{pH1}^{N1}$ at the l/l_B values indicated. The salt-free critical value of l/l_B is $l/(l_B)_{cr} = 0.195$. Top gray circles in (a)–(c) provide the upper critical salt concentrations $(\phi_s)_{cr}^{U}$, whereas the bottom gray circle in (c) provides the lower critical concentration $(\phi_s)_{cr}^{L}$. Each dashed line in (a)–(c) is a tie line connecting a pair of coexistent phases. The three phase boundaries in (a)–(c) are compared in (d).
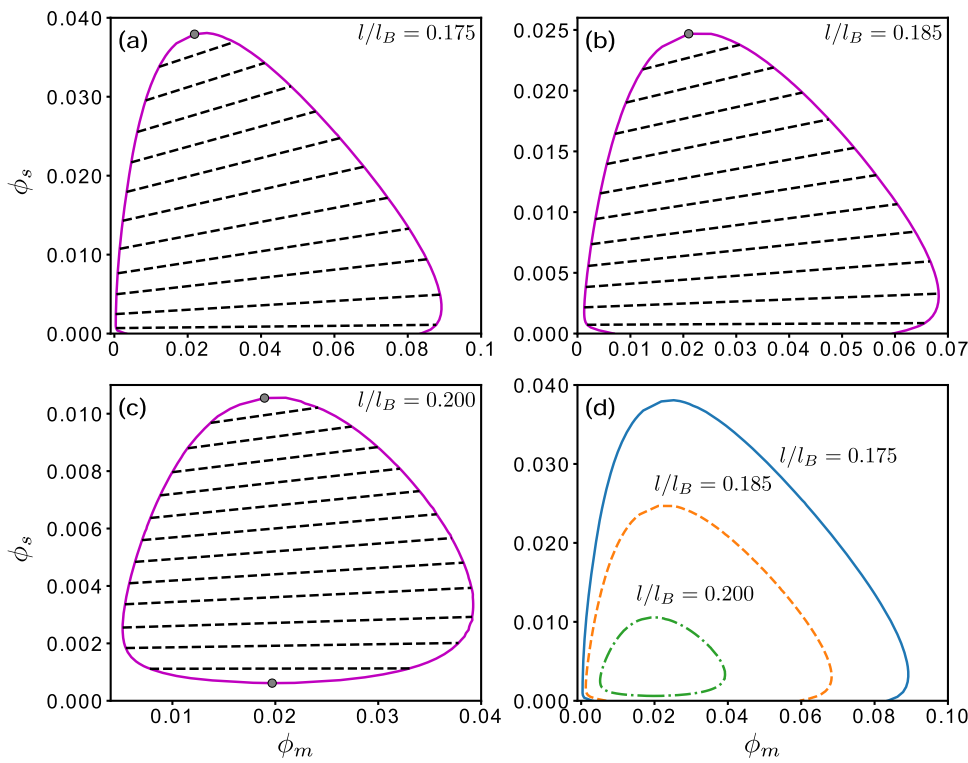

FIG. 9. Polymer-salt coexistence phase diagrams of $Ddx4^{N1}CS_{pH1}$ at the l/l_B values indicated. The salt-free critical value of l/l_B is $l/(l_B)_{cr} = 0.188$. Top gray circles in (a), (b), and (c) provide the upper critical salt concentrations $(\phi_s)_{cr}^{U}$, whereas the bottom gray circle in (c) provides the lower critical concentration $(\phi_s)_{cr}^{L}$. Each dashed line in (a)–(c) is a tie line connecting a pair of coexistent phases. The three phase boundaries in (a)–(c) are compared in (d).
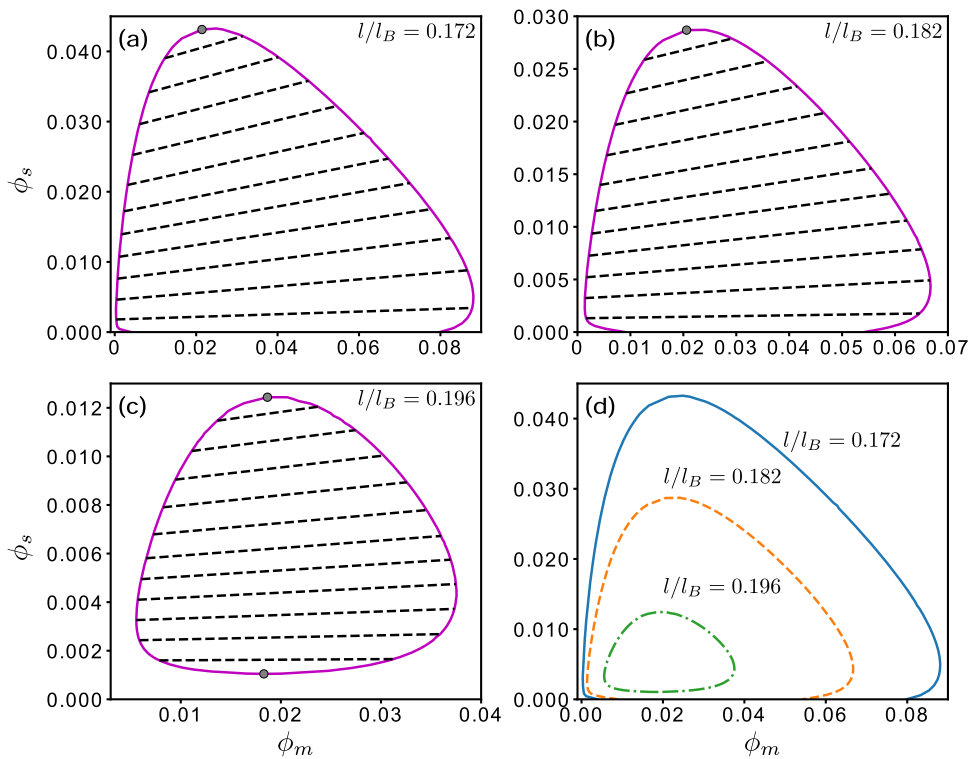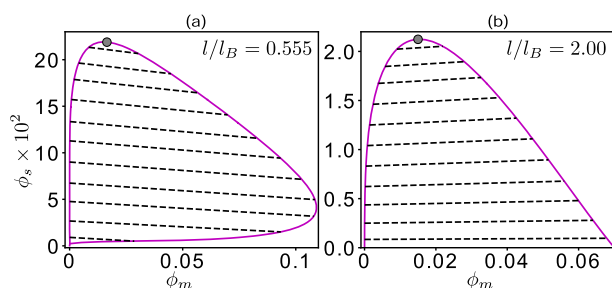
FIG. 10. Salt-dependent LLPS of polyelectrolytes and polyampholytes. rG-RPA phase diagrams for (a) an $N = 50$ homopolymer with monomer charge $= -1$ and (b) the $N = 40$ 4-block polyampholyte in Fig. 1. Note that salt-free $l/(l_B)_{cr} = 0.5$ for (a) and $= 3.63$ for (b). $(\phi_s)_{cr}^U$ is given by the gray circle. An unmarked $\phi_s = (\phi_s)_{cr}^L > 0$ exists for (a) but not for (b).

lines are negative in Fig. 10(a). Furthermore, consistent with literature reports on uniformly charged homopolymers (homopolyelectrolytes),[83–85] with addition of salt, rG-RPA predicts a one-to-two phase transition in the low salt regime as well as a two-to-one phase transition in the high salt regime. For comparison, Fig. 10(b) is the phase diagram of an overall neutral polyampholytes at a temperature substantially lower than the salt-free T_{cr} with all tie lines having positive slopes. A recent field theory simulation study of an overall neutral diblock polyampholyte also found tie lines with slightly positive slopes.[32] Since tie lines with exclusively positive slopes are also seen for the overall negatively charged low-pH Ddx4 IDRs above, the opposite-signed tie-line slopes in Fig. 10(a) for homopolymeric and those in Figs. 5(c) and 5(d) for heteropolymeric polyelectrolytes suggest a role of sequence heterogeneity in determining whether charged polymers tend to exclude or coalesce with salt ions. However, the precise origins of variation in tie-line slope remains to be ascertained. One idea is that the nonzero tie-line slopes arise from chain connectivity of polymers. If the polymers were not connected and behave like a collection of monomers, the salt concentrations in the dilute and condensed phases would simply follow that of the polymer leading to the positive slope.[46] However, chain connectivity can change the slope from positive to negative.

The nature of tie-line slopes has also received considerable attention in the salt-polymer phase diagrams observed during complex coacervation of symmetric polyelectrolytes.[40,42,44,46–49] Insights gleaned from these studies can yield clues to tie-line slope differences observed in our analysis. A recent theory[46] based on the concept of chain connectivity predicts a salt-concentration-dependent change of sign of the tie-line slope, exhibiting a behavior similar to that in Figs. 5(a) and 5(b). However, in this case of coacervation, the slope changes from positive to negative with addition of salt, opposite to the case of heteropolymers described here. Another idea is that the tie-line slope is determined by a competition between electrostatic interactions among polymers and configurational entropy of the salt ions,[47] whereby the magnitude of electrostatic interactions in the condensed phase are enhanced by reduced salt because of less screening but any difference in concentration in salt ions between the dilute and condensed phases is entropically unfavorable. It is intuitive that both of these proposed mechanisms—conjectured in modeling coacervation—would be affected by the charge pattern

of the polymers, but the manner in which the proposed mechanisms are modulated by sequence heterogeneity remains to be investigated.

F. rG-RPA rationalizes sequence-dependent LLPS of Ddx4 IDRs

Simple RPA theory and an extended RPA+FH theory with an augmented Flory-Huggins (FH) mean-field account of nonelectrostatic interactions were utilized to rationalize[24,55,56] experimental data on sequence- and salt-dependent LLPS of Ddx4 IDRs.[4,24] Because RPA accounts only for electrostatic interactions and a sequence-specific analytical treatment of other interactions is currently lacking, FH was used to provide an approximate account of nonelectrostatic interactions. These interactions can include hydrophobicity, hydrogen bonding, and especially cation-π and π-π interactions because π-related interactions play prominent roles in LLPS of biomolecular condensates.[88] To gain further insight into the semiquantitative picture emerged from these earlier studies[24,55,56] and to assess the generality of our rG-RPA theory, here we apply an augmented rG-RPA to the LLPS of the same Ddx4[N1] and Ddx4[N1]CS sequences by adding to the rG-RPA free energy in Eq. (1) an FH interaction term $-\chi \phi_m^2$, where $\chi = \Delta H(l_B/l) - \Delta S$ contains both enthalpic and entropic components, and refer to the resulting formulation as rG-RPA+FH.

To compare with experimental data,[24] we use this theory to compute the phase diagrams of Ddx4[N1] and Ddx4[N1]CS at pH 6.5 with 100 and 300 ml NaCl, which correspond, respectively, to $\phi_s = 0.0018$ and 0.0054. Naturally, pH-dependent behaviors can also be obtained by the same FH term together with Eq. (1) and Eq. (18) for rG-RPA free energy; but here we do not pursue a pH-dependent rG-RPA+FH analysis of Ddx4[N1] and Ddx4[N1]CS LLPS because no corresponding experimental data is currently available for comparison.

Our detailed rG-RPA study of salt-Ddx4[N1] and salt-Ddx4[N1]CS binary phase diagrams in Figs. 5–9 indicates that the difference between dilute- and condensed-phase salt concentrations is less than 15% for $\phi_s < 0.01$. Assuming that this trend is not much affected by nonelectrostatic interactions, here we make the simplifying assumption that salt concentration is constant when determining the rG-RPA+FH phase diagrams. Figure 11(a) shows that the resulting rG-RPA+FH theory with $\chi = 0.5(l_B/l)$ fits reasonably well with all four available experimental phase diagrams.

As control, phase diagrams are also computed without the augmented FH term (i.e., $\chi = 0$). These phase diagrams are shown as dashed lines in Fig. 11(b). Without the χ term, the critical temperatures of Ddx4[N1] and Ddx4[N1]CS with [NaCl] = 100 mM are both predicted to be below 0 °C [Fig. 11(b)]. This theoretical trend is consistent with the experimental observation that phenylalanine to alanine (F-to-A) and arginine to lysine (R-to-K) mutants of Ddx4[N1] do not undergo LLPS at physiologically relevant temperatures.[4,24,88] These mutations (F-to-A and R-to-K) are expected to significantly reduce π-related interactions[88] and therefore correspond to having a weaker FH term (i.e., a smaller χ).

One aforementioned experimentally observed feature that cannot be captured by the present rG-RPA+FH theory is that in the absence of salt, Ddx4[N1] at pH 6.5 does not phase separate at room temperature, but rG-RPA+FH with $\chi = 0.5(l_B/l)$ predicts phase
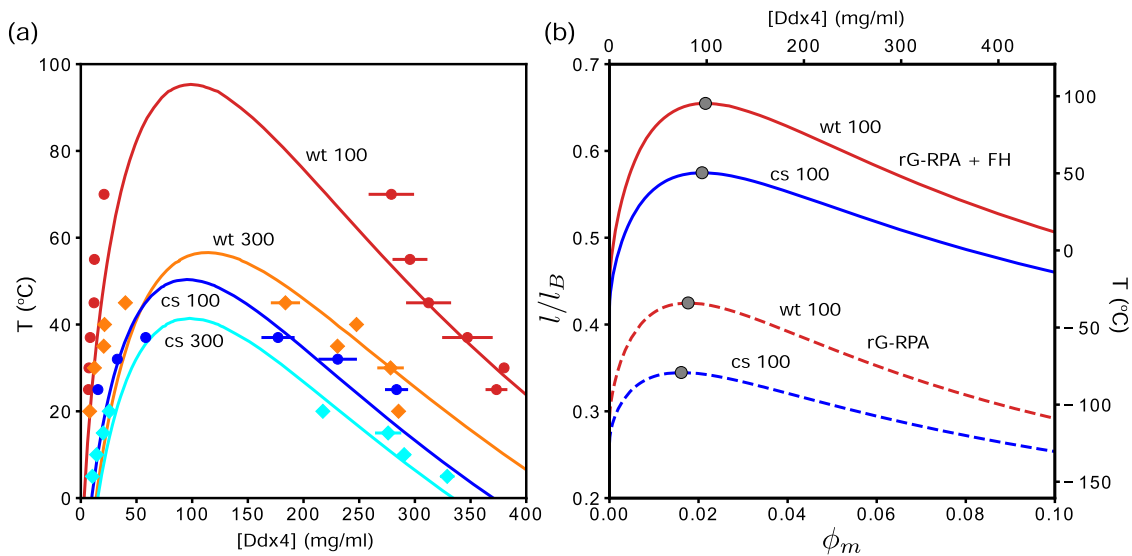
FIG. 11. Comparing rG-RPA+FH results with experimental data on Ddx4 IDRs. (a) Experimental data of Ddx4$^{N1}_{pH6.5}$ (wt) and Ddx4^{N1}CS$_{pH6.5}$ (cs) (chain length N = 241 for both sequences) in aqueous solutions with 100 and 300 mM NaCl (from Ref. 24; color symbols) are fitted, respectively, to the rG-RPA+FH theory with ϕ_s = 0.0018 and 0.0054 (continuous curves with the same color). For simplicity, the salt concentrations in the dilute and condensed Ddx4 phases are taken to be identical in this calculation. This is a reasonable approximation because the salt-Ddx4^{N1} binary phase diagrams in Fig. 5 indicate that the difference in salt concentration between the two phases is less than 15% for ϕ_s < 0.01. The fits yield an FH interaction parameter χ = 0.5(l_B/l) which is equivalent to an enthalpy ΔH = −0.56 kcal/mol favorable to polymer-polymer attraction. Model temperatures and model polymer volume fractions are converted, respectively, to °C and mg/ml by a procedure similar to that in Ref. 55 with an appropriately chosen model Kuhn length l that is quite similar to (though not identical with) the C$_\alpha$–C$_\alpha$ virtual bond length of polypeptides. (b) Phase diagrams of the two sequences with and without the augmented FH interaction. Without the FH term (i.e., χ = 0), the critical temperatures of both Ddx4$^{N1}_{pH6.5}$ and Ddx4^{N1}CS$_{pH6.5}$ at 100 mM NaCl are below 0 °C. The two χ = 0 systems may be interpreted as corresponding to sequences with reduced favorable nonelectrostatic interactions.[24,88] See the text for further discussion.

separation under the same conditions. There can be multiple reasons for this mismatch between the theory and experiment, a likely one of which is that the mean-field treatment of nonelectrostatic interactions does not take into possible coupling (cooperative effects) between sequence-specific electrostatic and nonelectrostatic interactions such as π-related interactions and hydrogen bonding that can be enhanced by proximate electrostatic attraction.

IV. CONCLUSIONS

In summary, we have developed a formalism for salt-, pH-, and sequence-dependent LLPS by combining RPA and Kuhn-length renormalization. The trends predicted by the resulting rG-RPA theory are consistent with established theoretical and experimental results. Importantly, unlike more limited previous analytical approaches, rG-RPA is generally applicable to both polyelectrolytes and neutral/near-neutral polyampholytes. In addition to providing physical rationalizations for experimental data on the pH-dependent LLPS of IP5 peptides and sequence and salt dependence of LLPS of Ddx4 IDRs, our theory offers several intriguing predictions of electrostatics-driven LLPS properties that should inspire further theoretical studies and experimental evaluations. One such observation is that in a salt-heteropolymer system, it is possible for the slope of the tie lines to shift from negative to positive by increasing salt. Although tie lines with exclusively positive or exclusively negative slopes were predicted for uniformly charged

polyelectrolytes and diblock polyampholytes,[32,42,43,72,86,87] a salt-dependent change in the sign of the tie-line slope for a single species of heteropolymer—specifically from negative to positive with increasing salt—is a notable prediction. In future studies, it would be interesting to explore how this property might have emerged from the intuitively higher degree of sequence heterogeneity of the Ddx4^{N1} IDR vis-à-vis that of simple diblock or few-block polyampholytes. In general, the interplay between sequence heterogeneity and a proposed chain connectivity effect[46] as well as a proposed screening-configurational entropy competition effect[47] on the salt partitioning slope between dilute and condensed phases remains to be elucidated. Another observation of our work is that inverse S-shape coexistence curves can arise from sequence heterogeneity not only for polyampholytes[55–57] but also for polyelectrolytes. As emphasized recently,[28] an inverse S-shape coexistence curve allows for a less concentrated condensed phase, which can be of biophysical relevance because it would enable a condensate with higher permeability.[89]

Because rG-RPA is an analytical theory, pertinent numerical computations are much more efficient than field-theory or explicit-chain simulations. Thus, in view of the above advances and despite its approximate nature, rG-RPA should be useful as a high-throughput tool for assessing sequence-dependent LLPS properties in developing basic biophysical understanding and in practical applications such as design of new heteropolymeric materials.

Future development of LLPS theory should address a number of physical properties not tackled by our current theories. These include, but are not necessarily limited to, the following: (i) Sequence-dependent effects of nonelectrostatic interactions, which is neglected in rG-RPA+FH. (ii) Counterion condensation.[72,74,90,91] (iii) Dependence of relative permittivity (dielectric constant) on polymer density[56,58] and salt.[92] (iv) A more accurate treatment of conformational heterogeneity to compute the structure factor. The present approach accounts approximately for the sequence-dependent end-to-end distance, but it fails to capture conformational heterogeneities at smaller length scales.[72] A formalism for the residue-pair-specific renormalized Kuhn length[68,93] should afford improvement in this regard. (v) Higher-order density fluctuations beyond the quadratic fluctuations[59] treated by rG-RPA. The rapidly expanding repertoire of experimental data on biomolecular condensates is providing impetus for theoretical efforts in all these directions.

ACKNOWLEDGMENTS

APPENDIX A: DERIVATION OF POLYMER SOLUTION FREE ENERGY

As described in the main text, we consider a neutral solution of n_p charged polymers of N monomers (residues) with charge sequence $|\sigma\rangle = [\sigma_1, \sigma_2, \ldots \sigma^N]^T$. Averaged net charge per monomer is defined as $q_c = (\sum_\tau \sigma_\tau)/N$. In addition, there are n_s salt ions (coions) carrying z_s charges and n_c counterions carrying z_c charges. Charge neutrality $|q_c|n_p + z_s n_s = z_c n_c$ is always preserved. Monomer and ion densities are defined as $\rho_m = n_p N/\Omega$, $\rho_s = n_s/\Omega$, and $\rho_c = n_c/\Omega$, respectively, with Ω being the solution volume.

We label the polymers by $\alpha = 1, 2, \ldots, n_p$ and residues in a polymer by $\tau = 1, 2, \ldots, N$ and denote the spatial coordinate of the τth monomer in the αth polymer by $\mathbf{R}_{\alpha,\tau}$. Similarly, the small ions are labeled by $a = 1, 2, \ldots, n_s + n_c$, in which $1 \le a \le n_s$ are for salt ions and $n_s + 1 \le a \le n_s + n_c$ are for counterions, with the coordinate of the ath small ion denoted by $\mathbf{r}_a$. The implicit-solvent partition function is then expressed as an integral over all solute coordinates divided by factorials that account for the indistinguishability of the molecules within each molecular species in the solution, viz.,

$$\mathcal{Z} = \frac{1}{n_p! n_c! n_s! n_w!} \int \prod_{\alpha=1}^{n_p} \prod_{\tau=1}^{N} d\mathbf{R}_{\alpha,\tau} \prod_{a=1}^{n_s+n_c} d\mathbf{r}_a e^{-\mathcal{T}[\mathbf{R}] - \mathcal{U}[\mathbf{R},\mathbf{r}]}, \quad (A1)$$

where n_w denotes the number of water molecules, $\mathcal{T}$ accounts for chain connectivity of the polymers, and $\mathcal{U}$ accounts for interactions among all solute molecules, $[\mathbf{R}]$ is shorthand for $[\{\mathbf{R}_{\alpha,\tau}\}]$, and

$[\mathbf{R}, \mathbf{r}]$ is shorthand for $[\{\mathbf{R}_{\alpha,\tau}\}, \{\mathbf{r}_a\}]$. Connectivity is enforced by a sum of Gaussian potentials sharing the same Kuhn length l, which is given by

$$\mathcal{T}[\mathbf{R}] = \frac{3}{2l^2} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2. \quad (A2)$$

For simplicity, we assume that interactions in $\mathcal{U}$ are all pairwise, in which case it takes the form

$$\mathcal{U}[\mathbf{R},\mathbf{r}] = \frac{1}{2} \sum_{\alpha,\beta=1}^{n_p} \sum_{\tau,\mu=1}^{N} \mathcal{U}_{pp}^{\tau\mu}(\mathbf{R}_{\alpha,\tau} - \mathbf{R}_{\beta,\mu}) + \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \sum_{a=1}^{n_s+n_c} \mathcal{U}_{ps}^{\tau a}(\mathbf{R}_{\alpha,\tau} - \mathbf{r}_a)$$
$$+ \frac{1}{2} \sum_{a,b=1}^{n_s+n_c} \mathcal{U}_{ss}^{ab}(\mathbf{r}_a - \mathbf{r}_b), \quad (A3)$$

where $\mathcal{U}_{pp}$, $\mathcal{U}_{ps}$, and $\mathcal{U}_{ss}$ are, respectively, monomer-monomer, monomer-ion, and ion-ion interaction potentials. It should be noted that although self-interactions, that is, the $(\alpha, \tau) = (\beta, \mu)$ terms for monomers and the $a = b$ terms for small ions, are included in the above summation to facilitate subsequent formal development of a field-theory description, these divergent terms will be regularized in the final free energy expression and thus have no bearing on the outcome of our theory. By introducing

$$\rho_{\mathbf{k}}^\tau = \sum_{\alpha=1}^{n_p} e^{i\mathbf{k}\cdot\mathbf{R}_{\alpha,\tau}}, \quad (A4a)$$

$$c_{\mathbf{k}}^s = \sum_{a=1}^{n_s} e^{i\mathbf{k}\cdot\mathbf{r}_a}, \quad (A4b)$$

$$c_{\mathbf{k}}^c = \sum_{a=1}^{n_c} e^{i\mathbf{k}\cdot\mathbf{r}_{a+n_s}}, \quad (A4c)$$

as the $\mathbf{k}$-space density operators for the monomers and small ions, we rewrite Eq. (A3) in $\mathbf{k}$-space as

$$\mathcal{U} = \frac{1}{2\Omega} \sum_{\mathbf{k}} \left[\sum_{\tau,\mu=1}^{N} \rho_{\mathbf{k}}^\tau \mathcal{U}_{pp}^{\tau\mu}(\mathbf{k})\rho_{-\mathbf{k}}^\mu + 2\sum_{\tau=1}^{N} \sum_{\gamma=s,c} \rho_{\mathbf{k}}^\tau \mathcal{U}_{ps}^{\tau\gamma}(\mathbf{k})c_{-\mathbf{k}}^\gamma \right.$$
$$\left. + \sum_{\gamma,\gamma'=s,c} c_{\mathbf{k}}^\gamma \mathcal{U}_{ss}^{\gamma\gamma'}(\mathbf{k})c_{-\mathbf{k}}^{\gamma'} \right], \quad (A5)$$

where $1/\Omega$ is the standard normalization factor for the Fourier transformation, and the general form $\mathcal{U}(\mathbf{k}) = \int d\mathbf{r}\mathcal{U}(\mathbf{r}) \exp(-i\mathbf{k} \cdot \mathbf{r})$ represents the interaction potentials in $\mathbf{k}$-space. As in Eq. (A3) for $\mathcal{U}[\mathbf{R},\mathbf{r}]$, the superscripts of $\mathcal{U}(\mathbf{k})$ are labels for monomers and ions, and the subscripts specify the interaction type. We further define interaction matrices $\hat{\mathcal{U}}(\mathbf{k})$'s by equating the matrix elements $[\hat{\mathcal{U}}(\mathbf{k})]_{\tau\mu}$ with $\mathcal{U}^{\tau\mu}(\mathbf{k})$ for $\mathcal{U}_{pp}$, $\mathcal{U}_{ps}$, and $\mathcal{U}_{ss}$. We also define the density operator vectors $|\rho_{\mathbf{k}}\rangle$ and $|c_{\mathbf{k}}\rangle$ such that $(|\rho_{\mathbf{k}}\rangle)_\tau = \rho_{\mathbf{k}}^\tau$ and $|c_{\mathbf{k}}\rangle = [c_{\mathbf{k}}^s, c_{\mathbf{k}}^c]^T$. $\mathcal{U}$ can then be expressed in the matrix representation as

$$\mathcal{U} = \frac{1}{2\Omega} \sum_{\mathbf{k}} \left[\langle \rho_{-\mathbf{k}}|\hat{\mathcal{U}}_{pp}(\mathbf{k})|\rho_{\mathbf{k}}\rangle + 2\langle \rho_{-\mathbf{k}}|\hat{\mathcal{U}}_{ps}(\mathbf{k})|c_{\mathbf{k}}\rangle + \langle c_{-\mathbf{k}}|\hat{\mathcal{U}}_{ss}(\mathbf{k})|c_{\mathbf{k}}\rangle \right]. \quad (A6)$$

The present study focuses on solution systems in which $\mathcal{U}_{ss}$ and $\mathcal{U}_{ps}$ are purely Coulombic, whereas $\mathcal{U}_{pp}$ has both Coulombic

and pairwise (two-body) excluded-volume repulsion components. Hence,

$$\hat{\mathcal{U}}_{ss}(\mathbf{k}) = \frac{4\pi l_B}{k^2}|z\rangle\langle z|, \tag{A7a}$$

$$\hat{\mathcal{U}}_{ps}(\mathbf{k}) = \frac{4\pi l_B}{k^2}|\sigma\rangle\langle z|, \tag{A7b}$$

$$\hat{\mathcal{U}}_{pp}(\mathbf{k}) = \frac{4\pi l_B}{k^2}|\sigma\rangle\langle\sigma| + v_2|1_N\rangle\langle 1_N|, \tag{A7c}$$

where $k \equiv |\mathbf{k}|$ and $l_B \equiv e^2/(4\pi\epsilon k_B T)$ is the Bjerrum length (e is the electronic charge, ϵ is the permittivity, k_B is the Boltzmann constant, T is the absolute temperature). $\langle z| = \text{sign}(q_c)[z_s, -z_c]$ is the vector representing the charge valences (number of electronic charges per ion) of salt ions and counterions, respectively, $v_2 > 0$ is the strength of the two-body excluded volume repulsion between monomers, and $|1_N\rangle$ is an N-dimensional vector in which every component is 1. All elements in the excluded volume matrix $|1_N\rangle\langle 1_N|$ take unity value because for simplicity all monomers are taken to be of equal size. Substituting the potentials given by Eq. (A7) into the $\mathcal{U}$ function in Eq. (A6) yields

$$\mathcal{U} = \frac{1}{2\Omega}\sum_{\mathbf{k}\neq 0}\lambda_k|\langle\sigma|\rho_{\mathbf{k}}\rangle + \langle z|c_{\mathbf{k}}\rangle|^2 + \frac{1}{2\Omega}\sum_{\mathbf{k}}v_2|\langle 1_N|\rho_{\mathbf{k}}\rangle|^2, \tag{A8}$$

where $\lambda_k = 4\pi l_B/k^2$ and $|A_{\mathbf{k}}|^2 \equiv A_{-\mathbf{k}}A_{\mathbf{k}}$ for arbitrary $\mathbf{k}$-dependent $A_{\mathbf{k}}$. The first summation does not need to include $\mathbf{k} = 0$ because this term is proportional to the overall net charge of the solution and therefore must be zero because of overall electric neutrality of the solution.

1. Field theory for polymer solution

The Hubbard-Stratonovich transformation is then applied to linearize the quadratic form $\mathcal{U}$ in Eq. (A8) by introducing conjugate fields $\psi_{\mathbf{k}}$ for charge density and $w_{\mathbf{k}}$ for mass density. The partition function $\mathcal{Z}$ in Eq. (A1) can then be rewritten in terms of

$$\mathcal{Z}' = \int\prod_{\alpha=1}^{n_p}\prod_{\tau=1}^{N}d\mathbf{R}_{\alpha,\tau}\prod_{a=1}^{n_s+n_c}d\mathbf{r}_a e^{-\mathcal{T}[\mathbf{R}]-\mathcal{U}[\mathbf{R},\mathbf{r}]}$$

$$= \exp\left\{-\frac{1}{2\Omega}v_2|\langle 1_N|\rho_{\mathbf{k}=0}\rangle|^2\right\}\prod_{\mathbf{k}\neq 0}\int\frac{d\psi_{\mathbf{k}}dw_{\mathbf{k}}}{2\pi\Omega\sqrt{\lambda_k v_2}}$$

$$\times\exp\left\{-\frac{1}{2\Omega}\sum_{\mathbf{k}\neq 0}\left[\frac{|\psi_{\mathbf{k}}|^2}{\lambda_k} + \frac{|w_{\mathbf{k}}|^2}{v_2}\right]\right\}\int\prod_{\alpha=1}^{n_p}\prod_{\tau=1}^{N}d\mathbf{R}_{\alpha,\tau}\prod_{a=1}^{n_s+n_c}d\mathbf{r}_a$$

$$\times\exp\left\{-\frac{i}{\Omega}\sum_{\mathbf{k}\neq 0}\left[(\langle\sigma|\rho_{-\mathbf{k}}\rangle + \langle z|c_{-\mathbf{k}}\rangle)\psi_{\mathbf{k}} + \langle 1_N|\rho_{-\mathbf{k}}\rangle w_{\mathbf{k}}\right] - \mathcal{T}[\{\mathbf{R}_{\alpha,\tau}\}]\right\}, \tag{A9}$$

where $\mathcal{Z} = \mathcal{Z}'/(n_p!n_c!n_s!n_w!)$. The first term in $\mathcal{Z}'$ is merely the $\mathbf{k} = 0$ component of $\mathcal{U}$, which by the definition of $\rho_{\mathbf{k}}^\tau$ is equal to

$$\mathcal{Z}_0 \equiv \exp\left\{-\frac{1}{2\Omega}v_2|\langle 1_N|\rho_{\mathbf{k}=0}\rangle|^2\right\} = \exp\left\{-\frac{v_2(Nn_p)^2}{2\Omega}\right\}. \tag{A10}$$

The remaining terms in $\mathcal{Z}'$ is a field integral of ψ and w. The first component [the first part of the third line in Eq. (A9)] is an

exponential of the quadratic self-correlations, and the second term [the fourth line and the latter part of the third line in Eq. (A9)] is a partition function for the polymers and the small ions under the influence of ψ and w, which we now symbolize as

$$\mathcal{Q}_{\text{sol}}[\psi, w] \equiv \int\prod_{\alpha=1}^{n_p}\prod_{\tau=1}^{N}d\mathbf{R}_{\alpha,\tau}\prod_{a=1}^{n_s+n_c}d\mathbf{r}_a\exp\left\{-\frac{i}{\Omega}\sum_{\mathbf{k}\neq 0}\left[(\langle\sigma|\rho_{-\mathbf{k}}\rangle + \langle z|c_{-\mathbf{k}}\rangle)\psi_{\mathbf{k}}\right.\right.$$

$$\left.\left. + \langle 1_N|\rho_{-\mathbf{k}}\rangle w_{\mathbf{k}}\right] - \mathcal{T}[\{\mathbf{R}_{\alpha,\tau}\}]\right\}. \tag{A11}$$

By the definitions of $c_{\mathbf{k}}$ and $\rho_{\mathbf{k}}^\tau$ in Eq. (A4), the exponent in the integrand of $\mathcal{Q}_{\text{sol}}$ may be expressed as

$$-\frac{i}{\Omega}\sum_{\mathbf{k}\neq 0}\left[(\langle\sigma|\rho_{-\mathbf{k}}\rangle + \langle z|c_{-\mathbf{k}}\rangle)\psi_{\mathbf{k}} + \langle 1_N|\rho_{-\mathbf{k}}\rangle w_{\mathbf{k}}\right] - \mathcal{T}[\{\mathbf{R}_{\alpha,\tau}\}]$$

$$= -\frac{i}{\Omega}\sum_{\mathbf{k}\neq 0}\psi_{\mathbf{k}}\left[(|z\rangle)_s\sum_{i=a}^{n_s}e^{-i\mathbf{k}\cdot\mathbf{r}_a} + (|z\rangle)_c\sum_{a=n_s+1}^{n_s+n_c}e^{-i\mathbf{k}\cdot\mathbf{r}_a}\right]$$

$$- \sum_{\alpha=1}^{n_p}\left[\frac{3}{2l^2}\sum_{\tau=1}^{N-1}(\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2\right.$$

$$\left. + \frac{i}{\Omega}\sum_{\mathbf{k}\neq 0}\sum_{\tau=1}^{N}(\sigma_\tau\psi_{\mathbf{k}} + w_{\mathbf{k}})e^{-i\mathbf{k}\cdot\mathbf{R}_{\alpha,\tau}}\right], \tag{A12}$$

where $|z\rangle_s = \text{sign}(q_c)z_s$ for salt ions and $|z\rangle_c = -\text{sign}(q_c)z_c$ for counterions as defined above. The coordinates of individual small ions and polymers are decoupled in this expression. Thus, the coordinate integrals in $\mathcal{Q}_{\text{sol}}$ are also decoupled, allowing it to be written as

$$\mathcal{Q}_{\text{sol}}[\psi, w] = (\mathcal{Q}_s[\psi])^{n_s}(\mathcal{Q}_c[\psi])^{n_c}(\mathcal{Q}_p[\psi, w])^{n_p}, \tag{A13}$$

where the n_s, n_c, and n_p superscripts are powers, with $\mathcal{Q}_s$ and $\mathcal{Q}_c$ being the single-molecule partition functions for salt ions and counterions, respectively; $[\psi]$ is shorthand for $[\{\psi_{\mathbf{k}}\}]$, and $[\psi, w]$ is shorthand for $[\{\psi_{\mathbf{k}}\}, \{w_{\mathbf{k}}\}]$. These single-molecule small-ion partition functions are given by

$$\mathcal{Q}_{s,c}[\psi] = \int d\mathbf{r}_{s,c}\exp\left\{-\frac{i(|z\rangle)_{s,c}}{\Omega}\sum_{\mathbf{k}\neq 0}\psi_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{r}_{s,c}}\right\}, \tag{A14}$$

where the expression for $\mathcal{Q}_s$ or $\mathcal{Q}_c$ corresponds, respectively, to choosing the subscript "s" or "c" for the "s, c" notation in Eq. (A14). The single-polymer partition function $\mathcal{Q}_p$ in Eq. (A13) equals

$$\mathcal{Q}_p[\psi, w] = \int\mathscr{D}[\mathbf{R}]e^{-\mathcal{H}_p[\mathbf{R};\psi,w]}, \tag{A15}$$

where $\mathscr{D}[\mathbf{R}] \equiv \prod_{\tau=1}^{N}d\mathbf{R}_\tau$, $[\mathbf{R}; \psi, w]$ is shorthand for $[\{\mathbf{R}_\tau\}, \{\psi_{\mathbf{k}}\}, \{w_{\mathbf{k}}\}]$, and

$$\mathcal{H}_p[\mathbf{R}; \psi, w] = \frac{3}{2l^2}\sum_{\tau=1}^{N-1}(\mathbf{R}_{\tau+1} - \mathbf{R}_\tau)^2 + \frac{i}{\Omega}\sum_{\mathbf{k}\neq 0}\sum_{\tau=1}^{N}(\sigma_\tau\psi_{\mathbf{k}} + w_{\mathbf{k}})e^{-i\mathbf{k}\cdot\mathbf{R}_\tau}. \tag{A16}$$

It should be noted that the small-ion label a and the polymer label α are not needed in the single-molecule partition functions in

Eqs. (A14) and (A15). Collecting results from Eqs. (A9), (A10), and (A13) yield the following formula for $\mathcal{Z}'$:

$$\mathcal{Z}' = \mathcal{Z}_0 \int \prod_{\mathbf{k} \neq \mathbf{0}} \frac{d\psi_{\mathbf{k}} dw_{\mathbf{k}}}{2\pi\Omega\sqrt{\lambda_k v_2}} \exp\left\{ -\frac{1}{2\Omega} \sum_{\mathbf{k} \neq \mathbf{0}} \left[\frac{|\psi_{\mathbf{k}}|^2}{\lambda_k} + \frac{|w_{\mathbf{k}}|^2}{v_2} \right] \right.$$
$$\left. + n_s \ln \mathcal{Q}_s + n_c \ln \mathcal{Q}_c + n_p \ln \mathcal{Q}_p \right\}, \tag{A17}$$

where $\mathcal{Z}_0$ is provided by Eq. (A10) and $\mathcal{Q}_s$, $\mathcal{Q}_c$, and $\mathcal{Q}_p$ are given by Eqs. (A14)–(A16).

2. Fluctuation expansion of partition function

To evaluate Eq. (A17) analytically, we first derive a mean-field solution at $(\psi, w) = (\overline{\psi}, \overline{w})$ in which the mean conjugated fields $\overline{\psi}$ and $\overline{w}$ satisfy the extremum condition $(\delta\mathcal{Z}'/\delta\psi_{\mathbf{k}}) = (\delta\mathcal{Z}'/\delta w_{\mathbf{k}}) = 0$, which leads to the equalities

$$\frac{\overline{\psi}_{\mathbf{k}}}{\Omega\lambda_k} = \frac{n_s}{\mathcal{Q}_s} \left(\frac{\delta\mathcal{Q}_s}{\delta\psi_{\mathbf{k}}} \right)_{(\overline{\psi},\overline{w})} + \frac{n_c}{\mathcal{Q}_c} \left(\frac{\delta\mathcal{Q}_c}{\delta\psi_{\mathbf{k}}} \right)_{(\overline{\psi},\overline{w})} + \frac{n_p}{\mathcal{Q}_p} \left(\frac{\delta\mathcal{Q}_p}{\delta\psi_{\mathbf{k}}} \right)_{(\overline{\psi},\overline{w})}, \tag{A18a}$$

$$\frac{\overline{w}_{\mathbf{k}}}{\Omega v_2} = \frac{n_p}{\mathcal{Q}_p} \left(\frac{\delta\mathcal{Q}_p}{\delta w_{\mathbf{k}}} \right)_{(\overline{\psi},\overline{w})}, \tag{A18b}$$

where the subscript $(\overline{\psi}, \overline{w})$ indicates that the functional (field) derivatives are evaluated at the to-be-solved mean conjugated fields. The ψ and w field are conjugates, respectively, to charge density and mass density. By using Eqs. (A14)–(A16) and the fact that the averages $\langle \cdots \rangle_t$ over the spatial coordinates of the given molecular species ($t = p$, s, or c) of $\mathbf{k}$-space density operators in Eq. (A4) are given by $\langle \rho_{\mathbf{k}}^\tau \rangle_p = n_p \langle e^{i\mathbf{k}\cdot\mathbf{R}_\tau} \rangle_p$, $\langle c_{\mathbf{k}}^s \rangle_s = n_s \langle e^{i\mathbf{k}\cdot\mathbf{r}_s} \rangle_s$, and $\langle c_{\mathbf{k}}^c \rangle_c = n_c \langle e^{i\mathbf{k}\cdot\mathbf{r}_c} \rangle_c$ because of the decoupling stated above by Eq. (A13), the first-order derivatives in Eq. (A18) are given by

$$\frac{n_{s,c}}{\mathcal{Q}_{s,c}} \frac{\delta\mathcal{Q}_{s,c}}{\delta\psi_{\mathbf{k}}} = -\frac{i(|z\rangle)_{s,c} n_{s,c}}{\Omega} \left(\langle e^{-i\mathbf{k}\cdot\mathbf{r}_{s,c}} \rangle_{s,c} \right)_{(\psi,w)} = -\frac{i(|z\rangle)_{s,c}}{\Omega} \left(\langle c_{-\mathbf{k}}^{s,c} \rangle_{s,c} \right)_{(\psi,w)}, \tag{A19a}$$

$$\frac{n_p}{\mathcal{Q}_p} \frac{\delta\mathcal{Q}_p}{\delta\psi_{\mathbf{k}}} = -\frac{in_p}{\Omega} \left(\left\langle \sum_{\tau=1}^{N} \sigma_\tau e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} \right\rangle_p \right)_{(\psi,w)} = -\frac{i}{\Omega} \sum_{\tau=1}^{N} \sigma_\tau \left(\langle \rho_{-\mathbf{k}}^\tau \rangle_p \right)_{(\psi,w)}, \tag{A19b}$$

$$\frac{n_p}{\mathcal{Q}_p} \frac{\delta\mathcal{Q}_p}{\delta w_{\mathbf{k}}} = -\frac{in_p}{\Omega} \left(\left\langle \sum_{\tau=1}^{N} e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} \right\rangle_p \right)_{(\psi,w)} = -\frac{i}{\Omega} \sum_{\tau=1}^{N} \left(\langle \rho_{-\mathbf{k}}^\tau \rangle_p \right)_{(\psi,w)}, \tag{A19c}$$

where $((\cdots)_t)_{(\psi,w)}$ denotes averaging over the spatial coordinates of the given molecular species evaluated for any given conjugate field ψ, w. With Eq. (A19), the relations in Eq. (A18) for the mean conjugate fields become

$$\overline{\psi}_{\mathbf{k}} = -i\lambda_k \left(\left\langle \left[\langle\sigma|\rho_{-\mathbf{k}}\rangle + \langle z|c_{-\mathbf{k}}\rangle \right] \right\rangle_{s,c,p} \right)_{(\overline{\psi},\overline{w})},$$
$$\overline{w}_{\mathbf{k}} = -iv_2 \left(\left\langle \left[\langle 1_N|\rho_{-\mathbf{k}}\rangle \right] \right\rangle_p \right)_{(\overline{\psi},\overline{w})}, \tag{A20}$$

which can now be solved self-consistently to determine $\overline{\psi}_{\mathbf{k}}$ and $\overline{w}_{\mathbf{k}}$.

We proceed to obtain an approximate solution by assuming that within regions where the system exists as a single phase, the mass density is rather homogeneous. In that case, the $\mathbf{k} \neq \mathbf{0}$ components of the density operators $\rho_{\mathbf{k}}^\tau$, $c_{\mathbf{k}}^s$, and $c_{\mathbf{k}}^c$ in Eq. (A4) are small (approximately zero). It then follows from Eq. (A20) that

$$\overline{\psi}_{\mathbf{k}} \approx \overline{w}_{\mathbf{k}} \approx 0, \quad \forall \mathbf{k} \neq \mathbf{0}. \tag{A21}$$

These considerations imply that the following approximate relations hold for the averaged densities on the right-hand side of Eq. (A19),

$$\langle c_{-\mathbf{k}}^{s,c} \rangle_{\approx 0} \approx n_{s,c}\delta_{\mathbf{k},\mathbf{0}}, \qquad \sum_{\tau=1}^{N} \sigma_\tau \langle \rho_{-\mathbf{k}}^\tau \rangle_{\approx 0} \approx q_c n_p N \delta_{\mathbf{k},\mathbf{0}},$$
$$\sum_{\tau=1}^{N} \langle \rho_{-\mathbf{k}}^\tau \rangle_{\approx 0} \approx n_p N \delta_{\mathbf{k},\mathbf{0}}, \tag{A22}$$

where the "≈ 0" subscript in $\langle \cdots \rangle_{\approx 0}$ signifies that the given average over the s, c, or p spatial coordinates is evaluated at the conjugate fields in Eq. (A21) for approximate homogeneous densities. Now, to arrive at a definite approximate description, we expand the logarithmic small-ion partition functions around $\psi_{\mathbf{k}\neq\mathbf{0}} = 0$ up to $O(\delta\psi^2)$. Utilizing the expressions for the averaged densities in Eq. (A22) and replacing the conjugate field $\overline{\psi}_{\mathbf{k}\neq\mathbf{0}} \approx 0$ [Eq. (A21)] at which the averages are evaluated by $\psi_{\mathbf{k}\neq\mathbf{0}} = 0$, we obtain

$$\ln \mathcal{Q}_{s,c}[\psi] \approx \ln \mathcal{Q}_{s,c}[\psi_{\mathbf{k}\neq\mathbf{0}} = 0] + \sum_{\mathbf{k}\neq\mathbf{0}} \left(\frac{\delta \ln \mathcal{Q}_{s,c}}{\delta\psi_{\mathbf{k}}} \right)_0 \delta\psi_{\mathbf{k}}$$
$$+ \frac{1}{2} \sum_{\mathbf{k},\mathbf{k}'\neq\mathbf{0}} \left(\frac{\delta^2 \ln \mathcal{Q}_{s,c}}{\delta\psi_{\mathbf{k}} \delta\psi_{\mathbf{k}'}} \right)_0 \delta\psi_{\mathbf{k}} \delta\psi_{\mathbf{k}'}$$
$$= \ln \Omega - \frac{i\langle z \rangle_{s,c}}{\Omega} \sum_{\mathbf{k}\neq\mathbf{0}} \langle e^{-i\mathbf{k}\cdot\mathbf{r}_{s,c}} \rangle_0 \delta\psi_{\mathbf{k}} - \frac{z_{s,c}^2}{2\Omega^2} \sum_{\mathbf{k},\mathbf{k}'\neq\mathbf{0}} \left[\langle e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}_{s,c}} \rangle_0 \right.$$
$$\left. - \langle e^{-i\mathbf{k}\cdot\mathbf{r}_{s,c}} \rangle_0 \langle e^{-i\mathbf{k}'\cdot\mathbf{r}_{s,c}} \rangle_0 \right] \delta\psi_{\mathbf{k}} \delta\psi_{\mathbf{k}'}$$
$$= \ln \Omega - \frac{z_{s,c}^2}{2\Omega^2} \sum_{\mathbf{k}\neq\mathbf{0}} |\psi_{\mathbf{k}}|^2, \tag{A23}$$

where the "0" subscript in $(\cdots)_0$ indicates that the derivatives are evaluated at $\psi_{\mathbf{k}\neq\mathbf{0}} = 0$. Similarly, replacing the "≈ 0" subscripts in Eq. (A22), here the "0" subscript in $\langle \cdots \rangle_0$ indicates that the average is evaluated at $\psi_{\mathbf{k}\neq\mathbf{0}} = 0$. In the last line of Eq. (A23), the expansion variable $\delta\psi_{\mathbf{k}}$ is written as $\psi_{\mathbf{k}}$ for every term in the $\sum_{\mathbf{k}\neq\mathbf{0}}$ summation because the expansion is around $\psi_{\mathbf{k}\neq\mathbf{0}} = 0$. Substituting Eq. (A23) for $\ln \mathcal{Q}_s$ and $\ln \mathcal{Q}_c$ into Eq. (A17) yields

$$\mathcal{Z}' \approx \mathcal{Z}_0 \int \prod_{\mathbf{k}\neq\mathbf{0}} \frac{d\psi_{\mathbf{k}} dw_{\mathbf{k}}}{2\pi\Omega\sqrt{\lambda_k v_2}} \exp\left\{ -\frac{1}{2\Omega} \sum_{\mathbf{k}\neq\mathbf{0}} \left[|\psi_{\mathbf{k}}|^2 \left(\frac{1}{\lambda_k} + z_s^2 \rho_s + z_c^2 \rho_c \right) \right. \right.$$
$$\left. \left. + \frac{|w_{\mathbf{k}}|^2}{v_2} \right] + n_p \ln \mathcal{Q}_p + C \right\}, \tag{A24}$$

where $C = (n_s + n_c) \ln \Omega$ will be dropped in subsequent consideration because it has no effect on the relative free energies of different configurational states. Let the exponent in Eq. (A24) without C be

denoted as $-\mathscr{H}$, then $\mathscr{H}$ may be seen as a Hamiltonian of a polymer system,

$$\mathscr{H}[\psi, w] = \frac{1}{2\Omega} \sum_{\mathbf{k} \neq \mathbf{0}} \left[v_k \psi_{-\mathbf{k}} \psi_{\mathbf{k}} + \frac{w_{-\mathbf{k}} w_{\mathbf{k}}}{v_2} \right] - n_p \ln \mathcal{Q}_p[\psi, w], \quad (A25)$$

where

$$\frac{1}{v_k} = \frac{1}{1/\lambda_k + z_s^2 \rho_s + z_c^2 \rho_c} \equiv \frac{4\pi l_B}{k^2 + \kappa^2} \quad (A26)$$

is merely a Fourier-transformed Coulomb potential with screening length $1/\kappa = [4\pi l_B (z_s^2 \rho_s + z_c^2 \rho_c)]^{-1/2}$. We may now express $\mathcal{Z}'$ as a product of three components, viz.,

$$\mathcal{Z}' = \mathcal{Z}_0 \mathcal{Z}_{\mathrm{ion}} \mathcal{Z}_p', \quad (A27)$$

where $\mathcal{Z}_0$ is defined in Eq. (A10),

$$\mathcal{Z}_{\mathrm{ion}} = \prod_{\mathbf{k} \neq \mathbf{0}} \frac{1}{\sqrt{v_k \lambda_k}} = \prod_{\mathbf{k} \neq \mathbf{0}} \left[1 + \frac{\kappa^2}{k^2} \right]^{-\frac{1}{2}}, \quad (A28)$$

and

$$\mathcal{Z}_p' = \prod_{\mathbf{k} \neq \mathbf{0}} \int \sqrt{\frac{v_k}{v_2}} \frac{d\psi_{\mathbf{k}} dw_{\mathbf{k}}}{2\pi\Omega} e^{-\mathscr{H}[\psi, w]}. \quad (A29)$$

Accordingly, the complete partition function $\mathcal{Z} = \mathcal{Z}'/(n_s! n_c! n_p! n_w!)$ provides free energy of the system in units $k_B T$ per volume l^3,

$$f = -\frac{l^3}{\Omega} \ln \mathcal{Z} = -s + f_{\mathrm{ion}} + f_P + f_0, \quad (A30)$$

where

$$-s = \frac{l^3}{\Omega} \ln(n_s! n_c! n_p! n_w!), \quad (A31)$$

$$f_0 = -\frac{l^3}{\Omega} \ln \mathcal{Z}_0 = \frac{v_2 l^3 (n_p N)^2}{2\Omega^2} = \frac{l^3}{2} v_2 \rho_m^2, \quad (A32)$$

$$f_{\mathrm{ion}} = -\frac{l^3}{\Omega} \ln \mathcal{Z}_{\mathrm{ion}} = \frac{l^3}{2} \sum_{\mathbf{k} \neq \mathbf{0}} \ln\left[1 + \frac{\kappa^2}{k^2} \right] = -\frac{(\kappa l)^3}{12\pi} + I_0, \quad (A33)$$

$$f_P = -\frac{l^3}{\Omega} \ln \mathcal{Z}_p'. \quad (A34)$$

3. Small-ion free energy

The first term of f_{ion} in Eq. (A33) is the standard Debye screening energy. The second term of f_{ion}, $I_0 = l^3 \kappa^2 k_{\max}$, is formally divergent ($k_{\max}$ is the maximum k value of the system, corresponding to the smallest length scale in coordinate space; $I_0 \to \infty$ as $k_{\max} \to \infty$), but since it is linearly proportional to n_s and n_c (through its dependence on κ^2; see above), this formally divergent term is irrelevant to the relative free energies of different configurational states of the system.[59] As in most analyses, the $\mathbf{k}$-summation is performed here by replacing it with a continuous integral over $\mathbf{k}$-space,

$$\frac{1}{\Omega} \sum_{\mathbf{k} \neq \mathbf{0}} \to \int \frac{d^3 k}{(2\pi)^3}. \quad (A35)$$

To make our model physically more realistic, however, we follow Muthukumar[73,94] who treated the charge of each small ion as distributed over a finite volume with a characteristic length scale comparable to the bare Kuhn length l of the polymers. In this treatment, the point-charge expression for f_{ion} in Eq. (A33) is replaced by

$$f_{\mathrm{ion}} = -\frac{1}{4\pi} \left[\ln(1 + \kappa l) - \kappa l + \frac{1}{2} (\kappa l)^2 \right], \quad (A36)$$

which reduces to $-(\kappa l)^3/(12\pi)$ in Eq. (A33), as it should, in the limit of $\kappa l \to 0$. In this regard, Eq. (A36)—which is used for all rG-RPA and fG-RPA applications in the present work—may be viewed as a regularized, more physical version of Eq. (A33).

4. Polymer free energy

We now proceed to derive an approximate, tractable analytical expression for $\mathcal{Z}_p'$ in Eq. (7) in the main text and Eq. (A29) by expanding $\ln \mathcal{Q}_p$ [defined in Eqs. (A15) and (A16)] around $\psi_{\mathbf{k} \neq \mathbf{0}} = w_{\mathbf{k} \neq \mathbf{0}} = 0$, viz.,

$$\ln \mathcal{Q}_p[\psi, w] = \ln \mathcal{Q}_p[\psi_{\mathbf{k} \neq \mathbf{0}} = w_{\mathbf{k} \neq \mathbf{0}} = 0] + \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{\tau=1}^{N} \left(\frac{\delta \ln \mathcal{Q}_p}{\delta \varphi_{\mathbf{k}}^\tau} \right)_0 \varphi_{\mathbf{k}}^\tau$$

$$+ \frac{1}{2} \sum_{\mathbf{k}, \mathbf{k}' \neq \mathbf{0}} \sum_{\tau, \mu=1}^{N} \left(\frac{\delta^2 \ln \mathcal{Q}_p}{\delta \varphi_{\mathbf{k}}^\tau \delta \varphi_{\mathbf{k}'}^\mu} \right)_0 \varphi_{\mathbf{k}}^\tau \varphi_{\mathbf{k}'}^\mu + O(\varphi^3)$$

$$= \ln \Omega + \frac{3(N-1)}{2} \ln\left(\frac{2\pi l^2}{3} \right) - \frac{1}{2\Omega^2} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{\tau, \mu=1}^{N}$$

$$\times \left\langle e^{-i\mathbf{k} \cdot (\mathbf{R}_\tau - \mathbf{R}_\mu)} \right\rangle_0 \varphi_{\mathbf{k}}^\tau \varphi_{-\mathbf{k}}^\mu + O(\varphi^3), \quad (A37)$$

where $\varphi_{\mathbf{k}}^\tau = \sigma_\tau \psi_{\mathbf{k}} + w_{\mathbf{k}}$ and the second term in the first line vanishes because of Eq. (A22). As in Eq. (A23), the first two constant terms in the third line of the above equation have no effect on the relative energies of different configurations of the system and therefore will be discarded for our present purpose. The $\langle \cdots \rangle_0$ in the last line of Eq. (A37) is the intrachain monomer-monomer correlation function evaluated at $\psi_{\mathbf{k} \neq \mathbf{0}} = w_{\mathbf{k} \neq \mathbf{0}} = 0$. This correlation function is equal to that of a Gaussian chain. However, in the presence of intrachain and interchain interactions, a Gaussian-chain description of the polymer chains in our system is unsatisfactory as has been demonstrated by theoretical and experimental studies[63,95–97] showing that polymers with different net charges and heteropolymers with different charge sequences—even when they have the same net charge—can have dramatically different conformational characteristics. Intuitively, this sequence-dependent conformational heterogeneity should apply not only to the case when a polymer chain is isolated but also to situations in which polymer chains are in semidilute solutions. To account for this fundamental property in the monomer-monomer correlation function, we need to include nonzero $\psi_{\mathbf{k} \neq \mathbf{0}}$ and $w_{\mathbf{k} \neq \mathbf{0}}$ fluctuations that arise from the higher-order terms in Eq. (A37). Accordingly, based on a rationale similar to that advanced in Refs. 59, 71, and 72, we replace the monomer-monomer correlation function in Eq. (A37) by

a correlation function involving *arbitrary* fields. This development leads to

$$\ln \mathcal{Q}_p[\psi, w] \simeq -\frac{N}{2\Omega^2} \sum_{\mathbf{k}\neq\mathbf{0}} [\xi_{\mathbf{k}}\psi_{\mathbf{k}}\psi_{-\mathbf{k}} + g_{\mathbf{k}}w_{\mathbf{k}}w_{-\mathbf{k}} + 2\zeta_{\mathbf{k}}w_{\mathbf{k}}\psi_{-\mathbf{k}}], \quad (A38)$$

where ξ, g, and ζ are structure factors of mass and charge densities,

$$\xi_{\mathbf{k}} = \frac{1}{N} \sum_{\tau,\mu=1}^{N} \sigma_\tau \sigma_\mu \left(\left\langle e^{i\mathbf{k}\cdot(\mathbf{R}_\tau - \mathbf{R}_\mu)} \right\rangle_p \right)_{(\psi,w)}, \quad (A39a)$$

$$g_{\mathbf{k}} = \frac{1}{N} \sum_{\tau,\mu=1}^{N} \left(\left\langle e^{i\mathbf{k}\cdot(\mathbf{R}_\tau - \mathbf{R}_\mu)} \right\rangle_p \right)_{(\psi,w)}, \quad (A39b)$$

$$\zeta_{\mathbf{k}} = \frac{1}{N} \sum_{\tau,\mu=1}^{N} \sigma_\tau \left(\left\langle e^{i\mathbf{k}\cdot(\mathbf{R}_\tau - \mathbf{R}_\mu)} \right\rangle_p \right)_{(\psi,w)}. \quad (A39c)$$

Substituting Eq. (A38) for $\ln \mathcal{Q}_p$ in Eq. (A25), we obtain

$$\mathcal{H}[\psi, w] = \frac{1}{2\Omega} \sum_{\mathbf{k}\neq\mathbf{0}} \langle \psi_{-\mathbf{k}} \ w_{-\mathbf{k}} | \begin{pmatrix} v_k + \rho_m\xi_{\mathbf{k}} & \rho_m\zeta_{\mathbf{k}} \\ \rho_m\zeta_{\mathbf{k}} & v_2^{-1} + \rho_m g_{\mathbf{k}} \end{pmatrix} \begin{vmatrix} \psi_{\mathbf{k}} \\ w_{\mathbf{k}} \end{vmatrix} \rangle$$

$$= \frac{1}{2\Omega} \sum_{\mathbf{k}\neq\mathbf{0}} \langle \Psi_{-\mathbf{k}} | \hat{\Delta}_{\mathbf{k}} | \Psi_{\mathbf{k}} \rangle, \quad (A40)$$

where $\langle \Psi_{-\mathbf{k}} | \equiv \langle \psi_{-\mathbf{k}} \ w_{-\mathbf{k}} |$, $|\Psi_{\mathbf{k}}\rangle = ((\langle \Psi_{-\mathbf{k}}|)^{*T}$, and $\hat{\Delta}_{\mathbf{k}}$ is the 2×2 matrix in the above equation. Thus, each term in the product given in Eq. (A29) can now be evaluated as a Gaussian integral to yield

$$\mathcal{Z}_p' = \prod_{\mathbf{k}\neq\mathbf{0}} \sqrt{\frac{v_k}{v_2 \det \hat{\Delta}_{\mathbf{k}}}}. \quad (A41)$$

Therefore, by Eqs. (A35) and (A41), the unit free energy is now formally given by

$$f_{\mathrm{P}} = -\frac{l^3 \ln \mathcal{Z}_p'}{\Omega} = \frac{l^3}{2} \int \frac{d^3k}{(2\pi)^3}$$

$$\times \ln\left[1 + \rho_m\left(\frac{\xi_{\mathbf{k}}}{v_k} + v_2 g_{\mathbf{k}}\right) + \frac{v_2}{v_k}\rho_m^2\left(\xi_{\mathbf{k}}g_{\mathbf{k}} - \zeta_{\mathbf{k}}^2\right) \right]. \quad (A42)$$

It should be noted, however, that the $k \equiv |\mathbf{k}| \to \infty$ behavior of the integrand in the above Eq. (A42) needs to be regularized. For point particles, the $k \to \infty$ limit of the pairwise correlation function is a Kronecker-δ,

$$\lim_{k\to\infty} \left\langle e^{i\mathbf{k}\cdot(\mathbf{R}_\tau - \mathbf{R}_\mu)} \right\rangle_p = \delta_{\tau\mu}. \quad (A43)$$

Thus, by Eq. (A39),

$$\lim_{k\to\infty} \xi_{\mathbf{k}} = \frac{1}{N} \sum_{\tau=1}^{N} \sigma_\tau^2, \quad (A44a)$$

$$\lim_{k\to\infty} g_{\mathbf{k}} = 1, \quad (A44b)$$

$$\lim_{k\to\infty} \zeta_{\mathbf{k}} = q_c. \quad (A44c)$$

Because $\lim_{k\to\infty}(1/v_k) = \lim_{k\to\infty} 4\pi l_B/k^2$ and $v_2 > 0$, Eq. (A44) indicates that the integral in Eq. (A42) has an ultraviolet (large-k) divergence. This divergence is physically irrelevant, however, because the

integral can be readily regularized by subtracting the unphysical Coulomb self-energy of the charged monomers,

$$f_{\mathrm{self}} = \frac{\rho_m l^3}{2N} \int \frac{d^3k}{(2\pi)^3} \frac{4\pi l_B}{k^2} \sum_{\tau=1}^{N} \sigma_\tau^2, \quad (A45)$$

that was included merely for formulational convenience in the first place. In the same vein as the charge smearing for the small ions [Eq. (A36)], we also smear the δ-function excluded volume repulsion by a Gaussian,[98,99] viz.,

$$v_2 \to v_2(k) = v_2 e^{-\frac{1}{6}(kl)^2}, \quad (A46)$$

and use $v_2(k)$ in the integral of Eq. (A42) of f_{P} to give a v_2-regularized $f_{\mathrm{P}}[v_2(k)]$. The regularized f_{P} resulting from these two procedures is then given by

$$f_{\mathrm{P}}[v_2(k)] - f_{\mathrm{self}} \to f_{\mathrm{P}}, \quad (A47)$$

where the last arrow signifies that this regularized version of f_{P} is the one used for our subsequent theoretical development in the present work.

As discussed above, the present separate treatments for small ions [Eq. (A36)] and polymers [Eqs. (A42) and (A47)] are needed in our formulation—which expresses the total partition function as a product consisting of separate factors for small ions and polymers [Eq. (A28)]—such that the polymer part of the partition function can be used to derive an effective Kuhn length. Not surprisingly, in the event that the bare chain length l is used instead of an effective Kuhn length and that the volume of small ions and the volume of the monomers of the polymers becomes negligible ($v_2 \to 0$), the free energy expression reduces to that of our simple RPA theory[55,56] as can be readily seen in the following. First, when the size of the small ions is assumed to be negligible, their free energy is given by the simple Debye-Hückel expression in Eq. (A33) instead of the finite-size expression in Eq. (A36). Second, as $v_2 \to 0$, all terms involving v_2 in Eq. (A42) vanish. Consequently, the resulting overall electrostatic free energy, denoted here as f_{el}, is given by

$$f_{\mathrm{el}} = f_{\mathrm{ion}}^{[\mathrm{Eq.\ (A33)}]} + f_{\mathrm{P}}^{[\mathrm{Eq.\ (A42)}]}(v_2 \to 0)$$

$$= \frac{l^3}{2} \int \frac{d^3k}{(2\pi)^3} \left\{ \ln\left[1 + \frac{\kappa^2}{k^2}\right] + \ln\left[1 + \rho_m\frac{\xi_{\mathbf{k}}}{v_k}\right] \right\}. \quad (A48)$$

Recalling that $\kappa^2 = 4\pi l_B(z_s^2\rho_s + z_c^2\rho_c)$ and $1/v_k = 4\pi l_B/(k^2 + \kappa^2)$ [Eq. (A26)], this quantity becomes

$$f_{\mathrm{el}} = \frac{l^3}{2} \int \frac{d^3k}{(2\pi)^3} \ln\left[\frac{\kappa^2 + k^2}{k^2} \times \frac{k^2 + \kappa^2 + 4\pi l_B\rho_m\xi_{\mathbf{k}}}{k^2 + \kappa^2}\right]$$

$$= \frac{l^3}{2} \int \frac{d^3k}{(2\pi)^3} \ln\left[1 + \frac{\kappa^2 + 4\pi l_B\rho_m\xi_{\mathbf{k}}}{k^2}\right]$$

$$= \frac{l^3}{2} \int \frac{d^3k}{(2\pi)^3} \ln\left[1 + \frac{4\pi l_B}{k^2}\left(z_s^2\rho_s + z_c^2\rho_c^2 + \rho_m\xi_{\mathbf{k}}\right)\right], \quad (A49)$$

which is exactly the same f_{el} expression in our previous simple RPA theory in a formulation that does not consider an explicit excluded-volume repulsion term and treats small ions and polymers on the same footing.[55,56]

5. Effective Gaussian-chain model for two-body correlation function

The (ψ, w)-dependence of the structure factors ξ, g, and ζ in Eq. (A42) for f_P allows for an account of sequence-dependent conformational heterogeneity by using a Gaussian chain with a renormalized Kuhn length[68] $l_1 = xl$ (instead of the "bare" Kuhn length l) to approximate the polymer partition function $\mathcal{Q}_p$ in Eq. (A15). Specifically, we make the approximation that

$$\mathcal{Q}_p \approx \int \mathscr{D}[\mathbf{R}] e^{-\mathcal{H}_p^0[\mathbf{R}]}, \quad \text{where} \quad \mathcal{H}_p^0[\mathbf{R}] = \frac{3}{2l^2 x} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\tau+1} - \mathbf{R}_\tau)^2. \tag{A50}$$

The structure factors ξ, g, and ζ in Eq. (A39) can then be readily expressed in terms of the yet-to-be-determined renormalization parameter x,

$$\xi_{\mathbf{k}} \to \xi_k^x = \frac{1}{N} \sum_{\tau,\mu=1}^{N} \sigma_\tau \sigma_\mu e^{-\frac{1}{6}(kl)^2 x|\tau-\mu|}, \tag{A51a}$$

$$g_{\mathbf{k}} \to g_k^x = \frac{1}{N} \sum_{\tau,\mu=1}^{N} e^{-\frac{1}{6}(kl)^2 x|\tau-\mu|}, \tag{A51b}$$

$$\zeta_{\mathbf{k}} \to \zeta_k^x = \frac{1}{N} \sum_{\tau,\mu=1}^{N} \sigma_\tau e^{-\frac{1}{6}(kl)^2 x|\tau-\mu|}. \tag{A51c}$$

The renormalization parameter $x = l_1/l$ is determined using a sequence-specific variational approach introduced by Sawle and Ghosh[68,100] as follows. We first express the Hamiltonian $\mathcal{H}_p[\mathbf{R}]$ in Eq. (A16) as $\mathcal{H}_p = \mathcal{H}_p^0 + \mathcal{H}_p^1$, where $\mathcal{H}_p^0$ [given by Eq. (A50)] is the principal term and

$$\mathcal{H}_p^1[\mathbf{R}; \psi, w] = \frac{3}{2l^2}\left(1 - \frac{1}{x}\right) \sum_{\tau=1}^{N-1} (\mathbf{R}_{\tau+1} - \mathbf{R}_\tau)^2$$
$$+ \frac{i}{\Omega} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{\tau=1}^{N} (\sigma_\tau \psi_{\mathbf{k}} + w_{\mathbf{k}}) e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} \tag{A52}$$

is the perturbative term. Then, for any given physical quantity $A[\mathbf{R}]$, the perturbation expansion of its thermodynamic average over polymer configurations $\{\mathbf{R}_\tau\}$ and field fluctuations $\Psi = (\psi, w)$ is given by[101]

$$\langle A[\mathbf{R}] \rangle = \frac{\left\langle e^{-\mathcal{H}_p^1[\mathbf{R};\Psi]} A[\mathbf{R}] \right\rangle_{0,\Psi}}{\left\langle e^{-\mathcal{H}_p^1[\mathbf{R};\Psi]} \right\rangle_{0,\Psi}}$$

$$= \langle A[\mathbf{R}] \rangle_0 + \left[\langle A[\mathbf{R}] \rangle_0 \langle \mathcal{H}_p^1[\mathbf{R};\Psi] \rangle_{0,\Psi} - \langle A[\mathbf{R}] \mathcal{H}_p^1[\mathbf{R};\Psi] \rangle_{0,\Psi} \right]$$

$$+ \frac{1}{2}\left[\left\langle A[\mathbf{R}] \left(\mathcal{H}_p^1[\mathbf{R};\Psi]\right)^2 \right\rangle_{0,\Psi} - \langle A[\mathbf{R}] \rangle_0 \left\langle \left(\mathcal{H}_p^1[\mathbf{R};\Psi]\right)^2 \right\rangle_{0,\Psi} \right]$$

$$+ \langle A[\mathbf{R}] \rangle_0 \left\langle \mathcal{H}_p^1[\mathbf{R};\Psi] \right\rangle_{0,\Psi}^2 - \left\langle A[\mathbf{R}] \mathcal{H}_p^1[\mathbf{R};\Psi] \right\rangle_{0,\Psi}$$

$$\times \left\langle \mathcal{H}_p^1[\mathbf{R};\Psi] \right\rangle_{0,\Psi} + O\left(\left(\mathcal{H}_p^1\right)^3\right), \tag{A53}$$

where the subscripts $0, \Psi$ for $\langle \cdots \rangle$ signify, respectively, that the average over $\{\mathbf{R}_\tau\}$'s is weighted by the Hamiltonian $\mathcal{H}_p^0[\mathbf{R}]$ in Eq. (A50)

and the average over field configurations is weighted by the Hamiltonian $\mathscr{H}[\psi, w]$ in Eq. (A25). [Note that the meaning of the "0" subscript here is different from that for the averages evaluated at $\psi_{\mathbf{k}\neq\mathbf{0}} = w_{\mathbf{k}\neq\mathbf{0}} = 0$ in Eq. (A23)]. An $\mathcal{H}_p^0[\mathbf{R}]$ that provides a good description of the thermal properties of A may then be obtained by minimizing $\langle A \rangle - \langle A \rangle_0$. This is accomplished by a partial optimization to seek a value of $x = l_1/l$ that would abolish the lowest-order nontrivial $\mathcal{H}_p^1$ contributions in Eq. (A53).

To obtain a partially optimized $x = l_1/l$ that provides a good approximation for the monomer-monomer correlation function, A is chosen to be the squared end-to-end distance of the polymer, i.e., $A = R_{ee}^2 \equiv |\mathbf{R}_N - \mathbf{R}_1|^2$, because R_{ee} is a simple yet effective measure of conformational dimensions of polymers.[59,68] To facilitate this calculation, we express $\mathcal{H}_p^1$ in Eq. (A52) as $\mathcal{H}_p^1 = \mathcal{X}_1 + \mathcal{X}_2$, where

$$\mathcal{X}_1[\mathbf{R}] = \frac{3}{2l^2}\left(1 - \frac{1}{x}\right) \sum_{\tau=1}^{N-1} (\mathbf{R}_{\tau+1} - \mathbf{R}_\tau)^2, \tag{A54a}$$

$$\mathcal{X}_2[\mathbf{R}; \Psi] = \frac{i}{\Omega} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{\tau=1}^{N} (\sigma_\tau \psi_{\mathbf{k}} + w_{\mathbf{k}}) e^{-i\mathbf{k}\cdot\mathbf{R}_\tau}, \tag{A54b}$$

such that $\mathcal{X}_1[\mathbf{R}]$ is independent of Ψ and all of $\mathcal{H}_p^1$'s dependence on Ψ is contained in $\mathcal{X}_2[\mathbf{R}; \Psi]$. It follows that the Ψ average is trivial (i.e., it produces a multiplicative factor of unity and therefore can be omitted) for any function of $\mathcal{X}_1[\mathbf{R}]$ only. In Eq. (A53), the only contributions from terms linear in $\mathcal{X}_1[\mathbf{R}]$ come from the second line on the right-hand side (the first line after the second equality), which equal

$$\langle R_{ee}^2 \rangle_0 \langle \mathcal{X}_1 \rangle_0 - \langle R_{ee}^2 \mathcal{X}_1 \rangle_0 = -l^2(N-1)x(x-1). \tag{A55}$$

For the $\mathcal{X}_2$-containing terms in Eq. (A53), we first consider their Ψ-averages before applying the $\langle \cdots \rangle_0$ averaging. For terms linear in $\mathcal{X}_2$, it is straightforward to see that

$$\langle \mathcal{X}_2 \rangle_\Psi = \frac{i}{\Omega} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{\tau=1}^{N} \left[\sigma_\tau \langle \psi_{\mathbf{k}} \rangle_\Psi + \langle w_{\mathbf{k}} \rangle_\Psi \right] e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} = 0 \tag{A56}$$

because $\langle \psi_{\mathbf{k}} \rangle_\Psi = \langle w_{\mathbf{k}} \rangle_\Psi = 0$ according to the quadratic-field Hamiltonian $\mathscr{H}[\psi, w]$ in Eq. (A40). Thus, $\mathcal{X}_2$ has zero contribution in the first and third lines on the right-hand side after the second equality of Eq. (A53). In contrast, terms quadratic in $\mathcal{X}_2[\mathbf{R}]$ are not identical zero because

$$\langle \mathcal{X}_2^2 \rangle_\Psi = -\frac{1}{\Omega^2} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{\tau,\mu=1}^{N} \left[\sigma_\tau \sigma_\mu \langle \psi_{-\mathbf{k}} \psi_{\mathbf{k}} \rangle_\Psi + \langle w_{-\mathbf{k}} w_{\mathbf{k}} \rangle_\Psi \right.$$
$$\left. + (\sigma_\tau + \sigma_\mu) \langle \psi_{-\mathbf{k}} w_{\mathbf{k}} \rangle_\Psi \right] e^{-i\mathbf{k}\cdot(\mathbf{R}_\tau - \mathbf{R}_\mu)}, \tag{A57}$$

and here $\langle \mathcal{X}_2^2 \rangle_\Psi$ is seen as depending on field-field correlation functions $\langle \psi\psi \rangle$, $\langle ww \rangle$, and $\langle \psi w \rangle$ averaged over Ψ. Thus, the $\mathcal{X}_2^2$ factors in the averages in the second line on the right-hand side after the second equality of Eq. (A53) provide the only nonzero contribution through second order in $\mathcal{H}_p^1$. Following Ref. 59, we only consider lowest-order nonzero contributions from $\mathcal{X}_1$ and from $\mathcal{X}_2$, separately, i.e., including only terms through $O(\mathcal{X}_1)$ and $O(\mathcal{X}_2^2)$ as discussed above. This approach to the perturbative analysis of Eq. (A53) may also be rationalized by an alternate analytical formulation put forth in Refs. 71 and 72.

As shown in Eq. (A40), the field configuration distribution may be approximated by a Gaussian distribution embodied by the quadratic Hamiltonian $\mathscr{H}[\psi, w]$. According to perturbation theory,[56,102] the field-field correlation functions in Eq. (A57) can now be obtained from the matrix $\hat{\Delta}_{\mathbf{k}}$ in Eq. (A40) via the relationships

$$\frac{\langle \psi_{-\mathbf{k}} \psi_{\mathbf{k}} \rangle}{\Omega} = \left(\hat{\Delta}_{\mathbf{k}}^{-1}\right)_{11} = \frac{v_2^{-1} + \rho_m g_{\mathbf{k}}}{\det \hat{\Delta}_{\mathbf{k}}}, \tag{A58a}$$

$$\frac{\langle w_{-\mathbf{k}} w_{\mathbf{k}} \rangle}{\Omega} = \left(\hat{\Delta}_{\mathbf{k}}^{-1}\right)_{22} = \frac{v_k + \rho_m \xi_{\mathbf{k}}}{\det \hat{\Delta}_{\mathbf{k}}}, \tag{A58b}$$

$$\frac{\langle \psi_{-\mathbf{k}} w_{\mathbf{k}} \rangle}{\Omega} = \frac{\langle \psi_{\mathbf{k}} w_{-\mathbf{k}} \rangle}{\Omega} = \left(\hat{\Delta}_{\mathbf{k}}^{-1}\right)_{12} = \left(\hat{\Delta}_{\mathbf{k}}^{-1}\right)_{21} = \frac{-\rho_m \zeta_{\mathbf{k}}}{\det \hat{\Delta}_{\mathbf{k}}}. \tag{A58c}$$

Hence, $\langle \mathcal{X}_2^2 \rangle$ is expressed in terms of $\hat{\Delta}_{\mathbf{k}}$ as

$$\langle \mathcal{X}_2^2 \rangle = -\frac{1}{\Omega} \sum_{\mathbf{k} \neq \mathbf{0}} \sum_{\tau,\mu=1}^{N} \frac{\langle \sigma_\tau \; 1 | \begin{pmatrix} v_2^{-1} + \rho_m g_{\mathbf{k}} & -\rho_m \zeta_{\mathbf{k}} \\ -\rho_m \zeta_{\mathbf{k}} & v_k + \rho_m \xi_{\mathbf{k}} \end{pmatrix} | \begin{matrix} \sigma_\mu \\ 1 \end{matrix} \rangle}{\det \hat{\Delta}_{\mathbf{k}}} e^{-i\mathbf{k}\cdot(\mathbf{R}_\tau - \mathbf{R}_\mu)}. \tag{A59}$$

It should be noted that the excluded volume interaction v_2 is not regularized by Eq. (A46) here because a k-independent v_2 is needed to guarantee a real solution for the renormalization parameter x for arbitrary charge sequence $|\sigma\rangle$ (Refs. 68 and 69). Thus, the regularized form of v_2 in Eq. (A46) applies only to the explicit v_2 dependence of f_p in Eq. (A42) but not the implicit v_2 dependence of x contained in the renormalized form of the structure factors ξ, g, and ζ. Substituting the x-dependent correlation functions in Eq. (A51) for the structure factors in Eq. (A59), we obtain the nonzero contribution from $\mathcal{X}_2$ in the second line of the right-hand side after the second equality of Eq. (A53) as

$$\frac{1}{2}\left[\langle R_{ee}^2 \mathcal{X}_2^2 \rangle_0 - \langle R_{ee}^2 \rangle_0 \langle \mathcal{X}_2^2 \rangle_0 \right] = \frac{Nl^4 x^2}{18} \int \frac{d^3 k}{(2\pi)^3} \frac{k^2 \Xi_k^x}{\det \Delta_k^x}, \tag{A60}$$

where

$$\det \Delta_k^x = \frac{v_k}{v_2} + \rho_m \left(\frac{\xi_k^x}{v_2} + v_k g_k^x\right) + \rho_m^2 \left[\xi_k^x g_k^x - (\zeta_k^x)^2\right], \tag{A61}$$

and

$$\Xi_k^x \equiv \frac{\bar{\xi}_k^x}{v_2} + v_k \bar{g}_k^x + \rho_m\left(\bar{\xi}_k^x g_k^x + \xi_k^x \bar{g}_k^x - 2\zeta_k^x \bar{\zeta}_k^x\right). \tag{A62}$$

Here, the renormalized $\bar{\xi}$, $\bar{g}$, and $\bar{\zeta}$ are given by

$$\bar{\xi}_k^x = \frac{1}{N} \sum_{\tau,\mu=1}^{N} \sigma_\tau \sigma_\mu |\tau - \mu|^2 e^{-\frac{1}{6}(kl)^2 x |\tau - \mu|}, \tag{A63a}$$

$$\bar{g}_k^x = \frac{1}{N} \sum_{\tau,\mu=1}^{N} |\tau - \mu|^2 e^{-\frac{1}{6}(kl)^2 x |\tau - \mu|}, \tag{A63b}$$

$$\bar{\zeta}_k^x = \frac{1}{N} \sum_{\tau,\mu=1}^{N} \sigma_\tau |\tau - \mu|^2 e^{-\frac{1}{6}(kl)^2 x |\tau - \mu|}. \tag{A63c}$$

Finally, by combining Eqs. (A55) and (A60), we arrive at the variational equation

$$1 - \frac{1}{x} - \frac{Nl^2}{18(N-1)} \int \frac{d^3 k}{(2\pi)^3} \frac{k^2 \Xi_k^x}{\det \Delta_k^x} = 0 \tag{A64}$$

for solving x. In our numerical calculations, we take $v_2 = 4\pi l^3/3$. Inserting the solution of x into Eq. (A51) provides an improved accounting of the conformational heterogeneity in the free energy; and this improvement is central to the present rG-RPA theory.

6. Mixing entropy

The factorials in Eq. (A1) arise from the indistinguishability of the molecules belonging to the same species. Taking logarithm and using Stirling's approximation, one obtains

$$-\frac{S}{k_B} = \ln(n_p! n_s! n_c! n_w!)$$
$$\simeq n_p \ln n_p + n_s \ln n_s + n_c \ln n_c + n_w \ln n_w - n_p - n_c - n_s - n_w, \tag{A65}$$

where additive terms of the form $[\ln(2\pi n)]/2$ (where $n = n_p, n_s, n_c,$ or n_w) are omitted because for large n, their contributions is negligible in comparison with the terms included in Eq. (A65). As in Ref. 56, here we assume for simplicity that the size of a monomer, a small ion, or a water molecule all equals l^3. Assuming further, for simplicity, that the system is incompressible, i.e., the system volume Ω is fully occupied by polymers, small ions, and water, then

$$\frac{1}{\Omega}(Nn_p + n_s + n_c + n_w) = \rho_m + \rho_s + \rho_c + \rho_w = \frac{1}{l^3}. \tag{A66}$$

Following Flory's notation, volume fractions of polymers and salt ion are defined, respectively, as

$$\phi_m = \rho_m l^3, \quad \phi_s = \rho_s l^3, \tag{A67}$$

and the volume fraction ϕ_c of counterions and volume fraction ϕ_w of water are given by

$$z_c \phi_c = q_c \phi_m + z_s \phi_s, \quad \phi_w = 1 - \phi_m - \phi_s - \phi_c. \tag{A68}$$

Because the last four terms in Eq. (A65) are linear in numbers of molecules, they are irrelevant to phase separation.[56] Discarding these terms results in the mixing entropy

$$-s \equiv -\frac{Sl^3}{k_B \Omega} = \frac{\phi_m}{N} \ln \phi_m + \phi_s \ln \phi_s + \phi_c \ln \phi_c + \phi_w \ln \phi_w \tag{A69}$$

given in Eq. (2) of the main text.

APPENDIX B: TEMPERATURE SELECTION FOR POLYMER-SALT PHASE DIAGRAMS OF Ddx4 VARIANTS

Three temperatures, two below and one slightly above the respective salt-free critical temperature $T_{cr} \propto l/(l_B)_{cr}$ of each of the Ddx4 variants Ddx4$_{pH7}^{N1}$, Ddx4^{N1}CS$_{pH7}$, Ddx4$_{pH1}^{N1}$, and Ddx4^{N1}CS$_{pH1}$ are selected for the phase diagrams in Figs. 6–9. The l/l_B values are selected to compare salt dependence of the sequences under temperatures producing similar gaps between the dilute- and condensed-phase protein densities at or near $\phi_s = 0$ for the different sequences. Specifically, for the same part of the figures [(a), (b), and (c) separately], $l/(l_B)$'s are such that dilute-condensed density gaps are similar across Figs. 6–9.

REFERENCES

[1] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, and A. A. Hyman, Science **324**, 1729 (2009).

[2] P. Li, S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. X. Jiang, B. T. Nixon, and M. K. Rosen, Nature **483**, 2499 (2012).

[3] M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko, and S. L. McKnight, Cell **149**, 753 (2012).

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

[5] A. Molliex, J. Temirov, J. Lee, M. Coughlin, A. P. Kanagaraj, H. J. Kim, T. Mittag, and J. P. Taylor, Cell **163**, 123 (2015).

[6] C. W. Pak, M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, and M. K. Rosen, Mol. Cell **63**, 72 (2016).

[7] L. P. Bergeron-Sandoval, H. K. Heris, C. Chang, C. E. Cornell, S. L. Keller, A. G. Hendricks, A. J. Ehrlicher, P. Francois, R. V. Pappu, and S. W. Michnick, e-print bioRxiv:145664 (2018).

[8] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, Nature **547**, 236 (2017).

[9] A. J. Plys and R. E. Kingston, Science **361**, 329 (2018).

[10] W. K. Cho, J. H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and I. Cisse, Science **361**, 412 (2018).

[11] B. R. Sabari, A. Dall'Agnese, A. Boika, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, and R. A. Young, Science **361**, eaar3958 (2018).

[12] B. Tsang, J. Arsenault, R. M. Vernon, H. Lin, N. Sonenberg, L.-Y. Wang, A. Bah, and J. D. Forman-Kay, Proc. Natl. Acad. Sci. U. S. A. **116**, 4218 (2019).

[13] Y. Shin and C. P. Brangwynne, Science **357**, eaaf4382 (2017).

[14] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen, Nat. Rev. Mol. Cell Biol. **18**, 285 (2017).

[15] S. Boeynaems, S. Alberti, N. L. Fawzi, T. Mittag, M. Polymenidou, F. Rousseau, J. Schymkowitz, J. Shorter, B. Wolozin, L. Van Den Bosch, P. Tompa, and M. Fuxreiter, Trends Cell Biol **28**, 420 (2018).

[16] M. L. Broide, C. R. Berland, J. Pande, O. O. Ogun, and G. B. Benedek, Proc. Natl. Acad. Sci. U. S. A. **88**, 5660 (1991).

[17] N. Asherie, A. Lomakin, and G. B. Benedek, Phys. Rev. Lett. **77**, 4832 (1996).

[18] P. L. San Biagio, V. Martorana, A. Emanuele, S. M. Vaiana, M. Manno, D. Bulone, M. B. Palma-Vittorelli, and M. U. Palma, Proteins: Struct. Func. Bioinf. **37**, 211 (1999).

[19] H. X. Zhou and X. Pang, Chem. Rev. **118**, 1691 (2018).

[20] S. Qin and H.-X. Zhou, J. Phys. Chem. B **120**, 8164 (2016).

[21] S. Cinar, H. Cinar, H. S. Chan, and R. Winter, J. Am. Chem. Soc. **141**, 7347 (2019).

[22] J. D. Forman-Kay, R. W. Kriwacki, and G. Seydoux, J. Mol. Biol. **430**, 4603 (2018).

[23] H. Cinar, Z. Fetahaj, S. Cinar, R. M. Vernon, H. S. Chan, and R. H. A. Winter, Chem. - A Eur. J. **25**, 13049 (2019).

[24] J. P. Brady, P. J. Farber, A. Sekhar, Y.-H. Lin, R. Huang, A. Bah, T. J. Nott, H. S. Chan, A. J. Baldwin, J. D. Forman-Kay, and L. E. Kay, Proc. Natl. Acad. Sci. U. S. A. **114**, E8194 (2017).

[25] S. Alberti, J. Cell Sci. **130**, 2789 (2017).

[26] Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerye, R. O'Meally, G. L. Dignon, A. E. Conicella, W. Zheng, R. B. Best, R. N. Cole, J. Mittal, F. Shewmaker, and N. Fawzi, EMBO J. **36**, 2951 (2017).

[27] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, PLoS Comput Biol. **14**, e1005941 (2018).

[28] S. Das, A. N. Amin, Y.-H. Lin, and H. S. Chan, Phys. Chem. Chem. Phys. **20**, 28558 (2018).

[29] G. L. Dignon, W. Zheng, R. B. Best, Y. C. Kim, and J. Mittal, Proc. Natl. Acad. Sci. U. S. A. **115**, 9929 (2018).

[30] J. McCarty, K. T. Delaney, S. P. O. Danielsen, G. H. Fredrickson, and J. E. Shea, J. Phys. Chem. Lett. **10**, 1644 (2019).

[31] S. P. O. Danielsen, J. McCarty, J. E. Shea, K. T. Delaney, and G. H. Fredrickson, Proc. Natl. Acad. Sci. U. S. A. **116**, 8224 (2019).

[32] S. P. O. Danielsen, J. McCarty, J.-E. Shea, K. T. Dalaney, and G. H. Fredrickson, J. Chem. Phys. **151**, 034904 (2019).

[33] C. P. Brangwynne, P. Tompa, and R. Pappu, Nat. Phys. **11**, 899 (2015).

[34] Y.-H. Lin, J. D. Forman-Kay, and H. S. Chan, Biochemistry **57**, 2499 (2018).

[35] H. G. Bungenberg de Jong and H. R. Kruyt, Proc. K. Ned. Akad. Wet. **32**, 849 (1929).

[36] J. T. G. Overbeek and M. J. Voorn, J. Cell Comput. Physiol. **49**, 7 (1957).

[37] E. Spruijt, A. H. Westphal, J. W. Borst, M. A. C. Stuart, and J. van der Gucht, Macromolecules **43**, 6476 (2010).

[38] R. Chollakup, W. Smitthipong, C. D. Eisenbach, and M. Tirrell, Macromolecules **43**, 2518 (2010).

[39] S. L. Perry, Y. Li, D. Priftis, L. Leon, and M. Tirrell, Polymers **6**, 1756 (2014).

[40] S. L. Perry and C. E. Sing, Macromolecules **48**, 5040 (2015).

[41] S. Srivastava and M. V. Tirrell, Adv. Chem. Phys. **161**, 499 (2016).

[42] T. K. Lytle, M. Radhakrishna, and C. E. Sing, Macromolecules **49**, 9693 (2016).

[43] T. K. Lytle and C. E. Sing, Soft Matter **13**, 7001 (2017).

[44] M. Radhakrishna, K. Basu, Y. Liu, R. Shamsi, S. L. Perry, and C. E. Sing, Macromolecules **50**, 3030 (2017).

[45] P. Dubin and R. J. Stewart, Soft Matter **14**, 329 (2018).

[46] P. Zhang, K. Shen, N. M. Alsaifi, and Z.-G. Wang, Macromolecules **51**, 5586 (2018).

[47] S. Adhikari, M. A. Leaf, and M. Muthukumar, J. Chem. Phys. **149**, 163308 (2018).

[48] L. Li, S. Srivastava, M. Andreev, A. B. Marciel, J. J. D. Pablo, and M. V. Tirrell, Macromolecules **51**, 2988 (2018).

[49] J. J. Madinya, L. W. Chang, S. L. Perry, and C. E. Sing, "Sequence-dependent self-coacervation in high charge-density polyampholytes," Mol. Syst. Des. Eng. (in press).

[50] L. W. Chang, T. K. Lytle, M. Radhakrishnan, J. J. Madinya, J. Velez, C. E. Sing, and S. L. Perry, Nat. Commun. **8**, 1273 (2017).

[51] M. Dzuricky, S. Roberts, and A. Chilkoti, Biochemistry **57**, 2405 (2018).

[52] T. K. Lytle, L. W. Chang, N. Markiewicz, S. L. Perry, and C. E. Sing, ACS Cent. Sci. **5**, 709 (2019).

[53] K. A. Mahdi and M. Olvera de la Cruz, Macromolecules **33**, 7649 (2000).

[54] A. V. Ermoshkin and M. Olvera de la Cruz, Macromolecules **36**, 7824 (2003).

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

[56] Y.-H. Lin, J. Song, J. D. Forman-Kay, and H. S. Chan, J. Mol. Liq. **228**, 176 (2017).

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

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

[59] M. Muthukumar, J. Chem. Phys. **105**, 5183 (1996).

[60] M. Muthukumar, Polym Sci Ser A **58**, 852 (2018).

[61] M. Muthukumar, Macromolecules **50**, 9528 (2017).

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

[63] R. K. Das and R. V. Pappu, Proc. Natl. Acad. Sci. U. S. A. **110**, 13392 (2013).

[64] B. Schuler, A. Soranno, H. Hofmann, and D. Nettels, Annu. Rev. Biophys. **45**, 207 (2016).

[65] I. Konig, A. Zarrine-Afser, M. Aznauryan, A. Soranno, B. Wunderlich, F. Dingfelder, J. Stuber, A. Pluckthun, D. Nettles, and B. Schuler, Nat. Methods **12**, 773 (2015).

[66] A. Soranno, I. Koenig, M. Borgia, H. Hofmann, F. Zosel, D. Nettels, and B. Schuler, Proc. Natl. Acad. Sci. U. S. A. **111**, 4874 (2014).

[67] S. M. Sizemore, S. M. Cope, A. Roy, G. Ghirlanda, and S. M. Vaiana, Biophys. J. **109**, 1038 (2015).

[68] L. Sawle and K. Ghosh, J. Chem. Phys. **143**, 085101 (2015).

[69] T. Firman and K. Ghosh, J. Chem. Phys. **148**, 123305 (2018).

[70] J. Huihui, T. Firman, and K. Ghosh, J. Chem. Phys. **149**, 085101 (2018).

[71] K. Shen and Z.-G. Wang, J. Chem. Phys. **146**, 084901 (2017).

[72] K. Shen and Z.-G. Wang, Macromolecules **51**, 1706 (2018).

[73] M. Muthukumar, Macromolecules **35**, 9142 (2002).

[74] G. Orkoulas, S. K. Kumar, and A. Z. Panagiotopoulos, Phys. Rev. Lett. **90**, 048303 (2003).

[75] J. W. Jiang, L. Blum, O. Bernard, and J. M. Prausnitz, Mol. Phys. **99**, 1121 (2001).

[76] Y. A. Budkov, A. L. Kolesnikov, N. Georgi, E. A. Nogovitsyn, and M. G. Kiselev, J. Chem. Phys. **142**, 174901 (2015).

[77] J. Jiang, J. Feng, H. Liu, and Y. Hu, J. Chem. Phys. **124**, 144908 (2006).

[78] D. W. Cheong and A. Z. Panagiotopoulos, Mol. Phys. **103**, 3031 (2005).

[79] S. Das, A. Eisen, Y.-H. Lin, and H. S. Chan, J. Phys. Chem. B **122**, 5418 (2018).

[80] Y. Wang, A. Lomakin, S. Kanai, R. Alex, and G. B. Benedek, Langmuir **33**, 7715 (2017).

[81] *CRC Handbook of Chemistry and Physics*, 93rd ed., edited by W. M. Haynes (CRC Press, Inc., 2012).

[82] K. Ghosh and K. A. Dill, Proc. Natl. Acad. Sci. U. S. A. **106**, 10649 (2009).

[83] H. Eisenberg and G. R. Mohan, J. Phys. Chem. **63**, 671 (1959).

[84] L. Sabbagh and M. Delsanti, Eur. Phys. J. E. **1**, 75 (2000).

[85] V. M. Prabhu, M. Muthukumar, G. D. Wignall, and Y. B. Melnichenko, Polymer **42**, 8935 (2001).

[86] A. Moreira and R. Netz, Eur. Phys. J. D **13**, 61 (2001).

[87] P. Zhang, N. M. Alsaifi, J. Wu, and Z.-G. Wang, Macromolecules **49**, 9720 (2016).

[88] R. M. Vernon, P. A. Chong, B. Tsang, T. H. Kim, A. Bah, P. Farber, H. Lin, and J. D. Forman-Kay, eLife **7**, e31486 (2018).

[89] M. T. Wei, S. Elbaum-Garfinkle, A. S. Holehouse, C. C. Chen, M. Feric, C. B. Arnold, R. D. Priestley, R. V. Pappu, and C. P. Brangwynne, Nat. Phys. **9**, 1118 (2017).

[90] G. S. Manning, Acc. Chem. Res. **12**, 443 (1979).

[91] M. Muthukumar, J. Chem. Phys. **120**, 9343 (2004).

[92] A. Levy, D. Andelman, and H. Orland, Phys. Rev. Lett. **108**, 227801 (2012).

[93] K. Ghosh and M. Muthukumar, J. Polym. Sci. B **39**, 2644 (2001).

[94] C.-L. Lee and M. Muthukumar, J. Chem. Phys. **130**, 024904 (2009).

[95] A. V. Dobrynin, R. H. Colby, and M. Rubinstein, J. Polym. Sci., Part B: Polym. Phys. **42**, 3513 (2004).

[96] A. V. Dobrynin and M. Rubinstein, Prog. Polym. Sci. **30**, 1049 (2005).

[97] G. L. Dignon, W. Zheng, Y. C. Kim, and J. Mittal, ACS Cent. Sci. **5**, 821 (2019).

[98] Z.-G. Wang, Phys. Rev. E **81**, 021501 (2010).

[99] M. C. Villet and G. H. Fredrickson, J. Chem. Phys. **141**, 224115 (2014).

[100] M. Muthukumar, J. Chem. Phys. **86**, 7230 (1987).

[101] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).

[102] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, 1996).