PNAS
www.pnas.org

Supplementary Information for

Comparative Roles of Charge, π and Hydrophobic Interactions in Sequence-Dependent Phase Separation of Intrinsically Disordered Proteins

Suman Das, Yi-Hsuan Lin, Robert M. Vernon, Julie D. Forman-Kay, and Hue Sun Chan

Hue Sun Chan
Email: chan@arrhenius.med.utoronto.ca

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.2008122117

Supplementary Information

(SI Appendix)

for

Comparative Roles of Charge, π and Hydrophobic Interactions in Sequence-Dependent Phase Separation of Intrinsically Disordered Proteins

Suman Das,[a] **Yi-Hsuan Lin**,[a,b] **Robert M. Vernon**,[b] **Julie D. Forman-Kay**[b,a] and **Hue Sun Chan**[a,1]

[a]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada; [b]Molecular Medicine, Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada

[1]Corresponding author
E-mail: chan@arrhenius.med.utoronto.ca;
Tel: (416)978-2697; Fax: (416)978-8548
Department of Biochemistry, University of Toronto, Medical Sciences Building – 5th Fl., 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada.

Contents

SI Text

Models and Methods

Coarse-Grained Chain Models

The coarse-grained protein chain models in the present study basically follow those in Refs. S1, S2, but with modified and additional features. In accordance with our previous notation for explicit-chain simulation studies,[S2,S3] let $\mu, \nu = 1, 2, \ldots, n$ be the labels for the n IDP chains in the system, and $i, j = 1, 2, \ldots, N$ be the labels of the N residues in each IDP chain. The total potential energy $U_{\rm T}$ is a function of the residue positions, denoted here as $\{\mathbf{R}_{\mu i}\}$. Writing

$$U_{\rm T} = U_{\rm bond} + U_{\rm el} + U_{\rm aa} \ , \tag{S1}$$

where $U_{\rm bond}$ is the bond-length term for chain connectivity:

$$U_{\rm bond} = \frac{K_{\rm bond}}{2} \sum_{\mu=1}^{n} \sum_{i=1}^{N-1} (r_{\mu i,\mu i+1} - l)^2 \tag{S2}$$

with $r_{\mu i,\nu j} \equiv |\mathbf{R}_{\mu i} - \mathbf{R}_{\nu j}|$, $l = 3.8$ Å is the C_α-C_α virtual bond length [l is equivalent to a in Eq. (3) of Ref. S2], $K_{\rm bond} = 10$ kJ mol^{-1}Å^{-2} [this value would be identical to that used in Ref. S1 if the 10 kJ/Å^2 value quoted above Eq. (1) in this reference is a typographical error, i.e., it misses a "/mol"; by comparison, the much stiffer $K_{\rm bond}$ value used in Eq. (3) in Ref. S2, which follows Ref. S4 with the aim of comparing with fixed-bond-length Monte Carlo simulations, is equivalent to 23.7 MJ mol^{-1}Å^{-2}], and $U_{\rm el}$ is the electrostatic interaction:

$$U_{\rm el} = \sum_{\substack{\mu,\nu=1 \\ (\mu,i)\neq(\nu,j)}}^{n} \sum_{i,j=1}^{N} \frac{\sigma_{\mu i}\sigma_{\nu j}e^2}{4\pi\epsilon_0\epsilon_{\rm r} r_{\mu i,\nu j}} \exp\Big(- \kappa r_{\mu i,\nu j}\Big) \ , \tag{S3}$$

wherein $\sigma_{\mu i}$ is the charge of the ith residue in units of elementary electronic charge e, ($\sigma_{\mu i}$ is independent of μ), ϵ_0 is vacuum permittivity, $\epsilon_{\rm r}$ is relative permittivity (dielectric constant), and κ is the reciprocal of the Debye screening length, which is taken to be 10.0 Å in this study ($\kappa = 0.1$ Å^{-1}). Following Table S1 of Ref. S1, σ values for Arg and Lys are assigned to be $+1$, those of Asp and Glu are -1, and that of His is $+0.5$. All other residues are taken to be neutral, i.e., with $\sigma = 0$.

The $U_{\rm aa}$ in Eq. (S1) is the sum of pairwise interaction energies among the residues, viz.,

$$U_{\rm aa} = \sum_{\substack{\mu,\nu=1 \\ (\mu,i)\neq(\nu,j)}}^{n} \sum_{i,j=1}^{N} (U_{\rm aa})_{\mu i,\nu j} \ , \tag{S4}$$

where $(U_{\rm aa})_{\mu i,\nu j}$ is the interaction between the ith residue of the μth chain with the jth residue of the νth chain. We investigate several physically plausible $U_{\rm aa}$ functions, as follows:

The HPS model

The hydrophobicity scale (HPS) model is identical to the one introduced by Dignon et al.[S1] based on an atomic-level hydrophobicity scale devised by Kapcha and Rossky.[S5] The interaction between amino-acid pairs in this model is given by

$$(U_{\rm aa})_{\mu i,\nu j} = (U_{\rm aa|HPS})_{\mu i,\nu j} \equiv \begin{cases} (U_{\rm LJ})_{\mu i,\nu j} + (1-\lambda_{ij}^{\rm HPS})\epsilon_{ij} \ , & \text{if } r \le 2^{1/6} a_{ij} \\ \lambda_{ij}^{\rm HPS}(U_{\rm LJ})_{\mu i,\nu j} & \text{otherwise} \end{cases} \tag{S5}$$

where $\lambda_{ij}^{\rm HPS} \equiv (\lambda_i + \lambda_j)/2$, $a_{ij} \equiv (a_i + a_j)/2$, with λ_i and a_i being the hydrophobicity and diameter, respectively, of the model amino acid residue at sequence position i, as given, respectively, by the λ and σ values in Table S1 of Ref. S1; $(U_{\rm LJ})_{\mu i,\nu j}$ is the Lennard-Jones (LJ) potential,

$$(U_{\rm LJ})_{\mu i,\nu j} = 4\epsilon_{ij}\left[\left(\frac{a_{ij}}{r_{\mu i,\nu j}}\right)^{12} - \left(\frac{a_{ij}}{r_{\mu i,\nu j}}\right)^{6}\right] \ , \tag{S6}$$

where the LJ well depth ϵ_{ij} (not to be confused with the permittivities) is set to be $\epsilon_{ij} = 0.2$ kcal mol^{-1} irrespective of i, j for the HPS model, as in Ref. S1.

The HPS+cation-π models

In view of the importance of cation-π interactions in protein structure (see discussion in the main text), we consider also a class of model potentials, $U_{\rm aa|HPS+c\pi}$s, that augment the HPS potential with cation-π terms for Arg-Phe, Arg-Trp, Arg-Tyr, Lys-Phe, Lys-Trp, and Lys-Tyr residue pairs. In these interaction schemes,

$$(U_{\rm aa})_{\mu i,\nu j} = (U_{\rm aa|HPS+c\pi})_{\mu i,\nu j} \equiv (U_{\rm aa|HPS})_{\mu i,\nu j} + (U_{\rm aa|c\pi})_{\mu i,\nu j} \ , \tag{S7}$$

where

$$(U_{\rm aa|c\pi})_{\mu i,\nu j} = (\epsilon_{\rm c\pi})_{ij}\left[\left(\frac{a_{ij}}{r_{\mu i,\nu j}}\right)^{12} - \left(\frac{a_{ij}}{r_{\mu i,\nu j}}\right)^{6}\right] \ , \tag{S8}$$

and $(\epsilon_{\rm c\pi})_{ij}$ is the cation-π interaction strength, $(\epsilon_{\rm c\pi})_{ij} > 0$ only if residue pair $\mu i, \nu j$ is one of the aforementioned six cation-π pairs, otherwise $(\epsilon_{\rm c\pi})_{ij} = 0$. This simple form is adopted from the cation-π term in Eq. (S1) of Ref. S6.

Two sets of $(\epsilon_{\rm c\pi})_{ij}$ values are analyzed in the present study:

(i) $(\epsilon_{\rm c\pi})_{ij} = 3.0$ kcal mol^{-1} for all six cation-π pairs. The rationale for using a single $(\epsilon_{\rm c\pi})_{ij}$ value is the suggestion by statistical and other inferences that the variations of

interaction strengths among the six cation-π amino acid residue pairs could be relatively small,[S6,S7] though subsequently we will also explore scenarios in which significant variations in cation-π interaction strengths exist among the pairs. When combined with the $(U_{\mathrm{aa|HPS}})_{\mu i,\nu j}$ contribution in Eq. (S7), $(\epsilon_{\mathrm{c}\pi})_{ij}$ = 3.0 kcal mol^{-1} leads to well depths for $(U_{\mathrm{aa}})_{\mu i,\nu j} = (U_{\mathrm{aa|HPS+c}\pi})_{\mu i,\nu j}$ of $\approx$ 0.85 kcal mol^{-1} for Arg-Phe, Arg-Trp, Arg-Tyr, and corresponding well depths of $\approx$ 0.90 kcal mol^{-1} for Lys-Phe, Lys-Trp, and Lys-Tyr (see Fig. 2a of the main text). It should be noted here that we have chosen an $(\epsilon_{\mathrm{c}\pi})_{ij}$ value significantly smaller than those used in Ref. S6 in order for the model cation-π interactions to be more compatiable with the shallow well depths of the $(U_{\mathrm{aa|HPS}})_{\mu i,\nu j}$ potentials in the HPS model, which has a maximum well depth of 0.2 kcal mol^{-1}. Nonetheless, the $(\epsilon_{\mathrm{c}\pi})_{ij}$ = 3.0 kcal mol^{-1} value still entails a cation-π interaction strength which is about double that of electrostatic interaction when ϵ_{r} in Eq. (S3) corresponds to that of bulk water ($\epsilon_{\mathrm{r}} \approx 80$). This ratio between the strengths of cation-π and electrostatic interactions in an aqueous environment conforms to a similar ratio deduced computationally.[S8]

(ii) Different $(\epsilon_{\mathrm{c}\pi})_{ij}$ values for cation-π pairs involving Arg and pairs involving Lys, with $(\epsilon_{\mathrm{c}\pi})_{ij}$ = 1.85 kcal mol^{-1} for Arg-Phe, Arg-Trp, Arg-Tyr and $(\epsilon_{\mathrm{c}\pi})_{ij}$ = 0.65 kcal mol^{-1} for Lys-Phe, Lys-Trp, and Lys-Tyr. This alternate model cation-π interaction scheme is motivated by observed trends of statistical potentials derived from PDB protein structures such as the Miyazawa-Jernigan energies[S9,S10] used in the KH/MJ model[S1] (described below) and the new analysis presented in the main text as well as recent experimental evidence,[S11,S12] all of which suggest that cation-π interactions involving Arg is more favorable than those involving Lys. The $(\epsilon_{\mathrm{c}\pi})_{ij}$ values in this scheme are chosen such that the combined well depth of $(U_{\mathrm{aa|HPS+c}\pi})_{\mu i,\nu j}$ for cation-aromatic pairs are comparable to the deepest well depth of $\approx$ 0.5 kcal mol^{-1} in the KH/MJ model. In particular, $(\epsilon_{\mathrm{c}\pi})_{ij}$ = 1.85 kcal mol^{-1} leads to a combined well depth of $\approx$ 0.55 kcal mol^{-1} for terms in $(U_{\mathrm{aa|HPS+c}\pi})_{\mu i,\nu j}$ involving Arg-aromatic pairs, whereas $(\epsilon_{\mathrm{c}\pi})_{ij}$ = 0.65 kcal mol^{-1} leads to a corresponding combined well depth of $\approx$ 0.3 kcal mol^{-1} for Lys-aromatic pairs (Fig. 2b of the main text).

The KH (KH/MJ) model

The Kim-Hummer/Miyazawa-Jernigan (KH/MJ) model corresponds to the KH-D model used by Dignon et al.,[S1] and is based on the statistical potentials of Miyazawa and Jernigan (MJ).[S10] Following Ref. S1, we refer to this model as KH in the main text and hereafter. The basic functional form of the KH potential, $U_{\mathrm{aa|KH}}$, is similar to that for the HPS potential in Eq. (S5). For the KH model,

$$(U_{\mathrm{aa}})_{\mu i,\nu j} = (U_{\mathrm{aa|KH}})_{\mu i,\nu j} \equiv \begin{cases} (U_{\mathrm{LJ}})_{\mu i,\nu j} + (1-\lambda_{ij}^{\mathrm{KH}})\epsilon_{ij} \ , & \text{if } r \le 2^{1/6} a_{ij} \\ \lambda_{ij}^{\mathrm{KH}} (U_{\mathrm{LJ}})_{\mu i,\nu j} & \text{otherwise} \end{cases} \tag{S9}$$

where $(U_{\mathrm{LJ}})_{\mu i,\nu j}$ is given by Eq. (S6), but now ϵ_{ij} depends on i, j. Specifically, for the KH

model

$$\epsilon_{ij} = |\alpha(e_{\mathrm{MJ},ij} - e_0)| \,, \tag{S10}$$

where $e_{\mathrm{MJ},ij}$ is the MJ statistical potential between the residue type at position i and the residue type at position j, e_0 is a constant shift of the energies, and

$$\lambda_{ij}^{\mathrm{KH}} = \begin{cases} 1 & \text{if } e_{\mathrm{MJ},ij} \leq e_0 \\ -1 & \text{otherwise} \end{cases} \,. \tag{S11}$$

We use $\alpha = 0.228$ and $e_0 = -1.0$ kcal mol^{-1} in the present study. The resulting pairwise energies e_{MJ} correspond to the KH-D parameter set for IDRs in Table S3 of Ref. S1.

Simulation method

Molecular (Langevin) dynamics simulations are carried out using the protocol outlined in the "Simulation framework" section of Ref. S1, with parameters modified for the present applications. For each simulation, we consider 100 copies of one of the four Ddx4 IDR sequences (Fig. S1) or the three LAF-1 IDR sequences (Fig. S5), governed by one of the above coarse-grained model potential functions. At the initial step, all the IDR chains are randomly placed in a relatively large, $300 \times 300 \times 300$ Å^3 simulation box. Energy minimization is then applied to minimize unfavorable steric clashes among the amino acid residues. Equilibrating NPT simulation is then performed for 50 ns at a temperature of 100 K and pressure of 1 bar, maintained by Martyna-Tobias-Klein (MTK) thermostat and barostat[S13,S14] with a coupling constant of 1 ps. It should be noted that the simulation pressure does not correspond to physical pressure because solvent (water) pressure is not accounted for in the present coarse-grained, implicit-solvent model setup. In this regard, pressure is used entirely as an efficient computational device for achieving condensed configurations as starting point of subsequent simulations. Throughout the dynamics simulation, equations of motion are integrated with a timestep of 10 fs and periodic boundary conditions are applied to all three spatial dimensions. After the initial NPT step, the simulation box is compressed again for 50 ns along all three spatial dimensions at 100 K as successive NVT ensembles (P changes during the process) using Langevin thermostat with friction coefficient 1 ps^{-1}. The extent of compression varies for different systems. Then the dimension along one of the three Cartesian axes of the simulation box is expanded 20 times relative to its initial value for a period of 50 ns while maintaing the temperature at 100 K. Equilibration NVT simulation is then performed at the chosen temperature for 2 μs. Finally, production NVT runs are carried out for 4 μs and the chain configurations are saved every 0.5 ns for subsequent analysis. During the production run, the friction coefficient of the Langevin thermostat is decreased to 0.01 ps^{-1} for sampling efficiency. All simulations are performed by the HOOMD-blue software package.[S15,S16] After the snapshots of simulated chain configurations are collected, the procedure for constructing phase diagrams from the configurations follows that described

in the "Simulation framework" section of Ref. S1 and the "Results and discussion" section of Ref. S2.

Explicit-Water Simulation of IDR-Concentration-Dependent Permittivity

Computational procedure

We estimate the IDR-concentration-dependent relative permittivity[S17,S18] by atomistic explicit-water molecular dynamic simulations performed at six Ddx4 IDR (wildtype, WT) concentrations using GROMACS, version 2016.5.[S19] The simulation proceeds as follows. Initially, a fully extended configuration of a Ddx4 IDR is prepared by PYMOL,[S20] to be used as input for Packmol[S21] to place five Ddx4 IDRs at random locations in a cubic simulation box. The size of the box is varied to achieve different Ddx4 IDR concentrations. The Ddx4 IDRs are solvated by explicit water models in the simulation box. Each of the systems so constructed is then charge neutralized by adding appropriate number of Na^+ ions. This is followed by energy minimization by steepest descent to minimize steric clashes. Hydrogen bonds are constrained with the LINCS algorithm.[S22] Equation of motion is integrated using a time step of 2 fs with the leap-frog integrator[S23] and cubic periodic boundary conditions. Long spatial-range electrostatic interaction is treated with particle mesh Ewald (PME) method[S24] with a grid spacing of 0.16 nm and an interpolation order of 4. A cut-off of 1 nm is used for short-range van der Waals and electrostatic interactions. Initial equilibration is carried out for 2 ns under NVT conditions at 300 K. Temperature is maintained by Velocity-rescale thermostat[S25] with a time constant of 0.1 ps for all simulations. This is followed by equilibration for 2 ns at 300 K under NPT conditions under 1 atm pressure, which is maintained by a Berendsen barostat[S26] with a coupling constant of 2 ps. Since the Berendsen barostat does not always yield an NPT ensemble with high accuracy, the resulting system is equilibrated again for 1 ns as an NPT ensemble using the Parrinello-Rahman barostat[S27,S28] with the same coupling constant, after which the production NPT run is carried out for 20 ns using the same Parrinello-Rahman barostat. Configurations are saved every 1.0 ps during the production run for subsequent analysis. In addition to simulations of Ddx4 IDR in essentially pure water (except a few Na^+ ions), we also conduct simulations with Na^+ and Cl^- ions at [NaCl] = 100 mM.

In order to enable a potentially more direct comparison with analytical theory that does not include the charges of amino acid residues in the estimation of effective permittivity of the aqueoue medium,[S17,S18] we carry out another set of simulations with Ddx4 IDR concentrations similar to the ones for which the above protocol is applied but with the charges of the sidechains of the charged amino acids Arg, Lys, Asp, and Glu artifically turned off. This set of simulation data is referred to as artificial Ddx4 or aDdx4. The same aforementioned procedure for equilibration and production is applied for this set of

simulations. The amber99sb-ildn force field[S29] and the TIP3P water model[S30] are used for both sets of simulations. To assess the robustness of the computed ϵ_r values, all simulations are also repeated using the SPC/E water model.[S31]

Relative permittivity analysis

Static relative permittivity ϵ_r (dielectric constant) is determined by the fluctuation of the total dipole moment vector, $\mathbf{M}_T$, of the system via the relation[S32]

$$\epsilon_r = \frac{\langle M_T^2 \rangle - \langle M_T \rangle^2}{3V\epsilon_0 k_B T} + 1 \,, \tag{S12}$$

where $M_T \equiv (\mathbf{M}_T \cdot \mathbf{M}_T)^{1/2}$ is the magnitude of the system dipole moment, $\langle \dots \rangle$ denotes averaging over system configurations under equilibrium conditions, and V is the volume of the simulation box. This relation, Eq. (S12), has been used to compute the static dielectric constant of several biological systems.[S32–S34] Following the formulation in Ref. S33, $\mathbf{M}_T$ is obtained as sum of dipole moments of individual water molecules and individual Ddx4 IDR chain molecules. Irrespective of the net charge of the molecule (water has net charge 0 whereas Ddx4 IDR has net charge $\approx -4e$), the dipole moment, $\mathbf{m}$, of a molecule comprising of N_m atoms with masses m_s ($s = 1, 2, \dots, N_m$) at positions $\mathbf{r}_s$ with point charges q_s is given by $\mathbf{m} = \sum_s^{N_m} q_s(\mathbf{r}_s - \mathbf{r}_{cm})$, where $\mathbf{r}_{cm} \equiv \sum_s^{N_m} m_s \mathbf{r}_s / \sum_s^{N_m} m_s$ is the center-of-mass position of the molecule. Accordingly, atomic ions, Na^+s and Cl^-s in our case, have zero dipole moment in this formulation. Once the dipole moments of the water and Ddx4 molecules are determined in this manner, they are combined to yield $\mathbf{M}_T$ which in turn provides the system relative permittivity through Eq. (S12). Our computed ϵ_r for various concentrations of Ddx4 IDR at different salt concentrations using both the TIP3P and SPC/E water models are given in Table S1.

Random-Phase-Approximation (RPA) Theory of Phase Separation with IDR-Concentration-Dependent Permittivity

Background

Our group has previously considered, within our RPA theory of liquid-liquid phase separation (LLPS), the effects of relative permittivity $\epsilon_{\rm r}$ being dependent upon local protein concentration;[S17,S18] i.e., $\epsilon_{\rm r} = \epsilon_{\rm r}(\phi_m)$ where ϕ_m is polymer (IDR) volume fraction. An $\epsilon_{\rm r}(\phi_m)$ necessitates changes to our earlier RPA expressions for electrostatic interaction for a constant, position-independent $\epsilon_{\rm r}$, viz. [Eq. (33) of Ref. S17],

$$f_{\rm el} = \frac{1}{2}\int \frac{d^3(ka)^3}{(2\pi)^3}\left\{\ln[\det(1+\hat{G}_k\hat{U}_k)] - {\rm Tr}(\hat{\rho}\,\hat{U}_k)\right\}\ . \tag{S13}$$

Here, as in Ref. S17, a^3 is unit volume, $\hat{G}_k$ is the position correlation matrix, $\hat{\rho}$ is the density operator that provides the densities of various molecular species in the system (accounting for matter, not electric charge), and $\hat{U}_k$ accounts for sequence-dependent Coulumb interactions [the expression for $\hat{U}_k$ is provided by Eq. (35) of Ref. S17]. For the simple illustrative case here, which is a system of only IDR polymers without salt or counterions, $\hat{G}_k$ reduces to the monomer-monomer correlation $(\hat{G}_k)_{ij} = (\rho_m/N)(\hat{G}_{\rm M}(kl))_{ij} = \rho_m \exp[-(kl)^2|i-j|/6]/N$, where ρ_m is monomer density, l is the length of a polymer link (virtual bond length, denoted as b in Ref. S17 as noted above), $i,j = 1,2,\ldots N$ are monomer labels along the polymer chain with N being the length of a chain, and $\hat{\rho}\,\hat{U}_k = \rho_m\hat{U}_k/N$ [Eq. (4) of Ref. S35].

When $\epsilon_{\rm r} = \epsilon_{\rm r}(\phi_m)$, we applied the following modified version of $f_{\rm el}$ [Eq. (68) of Ref. S17]:

$$f_{\rm el} = \int \frac{d\tilde{k}\tilde{k}^2}{4\pi^2}\left\{\frac{1}{\eta}\ln\left[1+\eta\mathcal{G}_1(\tilde{k})\right] - \mathcal{G}_2(\tilde{k})\right\}\ , \tag{S14}$$

where $\tilde{k} = kb$, $\eta = (b/a)^3$ (note that virtual bond length b in Ref. S17 is denoted by l in Ref. S40) and, in the absence of salt and counterions, Eqs. (69a) and (69b) of Ref. S17 become

$$\mathcal{G}_1(\tilde{k}) = \frac{4\pi}{\tilde{k}^2[1+\tilde{k}^2]T_0^*\epsilon_r(\phi_m)}\left(\frac{\phi_m}{N}\langle\sigma|\hat{G}_{\rm M}(\tilde{k})|\sigma\rangle\right), \tag{S15a}$$

$$\mathcal{G}_2(\tilde{k}) = \frac{4\pi}{\tilde{k}^2[1+\tilde{k}^2]T_0^*\epsilon_r(\phi_m)}\left(\frac{\phi_m}{N}\sum_{i=1}^{N}|\sigma_i|\right)\ . \tag{S15b}$$

As in Refs. S17 and S18, column vector $|\sigma\rangle$ is the charge sequence—its ith element, σ_i, being the charge of the ith monomer (residue) of the IDR in units of the electronic charge e, and $\langle\sigma| \equiv |\sigma\rangle^{\rm T}$; $T_0^* \equiv 4\pi\epsilon_0 k_{\rm B}Tb/e^2$ is a reduced temperature. As noted above, ϵ_0 is vacuum permittivity, $k_{\rm B}$ is Boltzmann constant, and T is absolute temperature. Previously,[S17,S18] expressions such as above Eqs. (S14) and (S15) for $\epsilon_{\rm r}(\phi_m)$ were obtained heuristically by

replacing every instance of ϵ_r in the corresponding constant-ϵ_r expressions by $\epsilon_r(\phi_m)$.

Concentration-Dependent Permittivity in the RPA Context

We now examine whether—and if so what—restrictive conditions have to be satisfied for the heuristic prescription Eqs. (S14) and (S15) to be valid.

When ϵ_r is position-independent, the electrostatic interaction energy (potential), in units of k_BT, between two unit point charges e at positions $\mathbf{r}$ and $\mathbf{r}'$ is given by $\mathcal{U}(\mathbf{r},\mathbf{r}') = \mathcal{U}(\mathbf{r}-\mathbf{r}') = e^2/(4\pi\epsilon_0\epsilon_r k_BT|\mathbf{r}-\mathbf{r}'|)$. However, when ϵ_r is position-dependent, i.e., $\epsilon_r = \epsilon_r(\mathbf{r})$, in general the electrostatic potential $\mathcal{U}$ is not expressible in a simple closed form because it is the solution to the generalized Poisson equation

$$-\nabla_{\mathbf{r}} \cdot [\epsilon_r(\mathbf{r})\nabla_{\mathbf{r}}\, \mathcal{U}(\mathbf{r}-\mathbf{r}')] = 4\pi l_B \delta(\mathbf{r}-\mathbf{r}') \;, \tag{S16}$$

as noted by Wang,[S36] where $l_B = e^2/(4\pi\epsilon_0 k_BT)$ is vacuum Bjerrum length (unlike Ref. S17, here l_B does not include ϵ_r). Thus, position dependence of ϵ_r can entail rather complex modifications of the charge-charge interactions. It cannot be analytically treated, in general, by simply replacing the constant ϵ_r in $\mathcal{U}(\mathbf{r},\mathbf{r}') = e^2/(4\pi\epsilon_0\epsilon_r k_BT|\mathbf{r}-\mathbf{r}'|)$ by $\epsilon_r(\mathbf{r})$ or $\epsilon_r(\mathbf{r}')$.

Another concern is that, by construction, RPA theory accounts only for the lowest-order polymer density fluctuations beyond the mean-field homogeneous density. In contrast, some of the proposed IDR-concentration-dependent form of $\epsilon_r = \epsilon_r(\phi_m)$, such as the "slab"[S37] and Clasusius-Mossotti[S38] models and the effective medium approximations of Maxwell Garnett and of Bruggeman[S39] considered in Refs. S17, S18 involve higher-order dependence on ϕ_m, raising questions as to whether application of these $\epsilon_r(\phi_m)$ formula in the context of RPA is consistent with the basic premises of RPA. We address these issues below.

Derivation of RPA with Concentration-Dependent Permittivity

Unless specified otherwise, the notation in this subsection follows that of Ref. S40, as the following formal development is, on one hand, a restricted case of the theory in Ref. S40 in that here we do not consider salt, counterions or Kuhn-length renormalization. On the other hand, the present analysis is an extension of the theory in Ref. S40, which is limited to constant ϵ_rs, to case with a position-dependent $\epsilon_r(\mathbf{r})$. Accordingly, we note that the number of chains in the system, which is symbolized by n in the main text and elsewhere in this Supporting Information, is denoted by n_p (following Ref. S40) in the derivation below.

In general, the Boltzmann factor for the electrostatic interaction energy of a system with charge density $\rho(\mathbf{r})$ is given by $\exp[-(1/2)\int d\mathbf{r}d\mathbf{r}'\rho(\mathbf{r})\mathcal{U}(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')]$. (Note that the electric charge density $\rho(\mathbf{r})$ here and in subsequent development in this section should not be confused with the matter density operator $\hat{\rho}$ or its matrix elements.) We focus first on obtaining an equivalent mathematical form of this factor that is amenable to RPA analyses. By standard field-theoretic Hubbard-Stratonovich transformation, this factor may

be expressed as a functional integral over a conjugate field $\psi(\mathbf{r})$:

$$\frac{1}{(\det\hat{\mathcal{U}})^{1/2}}\left\{\prod_{\mathbf{r}}\int\frac{d\psi(\mathbf{r})}{\sqrt{2\pi}}\right\}\,\exp\left[-\frac{1}{2}\int d\mathbf{r}'d\mathbf{r}''\psi(\mathbf{r}')\mathcal{U}^{-1}(\mathbf{r}',\mathbf{r}'')\psi(\mathbf{r}'')-i\int d\mathbf{r}'\rho(\mathbf{r}')\psi(\mathbf{r}')\right]\,, \tag{S17}$$

where $\hat{\mathcal{U}}$ denotes, in matrix notation, the operator $\mathcal{U}(\mathbf{r},\mathbf{r}')$ [i.e., the matrix element $\hat{\mathcal{U}}_{\mathbf{r},\mathbf{r}'}=\mathcal{U}(\mathbf{r},\mathbf{r}')$], $\mathcal{U}^{-1}(\mathbf{r}',\mathbf{r}'')$ is the $\mathbf{r}',\mathbf{r}''$ matrix element of the inverse operator $\hat{\mathcal{U}}^{-1}$ of $\hat{\mathcal{U}}$. By definition, $\int d\mathbf{r}''\mathcal{U}^{-1}(\mathbf{r},\mathbf{r}'')\mathcal{U}(\mathbf{r}'',\mathbf{r}')=\delta(\mathbf{r}-\mathbf{r}')$. Consider now the operator $-\nabla_{\mathbf{r}''}\cdot[\epsilon_{\rm r}(\mathbf{r}'')\nabla_{\mathbf{r}''}\delta(\mathbf{r}-\mathbf{r}'')]/(4\pi l_{\rm B})$. Since

$$\begin{aligned}\int d\mathbf{r}''\{\nabla_{\mathbf{r}''}\cdot[\epsilon_{\rm r}(\mathbf{r}'')\nabla_{\mathbf{r}''}\delta(\mathbf{r}-\mathbf{r}'')]\}\mathcal{U}(\mathbf{r}'',\mathbf{r}')&=\int d\mathbf{r}''[\epsilon_{\rm r}(\mathbf{r}'')\nabla_{\mathbf{r}''}\delta(\mathbf{r}-\mathbf{r}'')]\cdot\nabla_{\mathbf{r}''}\mathcal{U}(\mathbf{r}'',\mathbf{r}')\\&=\int d\mathbf{r}''\delta(\mathbf{r}-\mathbf{r}'')\{\nabla_{\mathbf{r}''}\cdot[\epsilon_{\rm r}(\mathbf{r}'')\nabla_{\mathbf{r}''}\mathcal{U}(\mathbf{r}'',\mathbf{r}')]\}\end{aligned} \tag{S18}$$

follows from repeated applications of integration by parts under the reasonable assumption that the values of the integrand cancel or vanish at the pertinent boundaries of integration, and by Eq. (S16) the quantity in curly brackets in the last term in Eq. (S18) is $-4\pi l_{\rm B}\delta(\mathbf{r}''-\mathbf{r}')$, Eq. (S18) is evaluated as $-4\pi l_{\rm B}\int d\mathbf{r}''\;\delta(\mathbf{r}-\mathbf{r}'')\delta(\mathbf{r}''-\mathbf{r}')=-4\pi l_{\rm B}\delta(\mathbf{r}-\mathbf{r}')$ and therefore $-\nabla_{\mathbf{r}''}\cdot[\epsilon_{\rm r}(\mathbf{r}'')\nabla_{\mathbf{r}''}\delta(\mathbf{r}-\mathbf{r}'')]/(4\pi l_{\rm B})$ is the $\mathbf{r},\mathbf{r}''$ matrix element of the inverse of $\hat{\mathcal{U}}$, viz.,

$$\mathcal{U}^{-1}(\mathbf{r},\mathbf{r}'')=-\frac{1}{4\pi l_{\rm B}}\nabla_{\mathbf{r}''}\cdot[\epsilon_{\rm r}(\mathbf{r}'')\nabla_{\mathbf{r}''}\delta(\mathbf{r}-\mathbf{r}'')]\;. \tag{S19}$$

Equivalently, the $\mathbf{r}'',\mathbf{r}$ matrix element of $\hat{\mathcal{U}}^{-1}$ takes the form

$$\mathcal{U}^{-1}(\mathbf{r}'',\mathbf{r})=-\frac{1}{4\pi l_{\rm B}}\nabla_{\mathbf{r}}\cdot[\epsilon_{\rm r}(\mathbf{r})\nabla_{\mathbf{r}}\delta(\mathbf{r}''-\mathbf{r})]\;. \tag{S20}$$

It follows that the $-(1/2)\int d\mathbf{r}'d\mathbf{r}''\psi(\mathbf{r}')\mathcal{U}^{-1}(\mathbf{r}',\mathbf{r}'')\psi(\mathbf{r}'')$ factor in Eq. (S17) is given by

$$\begin{aligned}-\frac{1}{2}\int d\mathbf{r}'d\mathbf{r}''\psi(\mathbf{r}')\mathcal{U}^{-1}(\mathbf{r}',\mathbf{r}'')\psi(\mathbf{r}'')&=\frac{1}{8\pi l_{\rm B}}\int d\mathbf{r}d\mathbf{r}'\psi(\mathbf{r})\{\nabla_{\mathbf{r}'}\cdot[\epsilon_{\rm r}(\mathbf{r}')\nabla_{\mathbf{r}'}\delta(\mathbf{r}-\mathbf{r}')]\}\psi(\mathbf{r}')\\&=-\frac{1}{8\pi l_{\rm B}}\int d\mathbf{r}d\mathbf{r}'\psi(\mathbf{r})[\epsilon_{\rm r}(\mathbf{r}')\nabla_{\mathbf{r}'}\delta(\mathbf{r}-\mathbf{r}')]\cdot[\nabla_{\mathbf{r}'}\psi(\mathbf{r}')]\\&=\frac{1}{8\pi l_{\rm B}}\int d\mathbf{r}d\mathbf{r}'\psi(\mathbf{r})[\epsilon_{\rm r}(\mathbf{r}')\nabla_{\mathbf{r}}\delta(\mathbf{r}-\mathbf{r}')]\cdot[\nabla_{\mathbf{r}'}\psi(\mathbf{r}')]\\&=-\frac{1}{8\pi l_{\rm B}}\int d\mathbf{r}d\mathbf{r}'\epsilon_{\rm r}(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')[\nabla_{\mathbf{r}}\psi(\mathbf{r})]\cdot[\nabla_{\mathbf{r}'}\psi(\mathbf{r}')]\\&=-\frac{1}{8\pi l_{\rm B}}\int d\mathbf{r}\;\epsilon_{\rm r}(\mathbf{r})[\nabla_{\mathbf{r}}\psi(\mathbf{r})]\cdot[\nabla_{\mathbf{r}}\psi(\mathbf{r})]\\&=-\frac{1}{8\pi l_{\rm B}}\int d\mathbf{r}\;\epsilon_{\rm r}(\mathbf{r})|\nabla\psi(\mathbf{r})|^2\;,\end{aligned} \tag{S21}$$

where the first equality follows from a mere change in the integration variable, the second and fourth equalities from integration by parts assuming that boundary contribution vanishes, the third equality from $\nabla_{\mathbf{r}'}\delta(\mathbf{r}-\mathbf{r}') = -\nabla_{\mathbf{r}}\delta(\mathbf{r}-\mathbf{r}')$, and the $\mathbf{r}$ subscript of $\nabla_{\mathbf{r}}$ is dropped in the final expression because there is little danger of notational ambiguity. Equation (S21) is identical to the corresponding terms in the Hamiltonians in Eq. (3) of Ref. S41 and Eq. (2.7) of Ref. S36 for systems with an inhomogeneous dielectric medium.

We turn next to the $(\det\hat{\mathcal{U}})^{-1/2}$ factor in Eq. (S17). For any matrices A and B, $(\det A)^{-1} = (\det A^{-1})$ and $(\det AB) = (\det A)(\det B)$, we write $(\det\hat{\mathcal{U}})^{-1/2} = (\det\hat{\mathcal{U}}^{-1})^{1/2} = (\det\hat{\epsilon}_{\rm r})^{1/2}(\det\hat{\mathcal{U}}_0^{-1})^{1/2}$, where $\hat{\mathcal{U}}^{-1}$'s matrix elements $\mathcal{U}^{-1}_{\mathbf{r}\mathbf{r}'} \equiv \mathcal{U}^{-1}(\mathbf{r},\mathbf{r}')$ is given by Eq. (S19), the $\mathbf{r},\mathbf{r}'$ matrix elements of the operators $\hat{\epsilon}_{\rm r}$ and $\hat{\mathcal{U}}_0^{-1}$ are defined, respectively, by

$$(\hat{\epsilon}_{\rm r})_{\mathbf{r}\mathbf{r}'} \equiv \epsilon_{\rm r}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}') \;, \tag{S22}$$

$$(\hat{\mathcal{U}}_0^{-1})_{\mathbf{r}\mathbf{r}'} \equiv -\frac{1}{4\pi l_{\rm B}}\nabla^2_{\mathbf{r}}\delta(\mathbf{r}-\mathbf{r}') \;. \tag{S23}$$

Then, $\hat{\epsilon}_{\rm r}\,\hat{\mathcal{U}}_0^{-1} = \hat{\mathcal{U}}^{-1}$ can be ready verified using integration by parts:

$$\begin{aligned}
(\hat{\epsilon}_{\rm r}\,\hat{\mathcal{U}}_0^{-1})_{\mathbf{r}\mathbf{r}'} &= \int d\mathbf{r}''(\hat{\epsilon}_{\rm r})_{\mathbf{r}\mathbf{r}''}(\hat{\mathcal{U}}_0^{-1})_{\mathbf{r}''\mathbf{r}'} \\
&= -\frac{1}{4\pi l_{\rm B}}\int d\mathbf{r}''\epsilon_{\rm r}(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}'')\nabla^2_{\mathbf{r}''}\delta(\mathbf{r}''-\mathbf{r}') \\
&= \frac{1}{4\pi l_{\rm B}}\int d\mathbf{r}''\epsilon_{\rm r}(\mathbf{r}'')[\nabla_{\mathbf{r}''}\delta(\mathbf{r}-\mathbf{r}'')]\cdot[\nabla_{\mathbf{r}''}\delta(\mathbf{r}''-\mathbf{r}')] \\
&= -\frac{1}{4\pi l_{\rm B}}\int d\mathbf{r}''\nabla_{\mathbf{r}''}\cdot\{\epsilon_{\rm r}(\mathbf{r}'')[\nabla_{\mathbf{r}''}\delta(\mathbf{r}-\mathbf{r}'')]\}\delta(\mathbf{r}''-\mathbf{r}') \\
&= -\frac{1}{4\pi l_{\rm B}}\nabla_{\mathbf{r}'}\cdot[\epsilon_{\rm r}(\mathbf{r}')\nabla_{\mathbf{r}'}\delta(\mathbf{r}-\mathbf{r}')] \\
&= \hat{\mathcal{U}}^{-1}_{\mathbf{r}\mathbf{r}'} \;,\ \text{Q.E.D.}
\end{aligned} \tag{S24}$$

Because $\hat{\epsilon}_{\rm r}$ in Eq. (S22) is a diagonal matrix,

$$(\det\hat{\epsilon}_{\rm r}) = \prod_{\mathbf{r}}\epsilon_{\rm r}(\mathbf{r}) \;. \tag{S25}$$

Using Fourier transformation from $\mathbf{r}$ to $\mathbf{k}$ space,[S40]

$$(\det\hat{\mathcal{U}}_0^{-1}) = \prod_{\mathbf{k}\neq\mathbf{0}}\frac{k^2}{4\pi l_{\rm B}} \;, \tag{S26}$$

where $k^2 \equiv |\mathbf{k}|^2$. Note that the $\mathbf{k}=\mathbf{0}$ term is excluded in the above and subsequent considerations because it does not contribute to the exponential factor in Eq. (S17) for our electrically neutral system of overall neutral polyampholytes.

The free energy per unit volume l^3 in units of $k_{\rm B}T$ of our system is given by

$$f = \frac{\phi_m}{N}\ln\phi_m + (1-\phi_m)\ln(1-\phi_m) - \frac{l^3}{\Omega}\ln\mathcal{Z}_{\rm el}, \tag{S27}$$

where N is the chain length (number of monomers) of the polyampholyte, Ω is solution (system) volume, $\phi_m \equiv l^3 n_p N/\Omega$ is monomer volume fraction with n_p being the total number of identical polyampholyte chains in the solution [n_p corresponds to the variable n used above in the formulation for explicit-chain simulations; it should also be noted here that the alternately defined $\phi_m = a^3 n_p N/\Omega$ in Eq. (3) of Ref. S17—which applies to Eqs. (S14) and (S15) in the present work—is equal to polyampholyte volume fraction only when the size of a monomer is equal to the volume unit a^3 of the model, i.e., when $r_m = 1$; whereas polyampholyte volume fraction is given by $r_m\phi_m$ in general[S17]; for simplicity, $r_m = 1$ is assumed below unless specified otherwise], and $\mathcal{Z}_{\rm el}$ is the electrostatic partition function, which may be viewed as a special case of $\mathcal{Z}'$ in Eq. (A9) of Ref. S40 with no salt, no counterion, and $v_2 = 0$, but now extended to $\epsilon_{\rm r} = \epsilon_{\rm r}(\mathbf{r})$. $\mathcal{Z}_{\rm el}$ is given by integrals over monomer coordinates,

$$\mathcal{Z}_{\rm el} = \int \prod_{\alpha=1}^{n_p}\prod_{\tau=1}^{N} d\mathbf{R}_{\alpha,\tau} e^{-\mathscr{H}[\mathbf{R}]} \ , \tag{S28}$$

where $\mathbf{R}_{\alpha,\tau}$ denotes the coordinate of the τth monomer in the αth polyampholyte [$\mathbf{R}_{\alpha,\tau}$ corresponds to the position variable $\mathbf{R}_{\mu i}$ defined before Eq. (S1) in the formulation for explicit-chain simulations; the monomer label τ corresponds also to the label i in Eq. (S15b)], and

$$\mathscr{H}[\mathbf{R}] = \frac{3}{2l^2}\sum_{\alpha=1}^{n_p}\sum_{\tau=1}^{N-1}\left(\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau}\right)^2 + \frac{1}{2}\sum_{\alpha,\beta=1}^{n_p}\sum_{\tau,\mu=1}^{N}\mathcal{V}_{\alpha\beta}^{\tau\mu}\left(\mathbf{R}_{\alpha,\tau}, \mathbf{R}_{\beta,\mu}\right) \ . \tag{S29}$$

The first term of $\mathscr{H}[\mathbf{R}]$ is for Gaussian-chain connectivity of the polyampholyte chains and $\mathcal{V}_{\alpha\beta}^{\tau\mu}$ in the second term is the interaction potential energy between the τth monomer in the αth chain and the μth monomer in the βth chain, viz.,

$$\mathcal{V}_{\alpha\beta}^{\tau\mu}(\mathbf{r},\mathbf{r}') = l_{\rm B}\sigma_\tau\sigma_\mu\mathcal{U}(\mathbf{r},\mathbf{r}') \ , \tag{S30}$$

where σ_τ, σ_μ are the charges, respectively, of monomers τ, μ along each of the n_p polyampholyte chains. We may now rewrite Eq. (S28) as a functional integral over the charge density $\rho(\mathbf{r})$ by including in the integrand a δ-functional for $\rho(\mathbf{r})$:

$$\mathcal{Z}_{\rm el} = \int\prod_{\mathbf{r}} d\rho(\mathbf{r}) \int \prod_{\alpha=1}^{n_p}\prod_{\tau=1}^{N} d\mathbf{R}_{\alpha,\tau} e^{-\mathscr{H}[\rho,\mathbf{R}]}\ \delta[\rho(\mathbf{r}) - \sum_{\alpha=1}^{n_p}\sum_{\tau=1}^{N}\sigma_\tau\delta(\mathbf{r}-\mathbf{R}_{\alpha,\tau})] \ , \tag{S31}$$

which follows from $\rho(\mathbf{r}) = \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \sigma_\tau \delta(\mathbf{r} - \mathbf{R}_{\alpha,\tau})$, whereas $\mathscr{H}[\rho, \mathbf{R}]$ is defined as

$$\mathscr{H}[\rho, \mathbf{R}] = \frac{3}{2l^2} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2 + \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \; \rho(\mathbf{r}) \mathcal{U}(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}') \; . \tag{S32}$$

Now, by applying Eqs. (S17) and (S21), the partition function $\mathcal{Z}_{\rm el}$ in Eq. (S31) may be expressed as a functional integral over $\rho(\mathbf{r})$, $\mathbf{R}_{\alpha,\tau}$, and the conjugate fields $\psi(\mathbf{r})$:

$$\begin{aligned} \mathcal{Z}_{\rm el} = & \int \prod_{\mathbf{r}} d\rho(\mathbf{r}) \int \prod_{\alpha=1}^{n_p} \prod_{\tau=1}^{N} d\mathbf{R}_{\alpha,\tau} \exp\left[-\frac{3}{2l^2} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2 \right] \\ & \times \frac{1}{(\det \hat{\mathcal{U}})^{1/2}} \left\{ \prod_{\mathbf{r}} \int \frac{d\psi(\mathbf{r})}{\sqrt{2\pi}} \right\} \exp\left[-\frac{1}{8\pi l_{\rm B}} \int d\mathbf{r} \; \epsilon_{\rm r}(\mathbf{r}) |\nabla \psi(\mathbf{r})|^2 - i \int d\mathbf{r}' \rho(\mathbf{r}') \psi(\mathbf{r}') \right] \\ & \times \delta[\rho(\mathbf{r}) - \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \sigma_\tau \delta(\mathbf{r} - \mathbf{R}_{\alpha,\tau})] \; . \end{aligned} \tag{S33}$$

After performing the $\prod_{\mathbf{r}} d\rho(\mathbf{r})$ functional integrals in the above expression, $\mathcal{Z}_{\rm el}$ becomes

$$\mathcal{Z}_{\rm el} = \int \prod_{\alpha=1}^{n_p} \prod_{\tau=1}^{N} d\mathbf{R}_{\alpha,\tau} \frac{1}{(\det \hat{\mathcal{U}})^{1/2}} \left\{ \prod_{\mathbf{r}} \int \frac{d\psi(\mathbf{r})}{\sqrt{2\pi}} \right\} e^{-\mathscr{H}[\psi, \mathbf{R}]} \; , \tag{S34}$$

where

$$\mathscr{H}[\psi, \mathbf{R}] = \frac{3}{2l^2} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2 + \frac{1}{8\pi l_{\rm B}} \int d\mathbf{r} \; \epsilon_{\rm r}(\mathbf{r}) |\nabla \psi(\mathbf{r})|^2 + i \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \sigma_\tau \psi(\mathbf{R}_{\alpha,\tau}) \; . \tag{S35}$$

We now proceed to evaluate the $(\det \hat{\mathcal{U}})^{-1/2}$ factor in Eq. (S34) via the aforementioned relations $(\det \hat{\mathcal{U}})^{-1/2} = (\det \hat{\mathcal{U}}^{-1})^{1/2}$ and $\hat{\mathcal{U}}^{-1} = \hat{\epsilon}_{\rm r} \, \hat{\mathcal{U}}_0^{-1}$. Using Eq. (S25) and applying the correspondence

$$\sum_{\mathbf{r}} \to \frac{\mathcal{N}_{\mathbf{r}}}{\Omega} \int d\mathbf{r} \tag{S36}$$

where $\mathcal{N}_{\mathbf{r}}$ is formally the number of $\mathbf{r}$ positions in the system, we may write

$$\sqrt{\det \hat{\epsilon}_{\rm r}} = \prod_{\mathbf{r}} \sqrt{\epsilon_{\rm r}(\mathbf{r})} = \exp\left\{ \frac{1}{2} \sum_{\mathbf{r}} \ln[\epsilon_{\rm r}(\mathbf{r})] \right\} = \exp\left\{ \frac{\mathcal{N}_{\mathbf{r}}}{2\Omega} \int d\mathbf{r} \ln[\epsilon_{\rm r}(\mathbf{r})] \right\} \; . \tag{S37}$$

For reasons to be enunciated below, consider the case in which $\epsilon_{\rm r}(\mathbf{r})$ is a linear combination of polyampholyte and water relative permittivites, i.e.,

$$\epsilon_{\rm r}(\mathbf{r}) = \epsilon_{\rm p} \phi_m(\mathbf{r}) + \epsilon_{\rm w}[1 - \phi_m(\mathbf{r})] = \epsilon_{\rm w} + \epsilon' \phi_m(\mathbf{r}) \; , \tag{S38}$$

where $\epsilon_{\rm p}$ and $\epsilon_{\rm w}$ are, respectively, the relative permittivities of polymer and water, and $\epsilon' = \epsilon_{\rm p} - \epsilon_{\rm w}$. Since the position-dependent monomer density

$$\phi_m(\mathbf{r}) = l^3 \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \delta\left(\mathbf{r} - \mathbf{R}_{\alpha,\tau}\right) \ , \tag{S39}$$

$$\begin{aligned} \ln[\epsilon_{\rm r}(\mathbf{r})] =& \ln \epsilon_{\rm w} + \ln\left[1 + \frac{\epsilon'}{\epsilon_{\rm w}}\phi_m(\mathbf{r})\right] \\ =& \ln \epsilon_{\rm w} + \ln\left[1 + \frac{\epsilon' l^3}{\epsilon_{\rm w}} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \delta\left(\mathbf{r} - \mathbf{R}_{\alpha,\tau}\right)\right] \ . \end{aligned} \tag{S40}$$

To be consistent with RPA which accounts only for lowest-order polymer density fluctuations, we approximate the above expression for $\ln[\epsilon_{\rm r}(\mathbf{r})]$ by including terms only up to the one linear in ϕ_m, viz.,

$$\ln[\epsilon_{\rm r}(\mathbf{r})] \approx \ln \epsilon_{\rm w} + \frac{\epsilon' l^3}{\epsilon_{\rm w}} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \delta\left(\mathbf{r} - \mathbf{R}_{\alpha,\tau}\right) \ . \tag{S41}$$

Hence the argument of the exponential function in Eq. (S37) is given by

$$\begin{aligned} \frac{\mathcal{N}_{\mathbf{r}}}{2\Omega} \int d\mathbf{r} \ln[\epsilon_{\rm r}(\mathbf{r})] \approx& \frac{\mathcal{N}_{\mathbf{r}}}{2} \ln \epsilon_{\rm w} + \frac{\mathcal{N}_{\mathbf{r}} \epsilon' l^3 n_p N}{2\epsilon_{\rm w}\Omega} = \frac{\mathcal{N}_{\mathbf{r}}}{2} \ln \epsilon_{\rm w} + \frac{\mathcal{N}_{\mathbf{r}}\epsilon'}{2\epsilon_{\rm w}}\phi_m \\ \approx& \frac{\mathcal{N}_{\mathbf{r}}}{2} \ln \epsilon_{\rm w} + \frac{\mathcal{N}_{\mathbf{r}}}{2} \ln\left(1 + \frac{\epsilon'}{\epsilon_{\rm w}}\phi_m\right) \\ =& \frac{\mathcal{N}_{\mathbf{r}}}{2} \ln[\epsilon_{\rm r}(\phi_m)] \ , \end{aligned} \tag{S42}$$

where the position-independent $\phi_m \equiv (l^3/\Omega) \int d\mathbf{r} \sum_{\alpha=1}^{n_p} \sum_{\tau=1}^{N} \delta(\mathbf{r} - \mathbf{R}_{\alpha,\tau}) = l^3 n_p N/\Omega$ is the overall average monomer volume fraction, the second approximate relation is in line with that in Eq. (S41), and the last equality follows from definition Eq. (S38). In formulations involving a size-dependent mean-field lattice model with ϕ_m defined in terms of unit volume with $r_m \neq 1$ (Ref. S17), the actual average monomer volume fraction ϕ is given by $\phi = r_m \phi_m$ where r_m is the monomer size factor, in which case $\epsilon_{\rm r}(\phi_m)$ is understood to represent the $\epsilon_{\rm r}$ expression in which all ϕ_m is replaced by $\phi = r_m\phi_m$; i.e., $\epsilon_{\rm r}(\phi_m) \to \epsilon_{\rm r}(\phi_m \to \phi = r_m\phi_m)$. With Eq. (S42), further application of Eqs. (S26) and (S37) yields

$$\begin{aligned} (\det \hat{\mathcal{U}})^{-1/2} =& \sqrt{\det \hat{\epsilon}_{\rm r}} \sqrt{\det \hat{\mathcal{U}}_0^{-1}} \approx \left[\sqrt{\epsilon_{\rm r}(\phi_m)}\right]^{\mathcal{N}_{\mathbf{r}}} \prod_{\mathbf{k}\neq\mathbf{0}} \sqrt{\frac{k^2}{4\pi l_{\rm B}}} \\ =& \prod_{\mathbf{k}\neq\mathbf{0}} \sqrt{\frac{k^2[\epsilon_{\rm r}(\phi_m)]^{\mathcal{N}_{\mathbf{r}}/(\mathcal{N}_{\mathbf{r}}-1)}}{4\pi l_{\rm B}}} \approx \prod_{\mathbf{k}\neq\mathbf{0}} \sqrt{\frac{k^2\epsilon_{\rm r}(\phi_m)}{4\pi l_{\rm B}}} \end{aligned} \tag{S43}$$

for the $(\det\hat{\mathcal{U}})^{-1/2}$ factor in Eq. (S34). To arrive at this expression, we made use of the fact that the total number of reciprocal space positions $\mathbf{k}$ is $\mathcal{N}_{\mathbf{r}}$ (same as the total number of coordinate space positions $\mathbf{r}$ when $\mathbf{k}=\mathbf{0}$ is included in the count), and that $\mathcal{N}_{\mathbf{r}} \gg 1$. It follows that $\mathcal{Z}_{\rm el}$ in Eq. (S34) may be written as

$$\mathcal{Z}_{\rm el} = \left\{\prod_{\mathbf{k}\neq\mathbf{0}} \sqrt{\frac{k^2\epsilon_{\rm r}(\phi_m)}{4\pi l_{\rm B}}}\right\} \int \prod_{\alpha=1}^{n_p}\prod_{\tau=1}^{N} d\mathbf{R}_{\alpha,\tau} \left\{\prod_{\mathbf{r}} \int \frac{d\psi(\mathbf{r})}{\sqrt{2\pi}}\right\} e^{-\mathscr{H}[\psi,\mathbf{R}]} \,, \tag{S44}$$

where $\mathscr{H}[\psi,\mathbf{R}]$ is given by Eq. (S35) with $\epsilon_{\rm r}(\mathbf{r})$ given by Eq. (S38):

$$\begin{aligned}
\mathscr{H}[\psi,\mathbf{R}] =& \frac{\epsilon_{\rm w}}{8\pi l_{\rm B}} \int d\mathbf{r}\, [\nabla\psi(\mathbf{r})]^2 + \frac{\epsilon'}{8\pi l_{\rm B}} \int d\mathbf{r}\, \phi_m(\mathbf{r})\, [\nabla\psi(\mathbf{r})]^2 \\
&+ \frac{3}{2l^2} \sum_{\alpha=1}^{n_p}\sum_{\tau=1}^{N-1} (\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2 + i \sum_{\alpha=1}^{n_p}\sum_{\tau=1}^{N} \sigma_\tau \psi(\mathbf{R}_{\alpha,\tau}) \\
=& \frac{\epsilon_{\rm w}}{8\pi l_{\rm B}} \int d\mathbf{r}\, [\nabla\psi(\mathbf{r})]^2 + \sum_{\alpha=1}^{n_p} \left\{ \frac{3}{2l^2} \sum_{\tau=1}^{N-1} (\mathbf{R}_{\alpha,\tau+1} - \mathbf{R}_{\alpha,\tau})^2 \right. \\
&\left. + \sum_{\tau=1}^{N} \left[i\sigma_\tau \psi(\mathbf{R}_{\alpha,\tau}) + \frac{\epsilon' l^3}{8\pi l_{\rm B}} [\nabla\psi(\mathbf{R}_{\alpha,\tau})]^2 \right] \right\} ,
\end{aligned} \tag{S45}$$

where Eq. (S39) for $\phi_m(\mathbf{r})$ has been applied to yield the last equality. Utilizing the Fourier transformation $\psi_{\mathbf{k}} = (\Omega/\mathcal{N}_{\mathbf{r}}) \sum_{\mathbf{r}} \psi(\mathbf{r}) \exp(-i\mathbf{k}\cdot\mathbf{r})$ of the conjugate field $\psi(\mathbf{r})$ [which may then be expressed as the inverse transformation of $\psi_{\mathbf{k}}$, i.e., $\psi(\mathbf{r}) = (1/\Omega)\sum_{\mathbf{k}} \psi_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{r})$] and the $\sum_{\mathbf{r}} \leftrightarrow (\mathcal{N}_{\mathbf{r}}/\Omega)\int d\mathbf{r}$ correspondence in Eq. (S36), the first term in the above Eq. (S45) can be rewritten as

$$\begin{aligned}
\frac{\epsilon_{\rm w}}{8\pi l_{\rm B}} \int d\mathbf{r}\, [\nabla\psi(\mathbf{r})]^2 &\to \frac{\epsilon_{\rm w}}{8\pi l_{\rm B}} \left(\frac{\Omega}{\mathcal{N}_{\mathbf{r}}}\right) \sum_{\mathbf{r}} \left[\left(\frac{1}{\Omega} \sum_{\mathbf{k}} \psi_{\mathbf{k}} \nabla e^{-i\mathbf{k}\cdot\mathbf{r}} \right) \cdot \left(\frac{1}{\Omega} \sum_{\mathbf{k}'} \psi_{\mathbf{k}'} \nabla e^{-i\mathbf{k}'\cdot\mathbf{r}} \right) \right] \\
&= -\frac{\epsilon_{\rm w}\Omega}{8\pi l_{\rm B}} \frac{1}{\Omega^2} \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \psi_{\mathbf{k}} (\mathbf{k}\cdot\mathbf{k}') \psi_{\mathbf{k}'} \delta_{\mathbf{k}+\mathbf{k}'} \\
&= \frac{1}{2\Omega} \sum_{\mathbf{k}} \frac{\epsilon_{\rm w} k^2}{4\pi l_{\rm B}} \psi_{\mathbf{k}} \psi_{-\mathbf{k}} = \frac{1}{2\Omega} \sum_{\mathbf{k}\neq\mathbf{0}} \frac{\epsilon_{\rm w} k^2}{4\pi l_{\rm B}} \psi_{\mathbf{k}} \psi_{-\mathbf{k}} \,,
\end{aligned} \tag{S46}$$

where the last equality follows because the $\mathbf{k}=\mathbf{0}$ term vanishes by virtue of the k^2 factor. The remaining terms of $\mathscr{H}[\psi,\mathbf{R}]$ in Eq. (S45) can be rewritten as the summation of contributions from n_p independent polymers, as follows. Consider the partition function

$$\mathcal{Q}_p[\psi] = \int \mathscr{D}[\mathbf{R}] e^{-\mathcal{H}_p[\psi,\mathbf{R}]} \tag{S47}$$

for a single polymer, where $\mathscr{D}[\mathbf{R}] = \prod_{\tau=1}^{N} d\mathbf{R}_\tau$, and

$$\begin{aligned}\mathcal{H}_p[\psi,\mathbf{R}] \equiv & \frac{3}{2l^2}\sum_{\tau=1}^{N-1}(\mathbf{R}_{\tau+1}-\mathbf{R}_\tau)^2 \\ & + \sum_{\tau=1}^{N}\left[\frac{i}{\Omega}\sum_{\mathbf{k}}\sigma_\tau\psi_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} - \frac{\epsilon' l^3}{8\pi l_{\rm B}}\frac{1}{\Omega^2}\sum_{\mathbf{k}}\sum_{\mathbf{k}'}(\mathbf{k}\cdot\mathbf{k}')\psi_{\mathbf{k}}\psi_{\mathbf{k}'}e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}_\tau}\right] \\ = & \frac{3}{2l^2}\sum_{\tau=1}^{N-1}(\mathbf{R}_{\tau+1}-\mathbf{R}_\tau)^2 \\ & + \sum_{\tau=1}^{N}\left[\frac{i}{\Omega}\sum_{\mathbf{k}\neq\mathbf{0}}\sigma_\tau\psi_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} - \frac{\epsilon' l^3}{8\pi l_{\rm B}}\frac{1}{\Omega^2}\sum_{\mathbf{k},\mathbf{k}'\neq\mathbf{0}}(\mathbf{k}\cdot\mathbf{k}')\psi_{\mathbf{k}}\psi_{\mathbf{k}'}e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}_\tau}\right] .\end{aligned} \tag{S48}$$

Note that the label α in $\mathbf{R}_{\alpha,\tau}$ is dropped in $\prod_{\tau=1}^{N} d\mathbf{R}_\tau$ and Eq. (S48) because the pertinent integration is only over the monomer coordinates of a single polymer chain. The $\mathbf{k},\mathbf{k}'=\mathbf{0}$ terms can be excluded in the summations of the last line of Eq. (S48) because in the first summation $\sum_{\tau=1}^{N}\sigma_\tau = 0$ for the overall neutral polyampholytes considered here and the $(\mathbf{k}\cdot\mathbf{k}')$ factor in the second summation means that the $\mathbf{k},\mathbf{k}'=\mathbf{0}$ terms are identically zero.

Utilizing the definition of $\psi(\mathbf{r})$ to $\psi_{\mathbf{k}}$ Fourier transformation stated after Eq. (S45), it can readily be verified that $\mathcal{H}_p[\psi,\mathbf{R}]$ is precisely the $\mathbf{k}$-space version of the quantity enclosed in curly brackets on the right hand side of Eq. (S45). Upon changing the functional integration variables $\psi(\mathbf{r})$ in Eq. (S44) to $\psi_{\mathbf{k}}$ and including the $\mathbf{k}$-independent functional Jacobian $|\delta\psi(\mathbf{r})/\delta\psi_{\mathbf{k}}|$ (which have no effect on the configurational distribution of the system),

$$\left\{\prod_{\mathbf{r}}\int\frac{d\psi(\mathbf{r})}{\sqrt{2\pi}}\right\} \to \left\{\prod_{\mathbf{k}}\int\sqrt{\frac{\mathcal{N}_{\mathbf{r}}}{2\pi\Omega^2}}\,d\psi_{\mathbf{k}}\right\} \tag{S49}$$

formally, and thus Eq. (S44) can now be recast in the equivalent form

$$\begin{aligned}\mathcal{Z}_{\rm el} = & \left\{\prod_{\mathbf{k}\neq\mathbf{0}}\sqrt{\frac{k^2\epsilon_{\rm r}(\phi_m)}{4\pi l_{\rm B}}}\right\}\left\{\prod_{\mathbf{k}}\int\sqrt{\frac{\mathcal{N}_{\mathbf{r}}}{2\pi\Omega^2}}\,d\psi_{\mathbf{k}}\right\}\exp\left[-\frac{1}{2\Omega}\sum_{\mathbf{k}\neq\mathbf{0}}\frac{\epsilon_{\rm w}k^2}{4\pi l_{\rm B}}\psi_{\mathbf{k}}\psi_{-\mathbf{k}}\right] \\ & \times\int\prod_{\alpha=1}^{n_p}\left\{\prod_{\tau=1}^{N}d\mathbf{R}_\tau\exp\left(-\mathcal{H}_p[\psi,\mathbf{R}]\right)\right\} ,\end{aligned} \tag{S50}$$

where we have made use of the fact that in the above expression, the first exponential factor [from Eq. (S46)] is independent of $\mathbf{R}_{\alpha,\tau}$, and the quantity enclosed in the last set of curly brakets [from Eq. (S48)] is identical for all n_p values of α, thus the entire last line of Eq. (S50) is equal to $n_p \ln \mathcal{Q}_p[\psi]$ in accordance with Eq. (S47). Because, as argued above, there is no $\mathbf{k}=\mathbf{0}$ contribution to $\mathcal{H}_p[\psi,\mathbf{R}]$, the $\prod_{\mathbf{k}}(\mathcal{N}_{\mathbf{r}}/2\pi\Omega^2)^{1/2}\int d\psi_{\mathbf{k}}$ functional integral in Eq. (S50) may be restricted to $\prod_{\mathbf{k}\neq\mathbf{0}}(\mathcal{N}_{\mathbf{r}}/2\pi\Omega^2)^{1/2}\int d\psi_{\mathbf{k}}$ with no impact on configurational

distribution. Therefore, $\mathcal{Z}_{\rm el}$ takes the simplified form:

$$\mathcal{Z}_{\rm el} = \left\{ \prod_{\mathbf{k}\neq\mathbf{0}} \int \sqrt{\frac{\mathcal{N}_{\mathbf{r}}}{2\pi\Omega^2}}\, d\psi_{\mathbf{k}} \sqrt{\frac{\epsilon_{\rm r}(\phi_m)k^2}{4\pi l_{\rm B}}} \right\} e^{-\mathscr{H}[\psi_{\mathbf{k}}]} \,, \tag{S51}$$

where

$$\mathscr{H}[\psi_{\mathbf{k}}] = \frac{1}{2\Omega} \sum_{\mathbf{k}\neq\mathbf{0}} \frac{\epsilon_{\rm w} k^2}{4\pi l_{\rm B}} \psi_{\mathbf{k}}\psi_{-\mathbf{k}} - n_p \ln \mathcal{Q}_p[\psi] \,. \tag{S52}$$

We are now in a position to apply RPA by expanding $\ln \mathcal{Q}_p$ around $\psi_{\mathbf{k}} = 0$ up to second order in $\psi_{\mathbf{k}}$,[S40] namely

$$\ln \mathcal{Q}_p[\psi] \approx \ln \mathcal{Q}_p[\psi=0] + \sum_{\mathbf{k}} \left(\frac{\delta \ln \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}} \right)_{\psi=0} \psi_{\mathbf{k}} + \frac{1}{2} \sum_{\mathbf{k},\mathbf{k}'} \left(\frac{\delta^2 \ln \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}\delta\psi_{\mathbf{k}'}} \right)_{\psi=0} \psi_{\mathbf{k}}\psi_{\mathbf{k}'} \,, \tag{S53}$$

wherein the zeroth order term (first term on the right hand side) is a constant that plays no role in determining configurational distribution. The first order term

$$\begin{aligned}
\left(\frac{\delta \ln \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}} \right)_{\psi=0} =& \frac{1}{\mathcal{Q}_p[\psi=0]} \left. \frac{\delta \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}} \right|_{\psi=0} \\
=& \sum_{\tau=1}^{N} \left\langle -\frac{i}{\Omega}\sigma_\tau e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} + 2 \times \frac{\epsilon' l^3}{8\pi l_{\rm B}} \frac{1}{\Omega^2} \mathbf{k} \cdot \sum_{\mathbf{k}'\neq\mathbf{0}} \mathbf{k}' \psi_{\mathbf{k}'} e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}_\tau} \right\rangle_{\psi=0} \\
=& -\frac{i}{\Omega} \sum_{\tau=1}^{N} \sigma_\tau \left\langle e^{-i\mathbf{k}\cdot\mathbf{R}_\tau} \right\rangle_{\psi=0} \\
=& 0
\end{aligned} \tag{S54}$$

as well. Here, the average $\langle ... \rangle_{\psi=0}$ is over monomer coordinates $[\mathbf{R}]$ and evaluated at $\psi_{\mathbf{k}} = 0$, the third equality follows because the second term in the second line of the above equation contains a factor of ψ that is set to zero, and the last equality is a consequence of the overall neutrality of the polyampholytes in the system we considered ($\sum_{\tau=1}^{N} \sigma_\tau = 0$). The second order term in the above Eq. (S53) is given by

$$\begin{aligned}
\left(\frac{\delta^2 \ln \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}\delta\psi_{\mathbf{k}'}} \right)_{\psi=0} =& \frac{1}{\mathcal{Q}_p[\psi=0]} \left. \frac{\delta^2 \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}\delta\psi'_{\mathbf{k}}} \right|_{\psi=0} - \frac{1}{\mathcal{Q}_p[\psi=0]} \left. \frac{\delta \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}} \right|_{\psi=0} \times \frac{1}{\mathcal{Q}_p[\psi=0]} \left. \frac{\delta \mathcal{Q}_p}{\delta\psi'_{\mathbf{k}}} \right|_{\psi=0} \\
=& \frac{1}{\mathcal{Q}_p[\psi=0]} \left. \frac{\delta^2 \mathcal{Q}_p}{\delta\psi_{\mathbf{k}}\delta\psi'_{\mathbf{k}}} \right|_{\psi=0} \\
=& \frac{1}{\Omega^2} \frac{\epsilon' l^3}{4\pi l_{\rm B}} \mathbf{k}\cdot\mathbf{k}' \sum_{\tau=1}^{N} \left\langle e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}_\tau} \right\rangle_{\psi=0} - \frac{1}{\Omega^2} \sum_{\tau,\mu=1}^{N} \sigma_\tau\sigma_\mu \left\langle e^{-i(\mathbf{k}\cdot\mathbf{R}_\tau + \mathbf{k}'\mathbf{R}_\mu)} \right\rangle_{\psi=0} \,,
\end{aligned} \tag{S55}$$

where the second equality follows from Eq. (S54). The two $\mathbf{R}$-averages over Gaussian chain configurations in the above Eq. (S55) may be evaluated as follows. For $\left\langle e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}_\tau}\right\rangle_{\psi=0}$, only a single monomer coordinate variable $\mathbf{R}_\tau$ is involved and thus it is uncontrained and $\mathbf{R}$-averaging entails only a single integration of $\mathbf{R}_\tau$ over the entire system volume Ω. The correspondence $\int d\mathbf{R}_\tau \leftrightarrow (\Omega/\mathcal{N}_\mathbf{r})\sum_{\mathbf{R}_\tau}$ yields $\left\langle e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}_\tau}\right\rangle_{\psi=0} = \delta_{\mathbf{k},-\mathbf{k}'}$. Next, to compute $\left\langle e^{-i(\mathbf{k}\cdot\mathbf{R}_\tau+\mathbf{k}'\mathbf{R}_\mu)}\right\rangle_{\psi=0}$, we rewrite it as $\left\langle e^{-i\mathbf{k}\cdot(\mathbf{R}_\tau-\mathbf{R}_\mu)}e^{-i(\mathbf{k}+\mathbf{k}')\cdot\mathbf{R}_\mu)}\right\rangle_{\psi=0}$, which indicates that the $\mathbf{R}$-averaging involves integrating over two monomer coordinates, one is unconstrained and the other is constrained by the Gaussian chain statistics for two points separated by a contour length $l|\tau-\mu|$. Without loss of generality, we select $\mathbf{R}_\mu$ to be the unconstrained coordinates. As for the first average, summing over $\mathbf{R}_\mu$ using the $\int d\mathbf{R}_\mu \leftrightarrow (\Omega/\mathcal{N}_\mathbf{r})\sum_{\mathbf{R}_\mu}$ correspondence yields the Kronecker $\delta_{\mathbf{k},-\mathbf{k}'}$. In accordance with the Gaussian statistics governed by the $3/2l^2\sum_{\tau=1}^{N-1}(\mathbf{R}_{\tau+1}-\mathbf{R}_\tau)^2$ term of $\mathcal{H}_p[\psi,\mathbf{R}]$ in Eq. (S48), $\mathbf{R}_\tau-\mathbf{R}_\mu$ is weighted by $\exp(-3|\mathbf{R}_\tau-\mathbf{R}_\mu|^2/2l^2|\tau-\mu|)$, and therefore the $\mathbf{R}$-averaging of $e^{-i\mathbf{k}\cdot(\mathbf{R}_\tau-\mathbf{R}_\mu)}$ yields $\exp(-k^2l^2|\tau-\mu|/6)$. These considerations allow us to arrive at the expression

$$\left(\frac{\delta^2\ln\mathcal{Q}_p}{\delta\psi_\mathbf{k}\delta\psi_{\mathbf{k}'}}\right)_{\psi=0} = -\frac{\delta_{\mathbf{k},-\mathbf{k}'}}{\Omega^2}\left[\frac{\epsilon' N l^3 k^2}{4\pi l_\mathrm{B}} + \langle\sigma|\hat{G}_\mathrm{M}(kl)|\sigma\rangle\right] \tag{S56}$$

for Eq. (S55), where $[\hat{G}_\mathrm{M}(kl)]_{\tau\mu} = \exp[-(kl)^2|\tau-\mu|/6]$ as defined above. Therefore, according to Eqs. (S53) and (S54), the $n_p\ln\mathcal{Q}_p[\psi]$ term in Eq. (S52) is given by

$$\begin{aligned} n_p\ln\mathcal{Q}_p[\psi] &\approx \frac{1}{2}n_p\sum_{\mathbf{k},\mathbf{k}'}\left(\frac{\delta^2\ln\mathcal{Q}_p}{\delta\psi_\mathbf{k}\delta\psi_{\mathbf{k}'}}\right)_{\psi=0}\psi_\mathbf{k}\psi_{\mathbf{k}'} \\ &= -\frac{n_p}{2\Omega^2}\sum_{\mathbf{k},\mathbf{k}'}\delta_{\mathbf{k},-\mathbf{k}'}\left[\frac{\epsilon' N l^3 k^2}{4\pi l_\mathrm{B}} + \langle\sigma|\hat{G}_\mathrm{M}(kl)|\sigma\rangle\right]\psi_\mathbf{k}\psi_{\mathbf{k}'} \\ &= -\frac{1}{2\Omega}\sum_{\mathbf{k}\neq 0}\left[\frac{\epsilon'\phi_m k^2}{4\pi l_\mathrm{B}} + \frac{\phi_m}{Nl^3}\langle\sigma|\hat{G}_\mathrm{M}(kl)|\sigma\rangle\right]\psi_\mathbf{k}\psi_{-\mathbf{k}}\,, \end{aligned} \tag{S57}$$

where we have used the definition of polymer volume fraction $\phi_m = l^3 n_p N/\Omega$, and the fact that the $\mathbf{k}=\mathbf{0}$ terms vanishes: the first term because of the k^2 factor and the second term because of the overall neutrality of the polyampholytes, i.e., $\sum_\tau\sigma_\tau = 0$, and $[\hat{G}_\mathrm{M}(0)]_{\tau\mu} = 1$. Combining this result with Eq. (S52), we arrive at

$$\begin{aligned} \mathscr{H}[\psi_\mathbf{k}] &\approx \frac{1}{2\Omega}\sum_{\mathbf{k}\neq\mathbf{0}}\left[\frac{(\epsilon_\mathrm{w}+\epsilon'\phi_m)k^2}{4\pi l_\mathrm{B}} + \frac{\phi_m}{Nl^3}\langle\sigma|\hat{G}_\mathrm{M}(kl)|\sigma\rangle\right]\psi_\mathbf{k}\psi_{-\mathbf{k}} \\ &= \frac{1}{2\Omega}\sum_{\mathbf{k}\neq\mathbf{0}}\left[\frac{\epsilon_\mathrm{r}(\phi_m)k^2}{4\pi l_\mathrm{B}} + \frac{\phi_m}{Nl^3}\langle\sigma|\hat{G}_\mathrm{M}(kl)|\sigma\rangle\right]\psi_\mathbf{k}\psi_{-\mathbf{k}}\,, \end{aligned} \tag{S58}$$

where we have made use of the above definition of $\epsilon_\mathrm{r}(\phi_m)$ which is linear in ϕ_m. We may now evaluate $\mathcal{Z}_\mathrm{el}$ by performing the functional integral $\prod_{\mathbf{k}\neq\mathbf{0}}\int d\psi_\mathbf{k}$ in Eq. (S51). Because

the $\psi_{\mathbf{k}}$s are Fourier transformations of the real-valued field $\psi(\mathbf{r})$, $\psi^*_{\mathbf{k}} = \psi_{-\mathbf{k}}$ and $\prod_{\mathbf{k}\neq\mathbf{0}} \int d\psi_{\mathbf{k}}$ $= \prod_{\mathbf{k}>\mathbf{0}} \int d\psi_{\mathbf{k}} \int d\psi^*_{\mathbf{k}}$, where the $\mathbf{k} > \mathbf{0}$ notation means that the product or summation excludes the origin and is over $\mathbf{k} = (k_1, k_2, k_3)$ but not $-\mathbf{k} = (-k_1, -k_2, -k_3)$. This can be effectuated by first excluding $(k_1, k_2, k_3) = (0, 0, 0)$ and then restricting the product or sum to $k_1 \geq 0$ (or to $k_2 \geq 0$ or to $k_3 \geq 0$). Expressing $\psi_{\mathbf{k}}$ in terms of its real part $\psi^{\mathrm{R}}_{\mathbf{k}}$ and imaginary part $\psi^{\mathrm{I}}_{\mathbf{k}}$, i.e., $\psi_{\mathbf{k}} = \psi^{\mathrm{R}}_{\mathbf{k}} + i\psi^{\mathrm{I}}_{\mathbf{k}}$ and $\psi^*_{\mathbf{k}} = \psi^{\mathrm{R}}_{\mathbf{k}} - i\psi^{\mathrm{I}}_{\mathbf{k}}$ where $\psi^{\mathrm{R}}_{\mathbf{k}}$ and $\psi^{\mathrm{I}}_{\mathbf{k}}$ are real numbers, one obtains $\prod_{\mathbf{k}>\mathbf{0}} \int d\psi_{\mathbf{k}} \int d\psi^*_{\mathbf{k}} = \prod_{\mathbf{k}>\mathbf{0}} 2 \int_{-\infty}^{\infty} d\psi^{\mathrm{R}}_{\mathbf{k}} \int_{-\infty}^{\infty} d\psi^{\mathrm{I}}_{\mathbf{k}}$. Since $\psi_{\mathbf{k}}\psi_{-\mathbf{k}} = (\psi^{\mathrm{R}}_{\mathbf{k}})^2 + (\psi^{\mathrm{I}}_{\mathbf{k}})^2$,

$$\begin{aligned}
\mathcal{Z}_{\mathrm{el}} =& \left\{ \prod_{\mathbf{k}>\mathbf{0}} \left(\frac{\mathcal{N}_{\mathbf{r}}}{\pi\Omega^2} \right) \left[\frac{\epsilon_{\mathrm{r}}(\phi_m) k^2}{4\pi l_{\mathrm{B}}} \right] \int_{-\infty}^{\infty} d\psi^{\mathrm{R}}_{\mathbf{k}} \int_{-\infty}^{\infty} d\psi^{\mathrm{I}}_{\mathbf{k}} \right\} \\
& \times \exp \left\{ \frac{1}{\Omega} \sum_{\mathbf{k}>\mathbf{0}} \left[\frac{\epsilon_{\mathrm{r}}(\phi_m) k^2}{4\pi l_{\mathrm{B}}} + \frac{\phi_m}{N l^3} \langle \sigma | \hat{G}_{\mathrm{M}}(kl) | \sigma \rangle \right] \left[(\psi^{\mathrm{R}}_{\mathbf{k}})^2 + (\psi^{\mathrm{I}}_{\mathbf{k}})^2 \right] \right\} \\
=& \prod_{\mathbf{k}>\mathbf{0}} \left(\frac{\mathcal{N}_{\mathbf{r}}}{\pi\Omega^2} \right) \left[\frac{\epsilon_{\mathrm{r}}(\phi_m) k^2}{4\pi l_{\mathrm{B}}} \right] \times \pi\Omega \left[\frac{\epsilon_{\mathrm{r}}(\phi_m) k^2}{4\pi l_{\mathrm{B}}} + \frac{\phi_m}{N l^3} \langle \sigma | \hat{G}_{\mathrm{M}}(kl) | \sigma \rangle \right]^{-1} \\
=& \prod_{\mathbf{k}\neq\mathbf{0}} \sqrt{\frac{\mathcal{N}_{\mathbf{r}}}{\Omega}} \left[1 + \frac{4\pi l_{\mathrm{B}}}{\epsilon_{\mathrm{r}}(\phi_m) k^2} \frac{\phi_m}{N l^3} \langle \sigma | \hat{G}_{\mathrm{M}}(kl) | \sigma \rangle \right]^{-1/2} .
\end{aligned} \tag{S59}$$

Hence, up to an additive constant $\propto \mathcal{N}_{\mathbf{r}} \ln(\mathcal{N}_{\mathbf{r}}/\Omega)$ that does not affect configurational distribution, the electrostatic contribution to the free energy in Eq. (S27) is equal to

$$\begin{aligned}
f_{\mathrm{el}} \equiv -\frac{l^3}{\Omega} \ln \mathcal{Z}_{\mathrm{el}} &= -\frac{l^3}{\Omega} \sum_{\mathbf{k}\neq\mathbf{0}} \ln \left[1 + \frac{4\pi l_{\mathrm{B}}}{\epsilon_{\mathrm{r}}(\phi_m) k^2} \frac{\phi_m}{N l^3} \langle \sigma | \hat{G}_{\mathrm{M}}(kl) | \sigma \rangle \right]^{-1/2} \\
&\to \frac{l^3}{2} \int \frac{d^3k}{(2\pi)^3} \ln \left[1 + \frac{4\pi l_{\mathrm{B}}}{\epsilon_{\mathrm{r}}(\phi_m) k^2} \frac{\phi_m}{N l^3} \langle \sigma | \hat{G}_{\mathrm{M}}(kl) | \sigma \rangle \right] ,
\end{aligned} \tag{S60}$$

where we have applied the correspondence

$$\frac{1}{\Omega} \sum_{\mathbf{k}} \to \int \frac{d^3k}{(2\pi)^3} \tag{S61}$$

and noted that the $k \to 0$ contribution vanishes inside the integral in Eq. (S60) because $d^3k \propto k^2 dk$ and thus $\sum_{\mathbf{k}\neq\mathbf{0}}$ may be approximated by $\Omega \int d^3k/(2\pi)^3$ for this quantity. The last expression in Eq. (S60) is formally identical to the one we obtained previously by heuristically replacing the position- and ϕ_m-independent ϵ_{r} in simple RPA theory with $\epsilon_{\mathrm{r}}(\phi_m)$ [Eq. (S14)]. This can be readily verified by setting $b = a = l$, hence $\eta = 1$ in Eqs. (S14) and (S15), and noting that $(1/2)d^3k/(2\pi)^3 = k^2 dk/4\pi^2$, in which case the last line of Eq. (S60) is seen to be equal to Eq. (S14) with the $\mathcal{G}_1(\tilde{k})$ term [Eq. (S15a)] present but the $\mathcal{G}_2(\tilde{k})$ term [Eq. (S15b)] omitted (no subtraction of self interaction) as well as $\tilde{k}^2(1+\tilde{k}^2)$ $\to \tilde{k}^2$ (no short-range cutoff for Coulomb interaction).

In other words, the heuristic RPA formulas for $\epsilon_{\rm r} \to \epsilon_{\rm r}(\phi_m)$ in Eqs. (S14) and (S15) can be rigorously established in the context of RPA approximation provided that $\epsilon_{\rm r}$ is a linear function of ϕ_m. Indeed, if $\epsilon_{\rm r}$ was a more complicated function of ϕ_m, the last term in Eq. (S45) would have individual interaction terms, such as $\delta(\mathbf{R}_{\alpha,\tau} - \mathbf{R}_{\beta,\mu})$, etc., that involve different polymer chains, and that would necessitate an additional summation $\sum_\alpha$ over polymer chains instead of a single $\sum_\tau$ over monomers on a single chain. In that case, the subsequent simplification in terms of the single-chain partition function $\mathcal{Q}_p$ [Eq. (S50)] and thus the RPA expansion of $\ln \mathcal{Q}_p$ [Eq. (S53)] cannot proceed in the manner described above. Therefore, it remains unclear whether Eq. (S60) holds in general for $\epsilon_{\rm r}(\phi_m)$ that is not linear in ϕ_m.

In our previous applications, we considered a Coulomb potential with a physical short-range cutoff by the modification

$$\mathcal{U}(\mathbf{r},\mathbf{r}') = \frac{l_{\rm B}}{\epsilon_{\rm r}|\mathbf{r}-\mathbf{r}'|} \to \mathcal{U}(\mathbf{r},\mathbf{r}') = \frac{l_{\rm B}}{\epsilon_{\rm r}|\mathbf{r}-\mathbf{r}'|}\left(1 - e^{-|\mathbf{r}-\mathbf{r}'|/l}\right) \tag{S62}$$

[cf. Eq. (6) of Ref. S35; Eq. (34) of Ref. S17], which for constant, position-independent $\epsilon_{\rm r}$ results in a $f_{\rm el}$ with $1/k^2$ replaced by $1/[k^2(1+k^2b^2)]$. In the context of a general position-dependent $\epsilon_{\rm r}$, this feature can in principle be accounted for by introducing an $\epsilon_{\rm r}(|\mathbf{r}-\mathbf{r}'|)$, but the necessary formalism has not been developed. In the present work, we incorporate this feature by simply replacing the $1/k^2$ factor by $1/[k^2(1+k^2b^2)]$ in Eq. (S60) so as to capture this physical property as much as possible and place our present results on an essentially equal footing with our earlier results for position-independent $\epsilon_{\rm r}$. Mathematically, this procedure may be viewed as a regularization for "ultraviolet" large-k (i.e., small-$|\mathbf{r}-\mathbf{r}'|$) divergence. As such, it does serve to impart a physical short-spatial-range cutoff, though it may not correspond exactly to any particularly modified form of $f_{\rm el}$ in Eq. (S62) that is applicable to a general position-dependent $\epsilon_{\rm r}(\mathbf{r})$.

Taking all of the above into consideration, we use the general formula in Eqs. (S14) and (S15) above (which allows for $a \neq b = l$ and thus $\eta = (b/a)^3 \neq 1$ and $r_m \neq 1$) for comparing RPA theory against explicit-chain simulation, with the understanding that $\epsilon_{\rm r}$ must be a linear function of polymer volume fraction $\phi = r_m\phi_m$. Following previous practice,[S17,S35] the electrostatic self-interaction term $\mathcal{G}_2(\tilde{k}) = 4\pi l_{\rm B}\phi_m/[k^2(1+k^2b^2)\epsilon_{\rm r}(\phi_m)Nb^3]\sum_{\tau=1}^N|\sigma_\tau|$ is subtracted in Eq. (S14). In the context of a position-dependent $\epsilon_{\rm r}(\mathbf{r})$, however, we recognize that this term can be physically significant for capturing the polyampholyte chains' varying preference for different dielectric environments.[S44] Hence we consider also an electrostatic free energy

$$f_{\rm el}^{[\rm self]} \equiv \int \frac{d\tilde{k}\tilde{k}^2}{4\pi^2\eta}\ln\Big[1+\eta\mathcal{G}_1(\tilde{k})\Big] = a^3\int\frac{dkk^2}{4\pi^2}\ln\left[1+\frac{b^3}{a^3}\mathcal{G}_1(kb)\right] \tag{S63}$$

that includes (does not substract) electrostatic self-interaction, and use both Eq. (S14) and

Eq. (S63) in our comparison of analytical theory with chain simulation.

Unit Conversion for Comparison with Explicit-Chain Simulations

The theory-predicted phase diagrams (coexistence curves) in Fig. S7 of the Supporting Information for position- and IDR concentration-independent $\epsilon_{\rm r}$ are computed numerically using the RPA+FH model described in Ref. S17. Specifically, translational and mixing entropy is given by Eqs. (13) and (14), the RPA formula for $f_{\rm el}$ is provided by Eqs. (39) and (40), and the augmented FH term is the one in Eq. (61) of this reference. Values of the parameters in these formulas are adapted to the present application, as follows:

- a: Unit length of the model. We set the unit volume, a^3, to be that of the volume occupied by a water molecule in pure water, i.e., $\phi_{\rm w}^{\rm pure} = \rho_{\rm w}^{\rm pure} \times a^3 = 1$, where the number density of pure water $\rho_{\rm w}^{\rm pure} = 10^6$ g m$^{-3} N_{\rm A}/18.01528$ g where 10^6 g m^{-3} is density of water, $N_{\rm A} = 6.02214086 \times 10^{23}$ is Avogadro's constant and 18.01528 g is molar mass of water. Thus, $a = (1/\rho_{\rm w}^{\rm pure})^{1/3} = 3.104$ Å$= 3.104 \times 10^{-10}$m.
- b: The C_α–C_α virtual bond length of polypeptides $b = l = 3.8$ Å$= 3.8 \times 10^{-10}$m.
- η [in Eq. (39) of Ref. S17]: From the above values for a and b, $\eta = (b/a)^3 = (3.8/3.104)^3 = 1.835$.
- r_m (monomer size factor in Eq. (14) of Ref. S17): The r_m ratio between the size of one amino acid residue in Ddx4 IDR and the unit volume a^3 is obtained as follows. Because the density of pure protein $= 1,587$ mg ml^{-1}, number of amino acid residues (monomers) in Ddx4 IDR is $N = 241$, and the molar mass of Ddx4 IDR is $25,833$ (Ref. S43), the monomer (amino acid residue) number density of pure protein is given by
$$\rho_m^{\rm pure} = (1.587 \times 10^6) \text{ g m}^{-3} \times 241 \times N_{\rm A}/25,833 \text{ g} . \tag{S64}$$
Since the volume fraction ϕ of pure protein is unity by definition, i.e., $\phi = \rho_m \times r_m \times a^3$, it follows that
$$r_m = (a^3 \rho_m^{\rm pure})^{-1} = \rho_{\rm w}^{\rm pure}/\rho_m^{\rm pure} = \frac{25833}{18.0} \cdot \frac{1}{1.587} \cdot \frac{1}{241} = 3.752 . \tag{S65}$$
- r_s and r_c [size factors for salt and counterions, respectively, in Eq. (14) of Ref. S17]: Both r_s and r_c are set to 1.

The conversion between the $\phi_m = n_p N a^3/\Omega$ in analytical theory to Ddx4 concentration, [Ddx4], in units of mg/ml (mg ml^{-1}), is given by

$$\phi_m = \Big\{ [\text{Ddx4(mg/ml)}] \times 1000 \text{ g/mg} \times 236/(\text{Ddx4 molar mass in g}) \Big\} \times N_{\rm A} \times a^3 , \tag{S66}$$

where $N = 236$ is the chain length of the Ddx4 IDRs, (Ddx4 molar mass in g) of the four Ddx4 IDR sequences are 25412.48, 25412.48, 24346.80, and 24740.48, respectively, for WT, CS, FtoA, and RtoK.[S43] It should be noted that there is a slight mismatch in the lengths of Ddx4 IDRs (236 vs 241) because a Ddx4^{N1} sequence with six amino acids added to its C-terminus as a tag was used in experiments.[S42,S43] Nonetheless, $N = 236$ is adopted in Eq. (S66) because the $N = 236$ sequence published in Ref. S42 is used in our simulations. In the context of our approximate analytical theory and coarse-grained chain model, the numerical difference between using $N = 236$ and $N = 241$ is not expected to be significant.

The mean-field Flory-Huggins (FH) χ parameters of non-electrostatic interactions for the four Ddx4 IDR sequences are obtained from averaging the KH potential energies $\epsilon_{ij}(r_0)$ ($= E_{ij}(r_0)$ [KH] in Fig. 1a of main text) for a given sequence (seq) over all i, j pairs of sequence positions except those entailing a charge-charge interaction [i.e., RR (Arg-Arg), RK (Arg-Lys), RD (Arg-Asp), RE (Arg-Glu), KK (Lys-Lys), KD (Lys-Asp), KE (Lys-Glu), DD (Asp-Asp), DE (Asp-Glu), EE (Glu-Glu); see main-text], yielding $\langle E\rangle_{\rm KH,seq} = -0.1047$, -0.1047, -0.0689, and -0.0924 kcal mol^{-1}, respectively, for seq = WT, CS, FtoA, and RtoK. These average sequence-dependent mean-field non-electroatic interaction energies $\langle E\rangle_{\rm KH,seq}$s are converted to the FH $\chi = \varepsilon_h/T^*$ in Eq. (61) of Ref. S17 as follows:

1. Convert per-mole units to per-interaction units:

$$\langle E\rangle_{\rm KH,seq}[(\text{J/amino acid pair})] = \Big\{\langle E\rangle_{\rm KH,seq}[(\text{kcal/mole of amino acid pairs})]/N_{\rm A}\Big\} \times 1000 \text{ cal/kcal} \times 4.18 \text{ J/cal} . \tag{S67}$$

2. Convert to the reduced variables used in analytical theory:

$$(z/2) \times \langle E\rangle_{\rm KH,seq}[(\text{J/amino acid pair})]/(k_{\rm B}T) = -\varepsilon_h/T^* , \tag{S68}$$

where T^* is the reduced temperature given by Eq. (38) in Ref. S17 (see below) and z is an FH geometric factor representing the maximal number of monomers (amino acid residues) that are spatial nearest neighbors to a given monomer; e.g., $z = 6$ for three-dimensional simple cubic lattices. We obtain $z/2 = 3.512$ by fitting our RPA+FH predictions to our explicit-chain simulation results.

3. Convert absolute temperature T in K to the reduced temperature T^*:

$$\frac{1}{T^*} = \frac{e^2}{4\pi\epsilon_0\epsilon_{\rm r}k_{\rm B}b}\frac{1}{T} , \tag{S69}$$

where the electronic charge $e = 1.6 \times 10^{-19}$ C, $\epsilon_0 = 8.854 \times 10^{-12}$ C V^{-1}m^{-1}, $b = 3.8 \times 10^{-10}$ m, and $\epsilon_{\rm r} = 80$, 40, or 20 in accordance with the corresponding

simulations with position- and IDR concentration-independent relative permittivities. Note that $T^* = \epsilon_{\rm r} T_0^*$ where T_0^* is defined after Eq. (S15) above and in Eq. (67) of Ref. S17.

4. Convert $\langle E \rangle_{\rm KH,seq}$ to FH ε_h:
Based on the above consideration,

$$\begin{aligned}
\varepsilon_h &= -T^* \left(\frac{z}{2}\right) \frac{\langle E \rangle_{\rm KH,seq}[(\text{J/amino acid pair})]}{k_{\rm B}T} \\
&= -\left(\frac{4\pi\epsilon_0\epsilon_{\rm r} b}{e^2}\right)\left(\frac{z}{2}\right) \times \Big\{ \langle E \rangle_{\rm KH,seq}[(\text{kcal/mole of amino acid pairs})]/N_{\rm A} \Big\} \\
&\quad \times 1000 \text{ cal/kcal} \times 4.18 \text{ J/cal} \\
&= -\frac{4\pi \times (8.854 \times 10^{-12}) \times (3.8 \times 10^{-10})}{(1.6 \times 10^{-19})^2} \frac{3.512 \times 1000 \times 4.18}{(6.02214086 \times 10^{23})} \\
&\quad \times \epsilon_{\rm r} \langle E \rangle_{\rm KH,seq}[(\text{kcal/mole of amino acid pairs})] \\
&= -0.04026 \times \epsilon_{\rm r} \times \langle E \rangle_{\rm KH,seq}[(\text{kcal/mole of amino acid pairs})] \, .
\end{aligned} \tag{S70}$$

Accordingly, the ε_h values for WT, CS, FtoA, and RtoK Ddx4 IDRs are, respectively, $\varepsilon_h =$ 0.337, 0.337, 0.222, and 0.298 when $\epsilon_{\rm r} = 80$; $\varepsilon_h = 0.169$, 0.169, 0.111, and 0.149 when $\epsilon_{\rm r} = 40$; and $\varepsilon_h = 0.0843$, 0.0843, 0.0555, and 0.0744 when $\epsilon_{\rm r} = 20$.

Note that ε_h decreases with decreasing $\epsilon_{\rm r}$ because the reduced temperature T^* in Eq. (S69) is proportional to $\epsilon_{\rm r}$. In this formulation using T^*, the result of decreasing $\epsilon_{\rm r}$ is a reduction in the strength of favorable FH interactions relative to that of the electrostatic interactions, which is equivalent to the physical situation (with temperature measured in K) of enhanced electrostatic interactions under a reduced $\epsilon_{\rm r}$ while keeping the non-electrostatic interactions unchanged.

SI Figures

WT:
MGDEDWEAEINPHMSSYVPIFEKDRYSGENGDNFNRTPASSSEMDDGPSR
RDHFMKSGFASGRNFGNRDAGECNKRDNTSTMGGFGVGKSFGNRGFSNSR
FEDGDSSGFWRESSNDCEDNPTRNRGFSKRGGYRDGNNSEASGPYRRGGR
GSFRGCRGGFGLGSPNNDLDPDECMQRTGGLFGSRRPVLSGTGNGDTSQS
RSGSGSERGGYKGLNEEVITGSGKNSWKSEAEGGES

CS:
MGDRDWRAEINPHMSSYVPIFEKDRYSGENGRNFNDTPASSSEMRDGPSE
RDHFMKSGFASGDNFGNRDAGKCNERDNTSTMGGFGVGKSFGNEGFSNSR
FERGDSSGFWRESSNDCRDNPTRNDGFSDRGGYEKGNNSEASGPYERGGR
GSFDGCRGGFGLGSPNNRLDPRECMQRTGGLFGSDRPVLSGTGNGDTSQS
RSGSGSERGGYKGLNEKVITGSGENSWKSEARGGES

FtoA:
MGDEDWEAEINPHMSSYVPIAEKDRYSGENGDNANRTPASSSEMDDGPSR
RDHAMKSGAASGRNAGNRDAGECNKRDNTSTMGGAGVGKSAGNRGASNSR
AEDGDSSGAWRESSNDCEDNPTRNRGASKRGGYRDGNNSEASGPYRRGGR
GSARGCRGGAGLGSPNNDLDPDECMQRTGGLAGSRRPVLSGTGNGDTSQS
RSGSGSERGGYKGLNEEVITGSGKNSWKSEAEGGES

RtoK:
MGDEDWEAEINPHMSSYVPIFEKDKYSGENGDNFNKTPASSSEMDDGPSK
KDHFMKSGFASGKNFGNKDAGECNKKDNTSTMGGFGVGKSFGNKGFSNSK
FEDGDSSGFWKESSNDCEDNPTKNKGFSKKGGYKDGNNSEASGPYKKGGK
GSFKGCKGGFGLGSPNNDLDPDECMQKTGGLFGSKKPVLSGTGNGDTSQS
KSGSGSEKGGYKGLNEEVITGSGKNSWKSEAEGGES

Fig. S1: The amino acid sequences (residues given by one-letter code) of the 236-residue Ddx4 IDR (wildtype, WT) and its charge scrambled (CS) variant (introduced by Nott et al.[S42]), phenylalanine-to-alanine variant (FtoA) (corresponds to the 14FtoA in Brady et al.[S43] and Vernon et al.[S12]) and arginine-to-lysine (RtoK) variant[S12] considered in the present study. Residues of amino acids A, D, E, F, K, and R are shown in the same colors as those in Fig. 3a of the main text. Other residues, including Y which is considered in subsequent simulations, are shown in black.

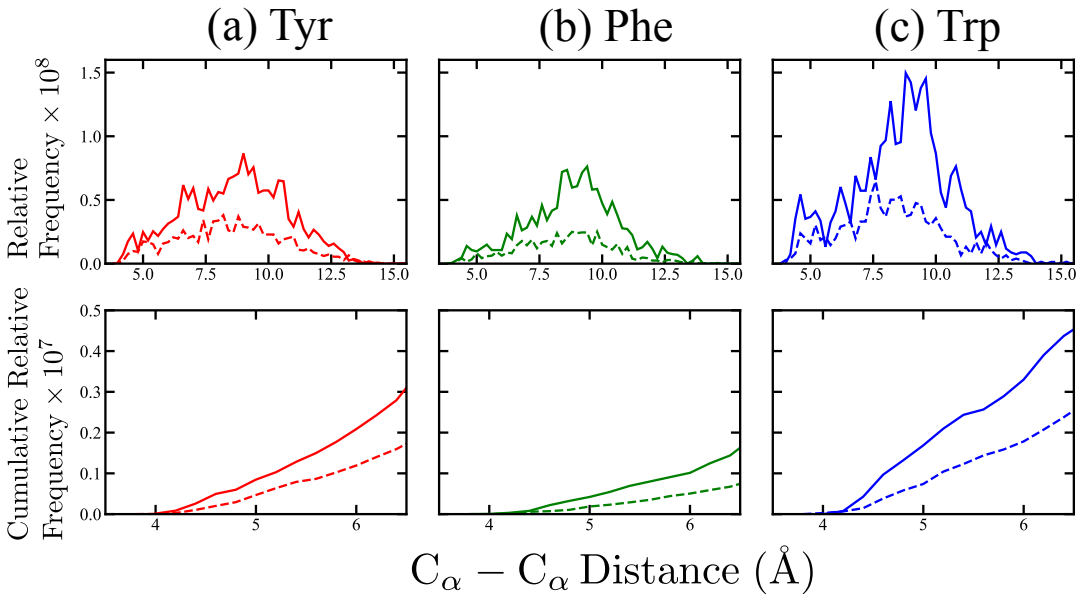

Fig. S2: Statistics of cation-π-like contacts. Distributions of C_α–C_α distance between a positively charged residue [arginine (solid curve) or lysine (dashed curve)] and an aromatic residue [tyrosine (a), phenylalanine (b), or tryptophan (c)] are obtained from the same dataset of 6,943 high-resolution X-ray structures (from a non-redundant set[S12]) used in Fig. 2 of the main text. The bin size for C_α–C_α distance and the color code for different residue pairs are also identical to those in Fig. 2 of the main text. For a given residue pair [Arg-Tyr, Lys-Tyr (a); Arg-Phe, Lys-Phe (b); or Arg-Trp, Lys-Trp (c)], the relative frequency of a given C_α–C_α distance bin is the total number of instances in the dataset in which the C_α–C_α distance between the given pair of residues falls within the bin, normalized (divided) by the product of the two total numbers of residues in the dataset for the two residues making up the pair. Cumulative relative frequency at a given distance is the sum of relative frequencies for distances lower or equal to the given distance. Here, cumulative relative frequencies are reported up to C_α–C_α distance of 6.5 Å, which is illustrative of common criteria for a residue-residue contact. The plotted distributions show clearly that arginine-aromatic contacts are consistently and significantly more numerous than lysine-aromatic contacts when compared on the same footing, suggesting strongly that the overall arginine-aromatic interactions are energetically more favorable than the overall lysine-aromatic interactions.[S9]

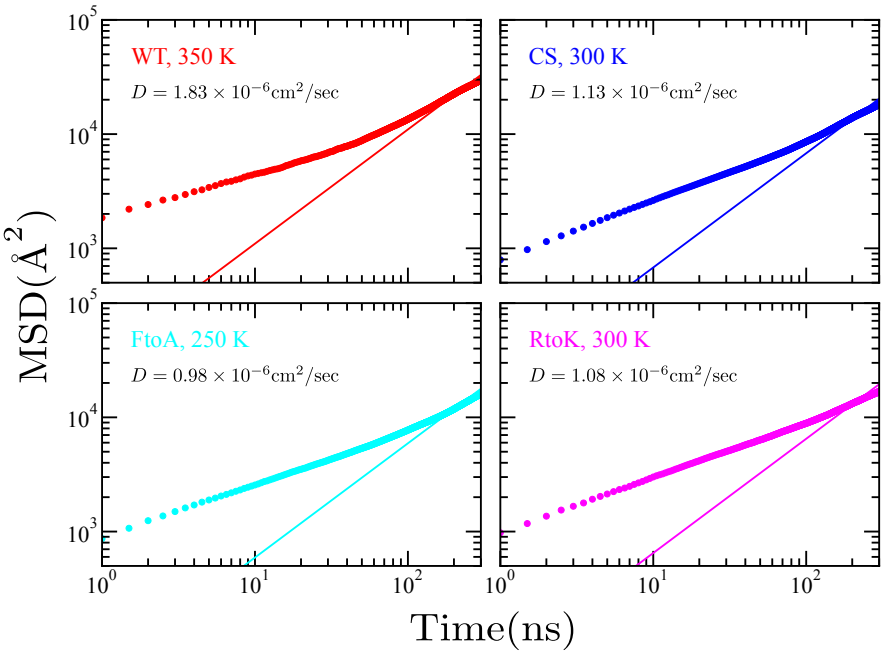

Fig. S3: Verification of liquid-like dynamics of simulated condensed phases. As in Dignon et al.,[S1] a relevant time-dependent mean-square deviation MSD(t) of molecular coordinates was simulated to provide evidence for liquid-like behavior in our model systems, viz.,[S45]

$$\mathrm{MSD}(t) = \frac{1}{n}\left\langle \sum_{\mu=1}^{n} \left| \left[\mathbf{r}_{\mu,\mathrm{CM}}(t+t_0) - \mathbf{r}_{\mathrm{CM}}(t+t_0)\right] - \left[\mathbf{r}_{\mu,\mathrm{CM}}(t_0) - \mathbf{r}_{\mathrm{CM}}(t_0)\right] \right|^2 \right\rangle_{t_0} ,$$

where $\mu = 1, 2, \ldots, n$ labels the model IDR chains, n is the total number of IDR chains in the simulation system, $\mathbf{r}_{\mu,\mathrm{CM}} = \sum_{i=1}^{N} m_i \mathbf{r}_{\mu i} / \sum_{i=1}^{N} m_i$ is the center-of-mass position of the μth chain, with m_i being the mass of the ith bead (residue) along an IDR chain, $\mathbf{r}_{\mathrm{CM}} = \sum_{\mu=1}^{n} \mathbf{r}_{\mu,\mathrm{CM}}/n$ is the center-of-mass of the entire collection of n chains, and the average is over the initial time point t_0. By subtracting drifts in molecular coordinates arising solely from the diffusion of the entire system's center of mass (see Fig. S4), the above-defined MSD(t) values, which are provided by the circles in the plots, are a useful measure of the liquidity of our simulated system. Diffusion coefficients, $D = \{\lim_{t\to\infty} d[\mathrm{MSD}(t)]/dt\}/6$, were then estimated, as indicated by the fitted straight line in each plot. Shown examples for the four Ddx4 IDR variants were simulated using the KH model with relative permittivity $\epsilon_\mathrm{r} = 40$ at the indicated temperatures, each of which is lower than the respective system's critical temperature. The magnitudes of our simulated Ds are similar to those simulated by Dignon et al. for their model FUS systems (Fig. S12 of Ref. S1). Note that our simulated Ds for the model Ddx4 IDR systems are, not unexpectedly, approximately three orders of magnitude higher than the corresponding experimental values[S43] because a unphysically low friction coefficient was necessitated in our Langevin dynamics simulations in order to accelerate sampling and also because a coarse-grained representation of the IDRs was used.

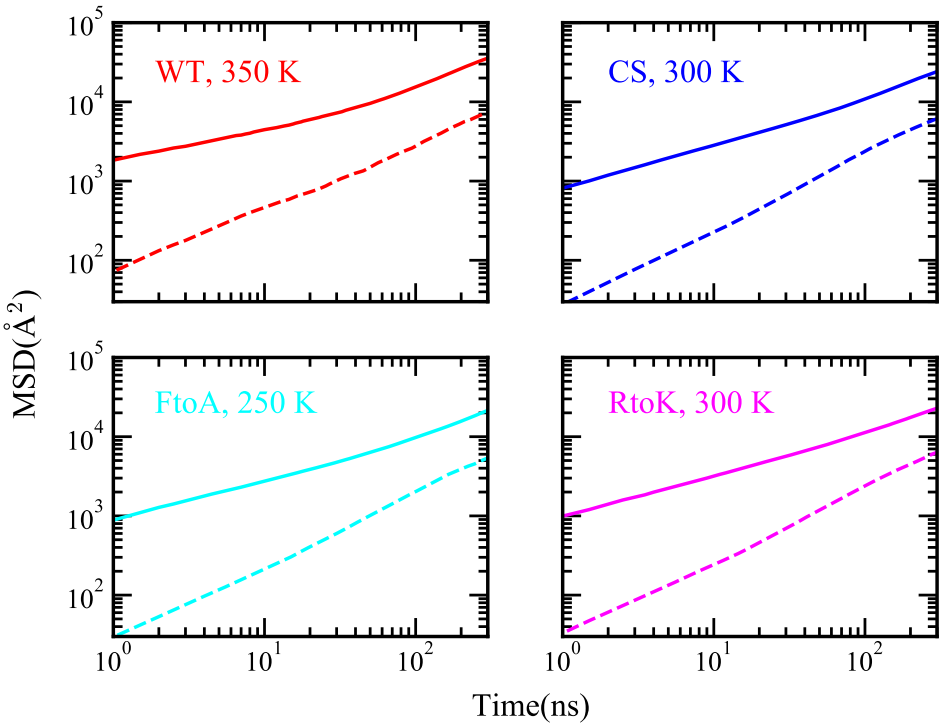

Fig. S4: Center-of-mass diffusion of the simulated Ddx4 IDR systems. Data are from the same systems as those in Fig. S3. The solid curves provide the mean-square deviation of the center-of-mass positions of the IDRs *without* subtracting the the center-of-mass position of the entire system, in which case

$$\mathrm{MSD}(t) = \frac{1}{n}\left\langle \sum_{\mu=1}^{n} \left| \mathbf{r}_{\mu,\mathrm{CM}}(t+t_0) - \mathbf{r}_{\mu,\mathrm{CM}}(t_0) \right|^2 \right\rangle_{t_0} ,$$

whereas the dashed curves represent the diffusion of the center of mass of the entire system of n IDRs, given by $\mathrm{MSD}(t) = \langle |\mathbf{r}_{\mathrm{CM}}(t+t_0) - \mathbf{r}_{\mathrm{CM}}(t_0)\rangle_{t_0}$. Echoing the findings in Fig. S3, a comparison of the solid and dashed curves in the present figure indicates that there is significant diffusion of individual IDRs relative to the center of mass of the entire collection of IDR chains.

LAF-1 RGG WT:
MESNQSNNGG SGNAALNRGG RYVPPHLRGG DGGAAAAASA GGDDRRGGAG
GGGYRRGGGN SGGGGGGGYD RGYNDNRDDR DNRGGSGGYG RDRNYEDRGY
NGGGGGGGNR GYNNNRGGGG GGYNRQDRGD GGSSNFSRGG YNNRDEGSDN
RGSGRSYNND RRDNGGDGLE HHHHHH

LAF-1 RGG RtoK:
MESNQSNNGG SGNAALNKGG KYVPPHLKGG DGGAAAAASA GGDDKKGGAG
GGGYKKGGGN SGGGGGGGYD KGYNDNKDDK DNKGGSGGYG KDKNYEDKGY
NGGGGGGGNK GYNNNKGGGG GGYNKQDKGD GGSSNFSKGG YNNKDEGSDN
KGSGKSYNND KKDNGGDGLE HHHHHH

LAF-1 RGG YtoF:
MESNQSNNGG SGNAALNRGG RFVPPHLRGG DGGAAAAASA GGDDRRGGAG
GGGFRRGGGN SGGGGGGGFD RGFNDNRDDR DNRGGSGGFG RDRNFEDRGF
NGGGGGGGNR GFNNNRGGGG GGFNRQDRGD GGSSNFSRGG FNNRDEGSDN
RGSGRSFNND RRDNGGDGLE HHHHHH

Fig. S5: LAF-1 IDR sequences simulated in Fig. 6 of the main text. Amino acid residues are given by one-letter code. The 176-residue LAF-1 RGG WT sequence and its RtoK and YtoF mutants (all with a polyhistidine tag) are the same as those used for in-vitro studies and given in the Supporting Information of Ref. S46. Underscoring the substitution positions considered in our simulations, K, F, R, and Y residues are shown, respectively, in cyan, magenta, blue, and pink, as in Fig. 6a of the main text. Other residues, including A, D, and E which are highlighted in Fig. 3a and Fig. S1 for Ddx4 IDRs, are shown here in black.

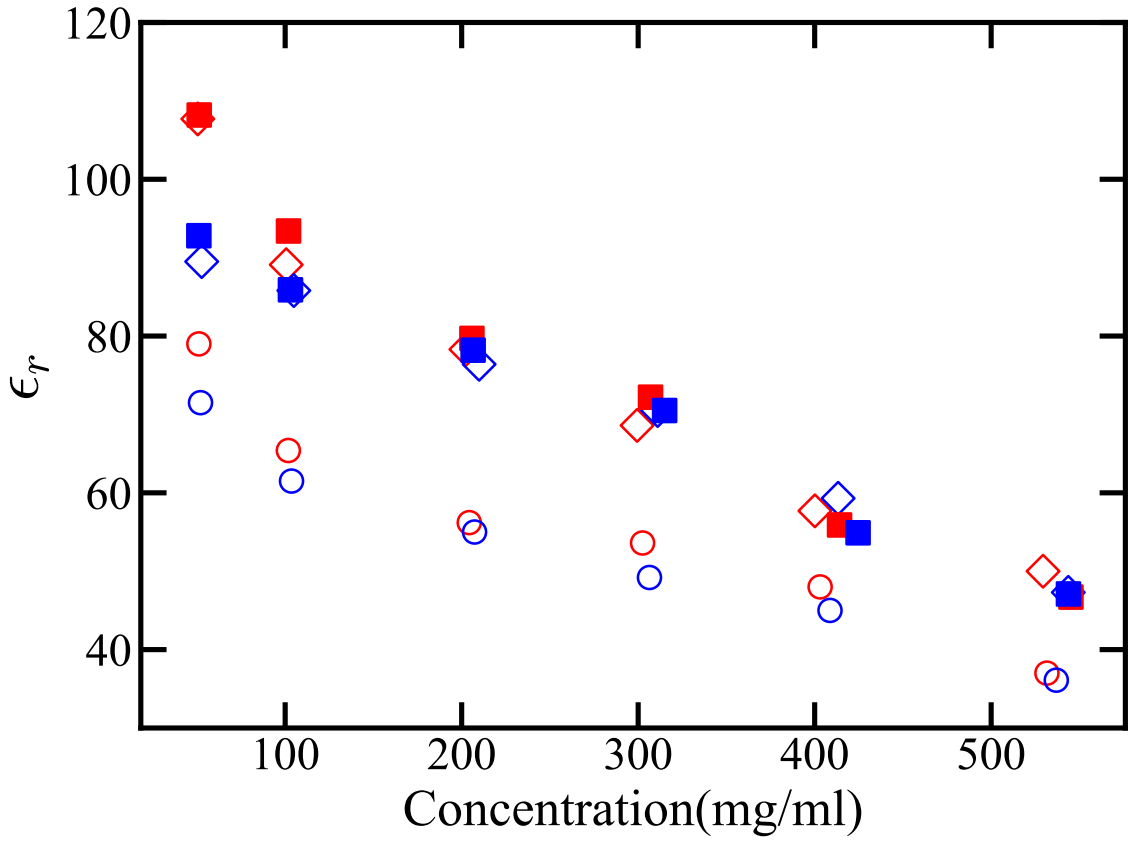

Fig. S6: Simulated IDR-concentration-dependent relative permittivity. Shown results—part of which are also provided in Fig. 7a of the main text—are for the WT Ddx4 IDR. Simulations were conducted using the SPC/E water model with 100 mM NaCl (circles), the TIP3P water model without salt (squares), and the TIP3P model with 100 mM NaCl (diamonds). Red symbols represent ϵ_r values simulated using the full force field, whereas blue symbols denote ϵ_r values simulated while the electric charges on the sidechains of arginine, lysine, glutamic acid, and aspartic acid are artifically turned off. The ϵ_r values plotted here are tabulated in Table S1.

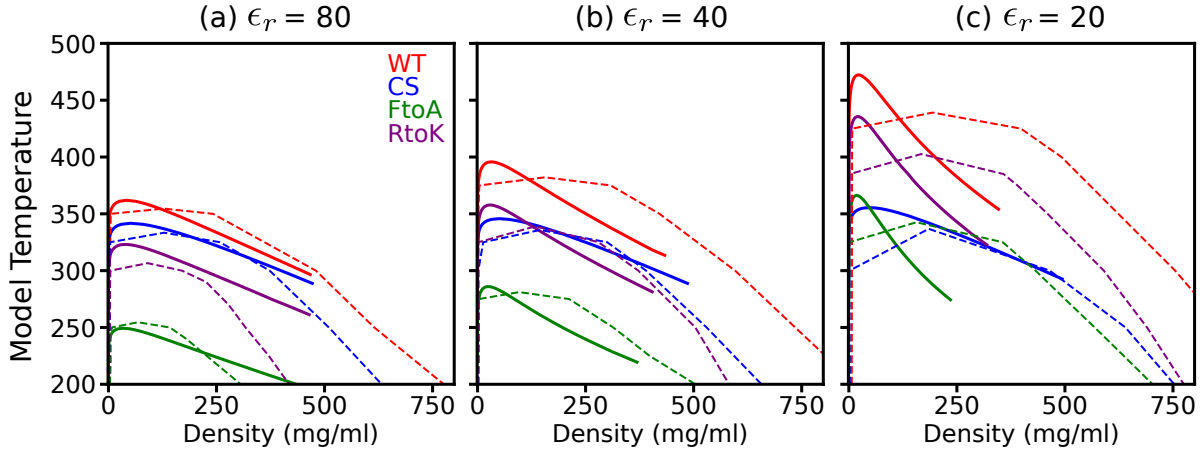

Fig. S7: Comparing analytical theory with simulation for sequence-dependent liquid-liquid phase separation of model Ddx4 systems. Phase diagrams simulated using the explicit-chain KH model under different permittivities (ϵ_r) for the four Ddx4 IDRs from Fig. 4 of the main text are replotted here as dashed curves. Predicted phase diagrams by the RPA+FH theory that afford the best overall fit, at $z/2 = 3.512$, are shown as solid curves.

SI Table

Table S1: IDR-concentration-dependent relative permittivity, ϵ_r, simulated for WT Ddx4 IDR using the SPC/E and TIP3P atomic models of water at $T = 300$ K.

SPC/E + salt[b]		**TIP3P, no salt**		**TIP3P + salt**[b]	
[Ddx4][a]	ϵ_r	[Ddx4][a]	ϵ_r	[Ddx4][a]	ϵ_r
51.1	**79.0**	**51.3**	**108.2**	**50.5**	**107.7**
(52.04)	(71.5)	(51.1)	(92.8)	(52.7)	(89.5)
101.8	**65.4**	**101.9**	**93.4**	**100.6**	**89.1**
(103.6)	(61.5)	(103.0)	(85.9)	(104.9)	(85.8)
204.3	**56.2**	**205.8**	**79.7**	**202.4**	**78.3**
(207.3)	(55.0)	(206.5)	(78.2)	(209.9)	(76.4)
302.5	**53.6**	**307.0**	**72.2**	**299.4**	**68.6**
(306.4)	(49.2)	(315.0)	(70.5)	(311.0)	(70.5)
403.1	**48.0**	**414.1**	**55.9**	**400.1**	**57.7**
(408.7)	(45.0)	(424.6)	(54.9)	(413.3)	(59.3)
531.6	**37.0**	**545.2**	**46.7**	**529.4**	**50.0**
(536.9)	(36.1)	(543.8)	(47.1)	(543.7)	(47.3)

[a] Concentrations (in mg ml^{-1}) and simulated ϵ_r values given in bold font are for systems that apply the full force field; those given in ordinary roman (non-bold) font and in parentheses are for systems in which the electric charges on the sidechains of arginine, lysine, glutamic acid, and aspartic acid of WT Ddx4 IDR are artifically turned off.

[b] [NaCl] = 100 mM.

SI References

[S1] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **14**, e1005941 (2018).

[S2] S. Das, A. N. Amin, Y.-H. Lin, H. S. Chan, Coarse-grained residue-based models of disordered protein condensates: Utility and limitations of simple charge pattern parameters. *Phys. Chem. Chem. Phys.* **20**, 28558–28574 (2018).

[S3] S. Das, A. Eisen, Y.-H. Lin, H. S. Chan, A lattice model of charge-pattern-dependent polyampholyte phase separation. *J. Phys. Chem. B* **122**, 5418–5431 (2018).

[S4] K. S. Silmore, M. P. Howard, A. Z. Panagiotopoulos, Vapor-liquid phase equilibrium and surface tension of fully flexible Lennard-Jones chains. *Mol. Phys.* **115**, 320–327 (2017).

[S5] L. H. Kapcha, P. J. Rossky, A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.* **426**, 484–498 (2014).

[S6] J. Song, S. C. Ng, P. Tompa, K. A. W. Lee, H. S. Chan, Polycation-π interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family *PLoS Comput. Biol.* **9**, e1003239 (2013).

[S7] J. P. Gallivan, D. A. Dougherty, Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9459–9464 (1999).

[S8] J. P. Gallivan, D. A. Dougherty, A computational study of cation-π interaction vs salt bridges in aqueous media: Implications for protein engineering. *J. Am. Chem. Soc.* **122**, 870–874 (2000).

[S9] S. Miyazawa, R. L. Jernigan, Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **18**, 534–552 (1985).

[S10] S. Miyazawa, R. L. Jernigan, Residue-residue potentials with a favourable contact pair term and an unfavourable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644 (1996).

[S11] J. Wang, J. M. Choi, A. S. Holehouse, H. O. Lee, X. Zhang, M. Jahnel, S. Maharana, R. Lemaitre, A. Pozniakovsky, D. Drechsel, I. Poser, R. V. Pappu, S. Alberti, A. A. Hyman, A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699 (2018).

[S12] R. M. Vernon, P. A. Chong, B. Tsang, T. H. Kim, A. Bah, P. Farber, H. Lin, J. D. Forman-Kay, Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *eLife* **7**, e31486 (2018).

[S13] G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).

[S14] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, G. J. Martyna, A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal-isobaric ensemble. *J. Phys. A: Math. Gen.* **39**, 5629–5651 (2006).

[S15] J. A. Anderson, C. D. Lorenz, A. Travesset, General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.* **227**, 5342–5359 (2008).

[S16] J. Glaser, T. D. Nguyen, J. A. Anderson, P. Lui, F. Spiga, J. A. Millan, D. C. Morse, S. C. Glotzer, Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Comput. Phys. Commun.* **192**, 97–107 (2015).

[S17] Y.-H. Lin, J. Song, J. D. Forman-Kay, H. S. Chan, Random-phase-approximation theory for sequence-dependent, biologically functional liquid-liquid phase separation of intrinsically disordered proteins. *J. Mol. Liq.* **228**, 176–193 (2017); Corrigendum: *J. Mol. Liq.* **273**, 676 (2019).

[S18] Y.-H. Lin, J. P. Brady, J. D. Forman-Kay, H. S. Chan, Charge pattern matching as a "fuzzy" mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J. Phys.* **19**, 115003 (2017).

[S19] M. J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team. *GROMACS User Manual version 2016.5*; 2018. http://www.gromacs.org.

[S20] W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: Palo Alto, California, USA, 2002. http://www.pymol.org/.

[S21] L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).

[S22] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.* **18**, 1463–1472 (1997).

[S23] R. W. Hockney, S. P. Goel, J. Eastwood, Quiet high resolution computer models of a plasma. *J. Comp. Phys.* **14**, 148–158 (1974).

[S24] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh Ewald potential. *J. Chem. Phys.* **103**, 8577–8592 (1995).

[S25] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

[S26] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684-3690 (1984).

[S27] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

[S28] S. Nosé, M. L. Klein, Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).

[S29] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, **78**, 1950–1958 (2010).

[S30] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

[S31] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).

[S32] J. W. Pitera, M. Falta, W. F. van Gunsteren, Dielectric properties of proteins from simulation: The effects of solvent, ligands, pH, and temperature. *Biophys. J.* **80**, 2546-2555 (2001).

[S33] T. Rudas, C. Schröder, S. Boresch, O. Steinhauser, Simulation studies of the protein-water interface. II. Properties at the mesoscopic resolution. *J. Chem. Phys.* **124**, 234908 (2006).

[S34] S. Floros, M. Liakopoulou-Kyriakides, K. Karatasos, G. E. Papadopoulos, Detailed study of the dielectric function of a lysozyme solution studied with molecular dynamics simulations. *Eur. Biophys. J.* **44**, 599–611 (2015).

[S35] Y.-H. Lin, J. D. Forman-Kay, H. S. Chan, Sequence-specific polyampholyte phase separation in membraneless organelles. *Phys. Rev. Lett.* **117**, 178101 (2016).

[S36] Z.-G. Wang, Fluctuation in electrolyte solutions: The self-energy. *Phys. Rev. E* **81**, 021501 (2010).

[S37] W. L. Bragg, A. B. Pippard, The form birefringence of macromolecules. *Acta Cryst.* **6**, 865–867 (1953).

[S38] J. D. Jackson, *Classical Electrodynamics*, 2nd edition (Wiley, New York 1975), pp. 154–155.

[S39] V. A. Markel, Introduction to the Maxwell Garnett approximation: Tutorial *J. Opt. Soc. Am. A* **33**, 1244–1256 (2016).

[S40] Y.-H. Lin, J. P. Brady, H. S. Chan, K. Ghosh, A unified analytical theory of heteropolymers for sequence-specific phase behaviors of polyelectrolytes and polyampholytes. *J. Chem. Phys.* **152**, 045102 (2020).

[S41] Q. Wang, T. Taniguchi, G. H. Fredrickson, Self-consistent field theory of polyelectrolyte systems. *J. Phys. Chem. B* **108**, 6733–6744 (2004).

[S42] 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, A. J. Baldwin, Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).

[S43] 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, L. E. Kay, Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8194–E8203 (2017).

[S44] C. N. Schutz, A. Warshel, What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **44**, 400–417 (2001).

[S45] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987).

[S46] B. S. Schuster, G. L. Dignon, W. S. Tang, F. M. Kelley, A. K. Ranganath, C. N. Jahnke, A. G. Simpkins, R. M. Regy, D. A. Hammer, M. C. Good, J. Mittal, Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11421–11431 (2020).