



Beyond monopole electrostatics in regulating conformations of intrinsically disordered proteins

Michael Phillips ^{a,*}, Murugappan Muthukumar ^{b,*} and Kingshuk Ghosh ^{a,c,*}

^aDepartment of Physics and Astronomy, University of Denver, Denver, CO 80208, USA

^bDepartment of Polymer Science and Engineering, University of Massachusetts, Amherst, MA 01003, USA

^cMolecular and Cellular Biophysics, University of Denver, Denver, CO 80208, USA

*To whom correspondence should be addressed: Email: To whom correspondence should be addressed. E-mail: michael.phillips@du.edu (M.P.); muthu@polysci.umass.edu (M.M.); kingshuk.ghosh@du.edu (K.G)

Edited By Ivet Bahar

Abstract

Conformations and dynamics of an intrinsically disordered protein (IDP) depend on its composition of charged and uncharged amino acids, and their specific placement in the protein sequence. In general, the charge (positive or negative) on an amino acid residue in the protein is not a fixed quantity. Each of the ionizable groups can exist in an equilibrated distribution of fully ionized state (monopole) and an ion-pair (dipole) state formed between the ionizing group and its counterion from the background electrolyte solution. The dipole formation (counterion condensation) depends on the protein conformation, which in turn depends on the distribution of charges and dipoles on the molecule. Consequently, effective charges of ionizable groups in the IDP backbone may differ from their chemical charges in isolation—a phenomenon termed charge-regulation. Accounting for the inevitable dipolar interactions, that have so far been ignored, and using a self-consistent procedure, we present a theory of charge-regulation as a function of sequence, temperature, and ionic strength. The theory quantitatively agrees with both charge reduction and salt-dependent conformation data of Prothymosin- α and makes several testable predictions. We predict charged groups are less ionized in sequences where opposite charges are well mixed compared to sequences where they are strongly segregated. Emergence of dipolar interactions from charge-regulation allows spontaneous coexistence of two phases having different conformations and charge states, sensitively depending on the charge patterning. These findings highlight sequence dependent charge-regulation and its potential exploitation by biological regulators such as phosphorylation and mutations in controlling protein conformation and function.

Keywords: heteropolymers, IDP, function

Significance Statement

How does the sequence of ionizable amino acids determine conformations of intrinsically disordered proteins (IDP)? Not merely in terms of charge-charge interactions treated so far in the literature! Omnipresent counterions in the solution condense on a certain number of ionizable amino acids to form dipoles along the protein backbone, resulting in charge regulation, which is self-consistently coupled to protein conformation. We present a dipole-based theory accounting for this phenomenon relevant to all charged macromolecules. After validation using experimental data on an IDP, theory predicts IDPs can coexist between two states with distinct net charge and conformation. These findings provide new insights to how mutations and modifications can modulate protein conformation with implications in protein design, function, and evolution.

Introduction

Intrinsically disordered proteins (IDPs), devoid of unique folded structure, generally have more abundant ionizable amino acids compared to their folded counterparts. Accordingly, the overall electrostatic interactions among all charged residues (both positive and negative) influence the overall dimension of the protein (1–5). Besides the average metrics such as charge composition, electrostatics of heteropolymers like IDPs also depends on the order in which the charges are linked in the protein backbone (6–15). This is due to long-range interactions between ionized side chains

of the amino acids which are also topologically correlated (due to chain connectivity) by their specific placement in the protein sequence. However, only a fraction of the ionizable groups will be ionized in a protein conformation, further complicating sequence dependent electrostatics interactions. The partial ionization is due to a phenomenon called charge-regulation, where charge-state of ionizable groups in the presence of other charges—as in the case of an IDP—self-adjusts, mediated by chain conformational entropy, and differs from their native charge state while in isolation. The degree of charge-regulation depends on

Competing Interest: The authors declare no competing interest.

Received: March 26, 2024. **Accepted:** August 13, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

the propensity of oppositely charged ions (dissociated from the chain or salt) to condense back to the protein chain, a process favored by attraction between opposite charges forming a dipole (16, 17), and opposed by loss in entropy of the dissociated ions. The formation of dipoles, due to ion-pair formation between opposite charges, will renormalize the effective charge, charge-charge interaction and give rise to additional sequence dependent charge-dipole and dipole-dipole interactions between distant parts of the chain. All these factors contribute to the overall free energy and collectively determine the resultant charge state.

How do these sequence-specific interactions balance with each other, mediated by conformational entropy, and ultimately determine the degree of ionization of ionizable amino acid residues and the size of the protein? To appreciate the complexity of this puzzle, consider three different classes of sequences: A polyelectrolytic segment (S1) with one type of charge, and two polyampholytes (S2 and S3) having equal number of positive and negative charges but arranged in different order (see Fig. 1).

Purely from electrostatics, ignoring conformational entropy and hydrophobic effects, a nominal uniformly charged polymer segment (S1) actually functions as a heteropolymer of charges (monopoles) and ion-pairs (dipoles) formed by counterion condensation on ionized groups. Even in this simplest polyelectrolytic sequence, the amino acid units exhibit charge-charge repulsion (u_{cc}), charge-dipole attraction (u_{cd}), and dipole-dipole attraction (u_{dd}). Therefore, optimization of these competing contributions of electrostatics is required to understand the emergence of the effective net charge and compatible conformations. Such antagonistic intra-chain electrostatic interactions are exacerbated when the protein segment contains both positive and negative charges (S2 and S3), due to additional attractive charge-charge interaction between the opposite charges functioning synergistically with dipolar interactions. Furthermore, dipoles can arise not

only from dissociated counterions but also from oppositely charged intra-chain units which puts a constraint on chain entropy. As cartooned for the sequence S2, the synergistic/antagonistic interactions among all units (repulsive and attractive u_{cc} , attractive u_{cd} and u_{dd}), and chain conformational entropy control charge-regulation and the resultant protein conformation. As an example of the significant role played by sequence (18), consider S3 which has blocks of similar charges which can favor a string of dipoles due to intra-chain ion-pair formation when oppositely charged blocks come in close proximity. This in turn puts additional contributions to conformational entropy and energetics of dipolar domains.

All of the above contributing factors, namely, interactions among charges and dipoles, hydrophobic interactions among all amino acids in the protein, and the accompanying conformational entropy, must be considered under different conditions of temperature and ionic strength in a self-consistent manner towards a fundamental understanding of net charge and conformation of IDPs. This is the primary objective of the present theory.

We provide a unified theory that accounts for these synergistic/antagonistic factors in a self-consistent manner to determine the degree of ionization (of positive and negative charges), and size of the chain (average end-to-end distance) as functions of sequence charge patterning including response to salt and temperature. In doing so, we also provide insights to the well-known problem of charge regulation but in polyampholytes and proteins (18–22). We build this unifying theory on two well developed but separate lines of investigation: one that models the degree of ionization of homopolyelectrolyte (20, 23–29) and the other determining sequence effects on the conformation of heteropolymers with different types of charges assuming full ionization (7, 8, 10, 30–32).

We first benchmark the validity of our theory against available data of charge reduction (19) and chain dimension of

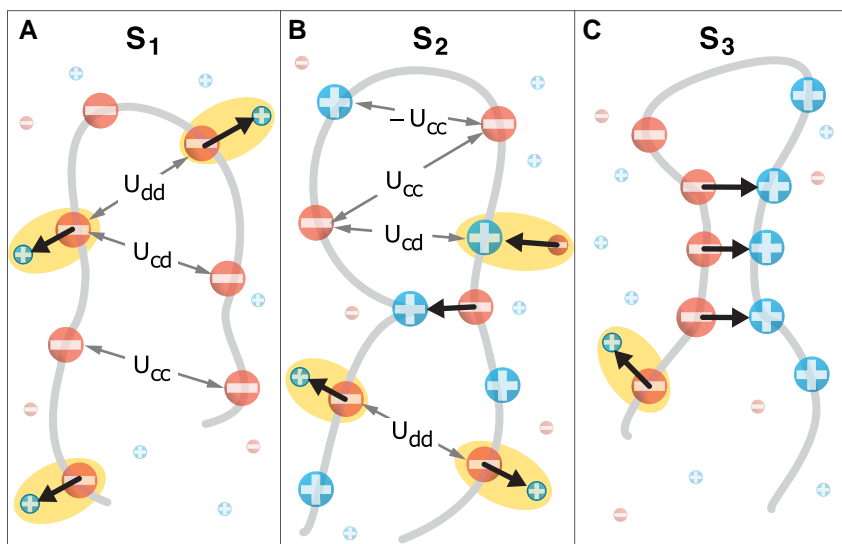


Fig. 1. Sequence dependent synergistic and antagonistic electrostatic interactions among native charges and dipoles formed by counterion condensation, resulting in charge and conformational regulation in IDPs. Sequence 1 (S1, A) has all negative charges (red) on the backbone but is subjected to an optimization between repulsive charge-charge interaction (U_{cc}) between fully ionized groups and attractive charge-dipole ($-U_{cd}$) and dipole-dipole ($-U_{dd}$) interactions involving ion-pairs formed by counterions. Sequences 2 (S2, B) and 3 (S3, C) have equal number of positive and negative charges but distributed differently. When the sequence has opposite charges, there is additional attractive charge-charge interaction ($-U_{cc}$) that is synergistic with the attractive interactions from dipoles and antagonistic with repulsive charge-charge interaction between similar charges. Dipoles can also form due to complexation between oppositely charged amino acids in the chain, in addition to those from counterion condensation. If the sequence is well mixed (S2), the propensity to form intra-chain dipole is small, compared to sequence S3 where charges are well segregated. Small light circles are background ions, bold circles denote counterions that condense on the polymer chain forming dipoles (black arrows with yellow shading), big circles are ionizable residues on the protein backbone.

Prothymosin-alpha measured in single-molecule experiments (33). Using parameters derived from this experiment, we make predictions for ionization and size for different heteropolymer sequences including real IDPs. We find ionization and conformation both depend on charge patterning. We also find IDP sequences can coexist in two states in the charge-conformation landscape giving rise to a cooperative transition with temperature. The origin of the cooperativity is primarily due to sequence dependent interaction between charges and dipoles, an outcome of partial ionization balancing entropic and enthalpic contributions to the free energy. Consequently, the emergence of the coexistence and the gap between the two stable solutions/states (in ionization and chain dimension space) can be controlled by modulating charge patterns in the sequence while keeping the same composition. This effect manifests in a nontrivial way in the context of phosphorylation, which adds negative charges on Serine/Threonine. We notice that while some phosphorylation sites cause only modest changes in conformation, other phosphorylation sites maintaining the same charge composition can cause drastic changes in ionization and conformation due to appearance of the two coexistent phases. We envision that our finding of dipolar forces and sequence effects on IDP conformations will enhance the growing body of research demonstrating the critical role of electrostatics in the function of IDPs (34–40).

Model

Degrees of ionization

We consider an IDP consisting of N residues, with N_+ positive and N_- negative charges upon full ionization (caused by release of counterions). However, counterions and/or salt ions can condense from solution onto some charged residues, neutralizing those charges and forming local dipoles. The partial ionization is modeled by introducing mean degrees of ionization for each charge type, $\alpha_{\pm} \in [0, 1]$. Each residue (m) is assigned a charge $q_m = \alpha_+$, $q_m = -\alpha_-$, or $q_m = 0$ depending on its classification (acidic/basic/neutral), generalizing earlier work on homopolymers (17, 27, 28). Correspondingly, degrees of condensation are given by $1 - \alpha_{\pm}$, and residue (m) specific dipole fraction is defined as $d_m = 1 - \alpha_+$, $d_m = 1 - \alpha_-$, or $d_m = 0$. Net charge of the entire chain is given by $q_{\text{net}} = N_+ \alpha_+ - N_- \alpha_-$.

Charge composition, patterning, and degrees of ionization dictate IDP size, given by ensemble averaged end-to-end distance, $R_{ee} = \sqrt{\langle r_{ee}^2 \rangle}$. It is convenient to measure the expansion or contraction of size with respect to the Flory Random Coil (FRC) limit by introducing a swelling factor x defined as $R_{ee}^2 = Nb\ell x$ where, $b = 3.8 \text{ \AA}$ is the bond length, $\ell = 8.0 \text{ \AA}$ is the Kuhn length (see Refs. (13, 41) for details). This swelling factor x depends on the degrees of ionization (α_+ , α_-), which in turn depend on x , necessitating a self-consistent formulation in which all three variables (x , α_+ , α_-) are determined together from a free energy. Below we describe this free energy as a function of all three variables as well as charge content, charge patterning, salt concentration, and temperature.

Free energy

The total free energy has five physically distinct contributions: $F(x, \alpha_+, \alpha_-) = F_1(\alpha_+, \alpha_-) + F_2(\alpha_+, \alpha_-) + F_3(\alpha_+, \alpha_-) + F_4(\alpha_+, \alpha_-) + F_5(x, \alpha_+, \alpha_-)$. F_1 , F_2 are the combinatorial, translational entropies of the chain; F_3 is the fluctuation contribution of all ions (see [Supplementary Material](#), eqs. S1-S6, for details of these three terms). F_4 is the energy (related to equilibrium constant) of ion pair formation arising

from each counterion (or salt ion) condensed with its oppositely charged partner on the chain, given in units of $\beta = 1/(k_B T)$ by (27)

$$\frac{\beta F_4}{N} = -[f_+(1 - \alpha_+) + f_-(1 - \alpha_-)] \frac{\tilde{\ell}_B}{\tilde{p}} \left(\delta + \frac{1}{2} \right) \quad (1)$$

where, $f_+ = N_+/N$ and $f_- = N_-/N$ are sequence charge fractions, $\tilde{p} \equiv p/b$ is the (nondimensional) distance between a pair of ions, $\tilde{\ell}_B \equiv \ell_B/b$, ℓ_B is Bjerrum length ($e^2/4\pi\epsilon_0\epsilon k_B T$), and $\delta = \epsilon/\epsilon_l$ is the dielectric mismatch between water's dielectric constant, $\epsilon = 80$, and that of the local chain environment, ϵ_l . Typical values of dielectric mismatch for IDPs are $\delta \in [1.3, 2.7]$, assuming $\epsilon_l \in [30, 60]$ (34). (See [Supplementary Material](#) for details on parameter definitions and estimation.) We assume the distance separating opposite ions is sufficiently small that screening can be neglected. Parameters describing the equilibrium constant of binding/unbinding of protons can be different from that due to the dissociated salt ions. However when fitted with data, we use a single set of effective parameters that are independent of the ion (salt or proton) type. When fitting single molecule data (protein concentration in pico molar) with salt concentration (in milli molar), as will be done below, we expect the fitted parameters to primarily arise from salt induced charge regulation. Contributions to the free energy from chain connectivity and intra-chain interactions are modeled by $F_5(x, \alpha_+, \alpha_-)$ derived using a variational approach (30, 41).

$$\beta F_5 = \frac{3}{2}(x - \ln(x)) + \frac{\omega_3 B}{2} \left(\frac{3}{2\pi x} \right)^3 + 2\tilde{\ell}_B Q \left(\frac{3}{2\pi x} \right)^{1/2} + \Omega \left(\frac{3}{2\pi x} \right)^{3/2} \quad (2)$$

The first term on the right-hand side captures chain connectivity. The second term describes three-body repulsive excluded volume with strength ω_3 , to prevent chain collapse in case of strong attraction. The three-body contribution is given by

$$B = \frac{1}{N} \sum_{l=3}^N \sum_{m=2}^{l-1} \sum_{n=1}^{m-1} \frac{(l-n)}{[(l-m)(m-n)]^{3/2}} \quad (3)$$

The third term accounts for sequence dependent charge-charge interactions with screening, where Q is defined as (31)

$$Q = \frac{1}{N} \sum_{m=2}^N \sum_{n=1}^{m-1} q_m q_n (m-n)^{1/2} A(\tilde{\kappa}^2 x (m-n)/6) \quad (4)$$

$$A(z) = 1 - \sqrt{\pi z} \exp(z) \operatorname{erfc}(\sqrt{z}), \quad (5)$$

where, the charge on each residue adopts degree of ionization, $q_m = \pm \alpha_{\pm}$. In the limit of zero screening, and full ionization Q reduces to the Sequence Charge Decoration (SCD) metric defined in prior work (8, 41). Inverse Debye screening length is given by $\tilde{\kappa}^2 = 4\pi\tilde{\ell}_B [f_+ \alpha_+ + f_- \alpha_-] \tilde{\rho} + 2\tilde{c}_s$ where densities of residues and salt ions are nondimensionalized as $\tilde{\rho} \equiv \rho b^3$ and $\tilde{c}_s \equiv c_s b^3$.

The final term with Ω includes all two-body short-range interactions, with three distinct contributions. The first contribution is nonelectrostatic, $\Omega_{\text{non-e}}$, given by

$$\Omega_{\text{non-e}} = \omega_2 \frac{1}{N} \sum_{m=2}^N \sum_{n=1}^{m-1} (m-n)^{-1/2} \quad (6)$$

where, ω_2 is a mean-field nonelectrostatic interaction among all residues. Sequence specificity of this term has also recently been proposed using a Sequence Hydrophathy Decoration metric to model simulated chain dimensions (13). We ignore temperature

dependence of ω_2 that would arise from temperature dependent solvation effects. Two additional contributions to Ω are, Ω_{c-d} and Ω_{d-d} , resulting from directionally averaged charge-dipole and dipole-dipole interactions approximated as delta function potentials (27, 29), with

$$\Omega_{c-d} = \omega_{cd} \frac{1}{N} \sum_{m=2}^N \sum_{n=1}^{m-1} (c_m d_n + c_n d_m) (m-n)^{-1/2} \quad (7)$$

$$\Omega_{d-d} = \omega_{dd} \frac{1}{N} \sum_{m=2}^N \sum_{n=1}^{m-1} d_m d_n (m-n)^{-1/2}, \quad (8)$$

where, for each residue, charge weight is dictated by degree of ionization, $c_m = \alpha_+$, and dipole weight is dictated by degree of condensation, $d_m = 1 - \alpha_+$. All nonionizable residues contribute zero. Despite treating charges of each sign with uniform degrees of ionization, sequence patterning is retained through the placement of positively and negatively charged residues. The exponent of $-1/2$ arises whenever the interaction is described by delta function potential, an approximation made to describe short-range charge-dipole and dipole-dipole terms in analytically tractable form. Consequently, these two contributions give rise to new patterning metrics: Sequence Charge-Dipole Decoration ($SCDD = \Omega_{c-d}/\omega_{cd}$) and Sequence Dipole Decoration ($SDD = \Omega_{d-d}/\omega_{dd}$). The strength of the effective interaction (pseudopotential) due to charge-dipole (ω_{cd}) and dipole-dipole (ω_{dd}) are given by (17, 27, 29),

$$\omega_{cd} = -\frac{\pi}{3} \delta^2 \tilde{\ell}_B^2 \tilde{p}^2 \exp(-2\tilde{\kappa}) [2 + \tilde{\kappa}] \quad (9)$$

$$\omega_{dd} = -\frac{\pi}{9} \delta^2 \tilde{\ell}_B^2 \tilde{p}^4 \exp(-2\tilde{\kappa}) [4 + 8\tilde{\kappa} + 4\tilde{\kappa}^2 + \tilde{\kappa}^3]. \quad (10)$$

Given an IDP sequence and parameters \tilde{p} , δ , ω_2 , ω_3 , total free energy is minimized to determine (x, α_+, α_-) as functions of salt concentration c_s and temperature via ℓ_B . Then, we obtain ensemble averaged end-to-end distance using $R_{ee} = \sqrt{N\ell_B x}$, and net charge using $q_{net} = N_+ \alpha_+ - N_- \alpha_-$. Some parameters of the model (\tilde{p} , δ , ω_3) can be either estimated from previous work or fitted from data while others (ω_2) can only be determined by fitting (see Fig. 2 caption and Supplementary Material).

Results and discussion

Condensation model quantitatively describes experimentally measured chain dimension and predicts charge of an IDP

We first provide a quantitative test of our model against two available data sets for a well studied IDP, Prothymosin-alpha. End-to-end distance at different salt concentrations was recently measured by Schuler and colleagues using single-molecule FRET (33). Net charge of the same protein at low salt is available from electrometry measurement (19). Our model quantitatively matches salt-dependent chain dimensions data as shown in Fig. 2 (see Supplementary Material for details of the fitting procedure, parameter values, its closeness to expectation and Table S1 for sequence). Net charge was predicted using these parameters across salt concentration and compared against the measured net charge at low salt value (magenta cross in Fig. 2). While our model well describes the salt dependent chain dimension and reasonably approximates charge at low salt, the discrepancy can be due to coarse-grain nature of the model that omits atomic contributions to chain conformation, site specific charge-regulation or due to indirect estimate of charge provided by electrometry. We note that existing theoretical models assuming full ionization ($\alpha_{\pm} = 1$) will highly overestimate the net charge (orange line) compared to data. Moreover, sequence dependent full ionization model (using only Eq. 2 with $\alpha_{\pm} = 1$) also overestimates salt dependent chain dimension at low salt values. Besides providing quantitative comparison with existing data, our monopole- and dipole-based model also predicts charge regulation at higher salt concentrations stimulating future experiments and simulation. It would be important to test these predictions with potentiometry experiments (20, 21) giving direct measurements of charge unlike electrometry.

The degree of ionization and chain dimension are self-consistently regulated by sequence patterning

To investigate the effects of specific charge sequence on conformation and ionization, we consider a set of sequences with identical charge composition but different arrangements. We use a subset of sequences originally designed by Das and Pappu (7) where 25

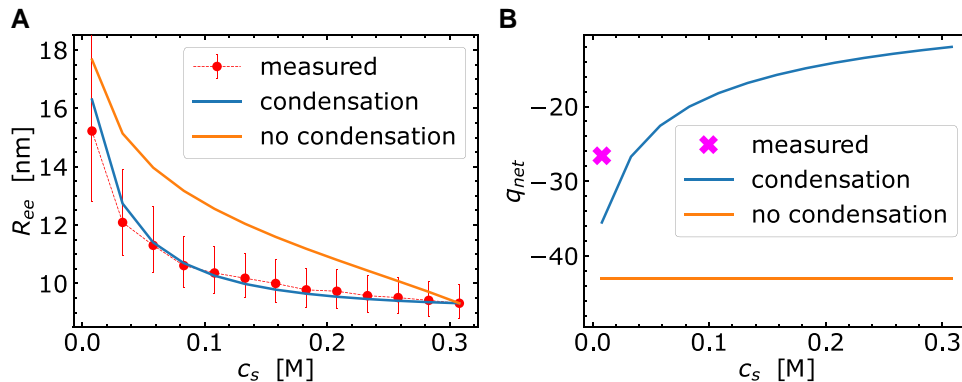


Fig. 2. Counterion condensation model quantitatively reproduces experimentally measured end-to-end distance data of Prothymosin-alpha, and predicts net charge of the same protein. A) Salt dependent end-to-end distance data (red points) is well described by the best fit counterion model (blue curve). Fitted parameters are: $\tilde{p} = 0.55$, $\delta = 1.3$, $\omega_2 = 1.275$. Other model parameters are: $\rho = 1 \times 10^{-6}$ mol / L ($\tilde{p} \approx 3.3 \times 10^{-8}$), $\ell_B = 7.12$ Å ($T = 20^\circ\text{C}$) ($\tilde{\ell}_B \approx 1.87$), and $\omega_3 = 0.1$. B) The best fit model predicts net charge (blue line) as a function of salt and is compared against effective charge measured at a single low salt concentration (19) (magenta point). The best fit model with no-condensation/full ionization (orange curve) over estimates both size at low salt concentration and charge. (See Supplementary Material for details on estimation of model parameters; sequences in Table S1 and fits for different choices of ω_3 in Figs. S2 and S3).

Glutamic acids and 25 Lysines were patterned in different order (see [Supplementary Material](#) and [Table S2](#) for details). We calculate degrees of ionization and end-to-end distance across a range in ℓ_B (inversely proportional to temperature) for these sequences. We make three important observations. First, sequences are fully ionized at low ℓ_B (high temperature) due to translational entropy, consistent with earlier studies (42). Subsequent increase in ℓ_B (decrease in temperature) reduces ionization, as ion pair formation is rewarded at the cost of entropy loss (see [Fig. 3](#)). Note that ionization will not approach zero even at very large ℓ_B (very low temperature), due to charge-dipole attraction which favor partial ionization.

Next, at moderate $\ell_B \approx 7 \text{ \AA}$ (near room temperature) we find that sequences with higher charge segregation (such as sv30) are more ionized compared to well mixed sequences (sv1 and others). Charge segregation in sv30 causes strong attraction (reflected in highly negative SCD) between oppositely charged blocks within the chain when amino acids are fully ionized. This intra-chain attraction is stronger than dipolar attractions that would form under significant condensation (see [Fig. S4](#)). Consequently, for well segregated block sequences, ionization is favored over condensation. Such sequences are also most compact due to this strong charge-charge attraction. In well mixed sequences (such as sv1), full ionization leads to substantially weaker overall charge-charge attraction (low SCD) due to partial cancelation by repulsion of like charges. As a result, well mixed sequences reduce enthalpy by forming more dipoles with solution ions enabling charge-dipole and dipole-dipole attractions. However, these dipolar attractions are weaker than charge-charge attraction in sv30, so well mixed sequences have larger chain dimensions compared to sv30 (see [Fig. 3](#)). We also verified the generality of these findings at other salt concentrations (see [Fig. S5](#)), although ionization and dimension become less sensitive to sequence patterning due to screening of electrostatics. Additionally, we varied ω_2 (equivalent of tuning the degree of hydrophobicity) to modulate R_{ee} while keeping the same charge patterning, and notice compact sequences tending to have higher condensation (see [Fig. S6](#)), possibly due to dipole and hydrophobicity induced enthalpic gain overcoming entropic loss due to collapse.

Finally, we find unequal degrees of ionization for positive and negative charges, $\alpha_+ \neq \alpha_-$, for all the sequence (except sv1 and sv30). This is surprising since each sequence has equal composition of positive and negative charges leading to the expectation

that the free energy should be invariant under charge reversal. However, the source of this symmetry breaking can be understood by focusing on the electrostatic interaction term between like charges and opposite charges that separately couple with the degrees of ionization giving rise to a special symmetry in patterning which is preserved only in sv1 and sv30 (See [Supplementary Material](#), discussion of eqs. S7–S11, for a detailed explanation).

Choice of phosphosites affects degree of ionization and chain dimension in IDPs

To probe the effects of sequence patterning beyond toy sequences, we consider an intrinsically disordered region of a protein (Uniprot ID P0A8H9) inhibiting binding of DNA gyrase to DNA. This protein was previously studied using all-atom simulation and a model assuming full ionization (30). The unmodified sequence has 7 positive and 12 negative charges. Two different sequence variants, S54S56 and S2T15, were created from two-site phosphorylation-mimic modifications (without actually adding a phosphate group) in which two S/T residues have been replaced by E resulting in 14 negative charges. These two sequence variants were selected because they have maximum and minimum SCD among all two-site modifications. Our earlier work (30) supported by all-atom Monte Carlo simulation (43, 44) has shown this is an effective strategy to detect sequences with maximum difference in size, since SCD serves as a reasonable indicator of electrostatic contribution to size, assuming other effects are negligible. These sequences also have different blockiness values (0.32 for S2T15, and 0.37 for S54S56) from the patterning metric proposed by Das and Pappu (7), supporting our choice of the phosphosites to generate appreciable difference in size. (See [Supplementary Material](#), [Table S3](#), for explicit sequences.)

We find the two phosphovariants, despite identical charge compositions, have markedly different sizes (see [Fig. 4](#)). In contrast to sv sequences discussed above, these two phosphovariants of P0A8H9 do not differ much in their degrees of ionization. This is because P0A8H9 variants are more polyelectrolyte like compared to sv sequences which are strong polyampholytes. Also, the modified chains with greater net charge (-7 ; in green and orange) are more expanded than the unmodified chain (net charge of -5 ; shown in blue) due to electrostatic repulsion. We conclude that sequences predicted to exhibit significant differences in size within full ionization models can maintain their differences even

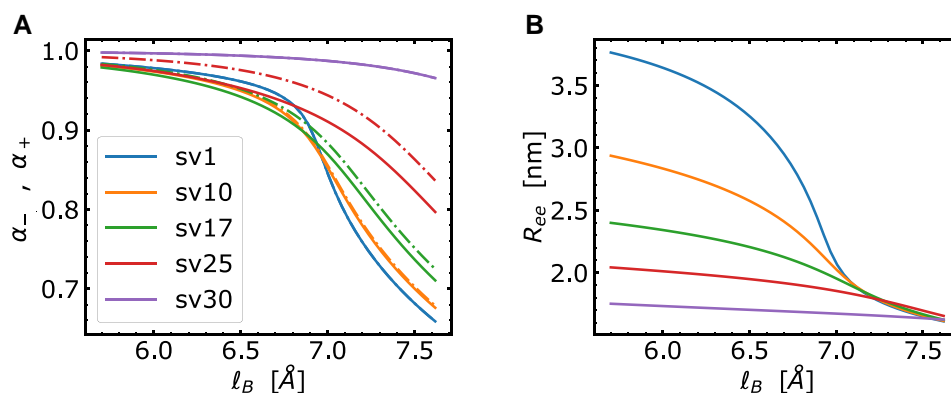


Fig. 3. Sequences with blocky charge patterns (sv30) are more ionized and compact compared to sequences with same charge composition where opposite charges are well mixed (sv1). A) Predicted degrees of ionization for negative (α_- ; solid lines) and positive charges (α_+ ; dashed lines), and B) end-to-end distances (R_{ee}), as functions of ℓ_B (relating temperature) shown for different toy sequences (different colors). We use parameters: $c_s \approx 1 \times 10^{-3} \text{ mol/L}$ ($\tilde{c}_s = 3 \times 10^{-5}$), $\rho \approx 5 \times 10^{-3} \text{ mol/L}$ ($\tilde{\rho} = 15 \times 10^{-5}$), with $\delta = 1.3$ and $\tilde{\rho} = 0.55$ using best fit values for Prothymosin-alpha ([Fig. 2](#)). Nonelectrostatic two-body and three-body interactions are $\omega_2 = 0$ and $\omega_3 = 0.1$. Ionization of positive and negative charges overlap only for sequences with special symmetry (sv1, sv30).

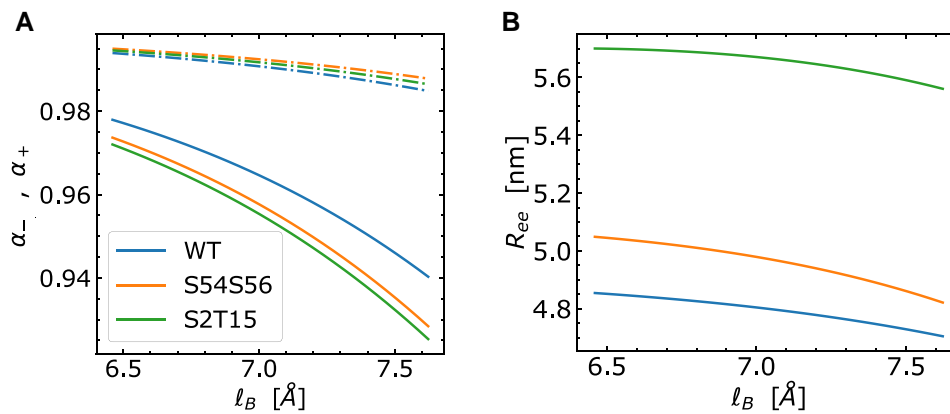


Fig. 4. Size and ionization are modulated by phosphorylation and specific choice of phosphosites, as demonstrated by IDP POA8H9 (WT), and its two phospho-mimic variants (S54S56 and S2T15) in which two S/T sites are replaced by E. A) Degrees of ionization for negative (solid lines) and positive charges (dashed lines) are significantly different at the same l_B for the same sequence. B) End-to-end distances as a function of l_B reveal strong dependence on charge patterning. WT is the most compact because its net charge is closest to neutrality; the phospho-mimic variants each have extra charge of -2 . Parameters are the same as in Fig. 3.

when charge condensation with dipole formation is included, although the extent of difference can depend on sequence properties.

Spontaneous appearance of coexisting phases in ionization and conformation landscape

While quantitative matching with Prothymosin- α data provide a reasonable estimate of the parameters (Fig. 2), they are not universal. Here, we again analyze POA8H9 and phosphorylation-mimic variants, with slightly different parameter choices; $\delta = 1.8$ and $\tilde{p} = 0.6$. The choice $\delta = 1.8$ is motivated by an estimate of dielectric constant in IDP condensate near $\epsilon_1 = 45$ (34), and $\tilde{p} = 0.6$ is close to the upper limit of the realistic estimate of the dipole length (see [Supplementary Material](#) for estimated range of realistic values). We have also increased the number of phosphosites to five allowing us to evaluate the effect of charge patterning from a large pool of sequences having same composition but different patterning. There are 792 possible sequences having five S/T replaced by E. We first focus on a subset of phospho-variants based on SCD, including the highest and lowest SCD along with two intermediates. (See [Supplementary Material](#), Table S3, for explicit sequences.)

The variant with the highest SCD (denoted as V1, green curve in Fig. 5) exhibits a first-order coil-globule transition with l_B (inversely proportional to temperature). Near this transition, two equally likely coexisting phases emerge, as is evident from the free energy landscape in ionization-conformation space showing two distinct and equal minima (see Fig. 6). One minimum corresponds to an expanded state with high ionization, while the other is a compact state with high condensation. Ionization/condensation are based on α_- , since the sequence has excess negative charges. The cooperative transition is due to sudden enthalpic decrease from a highly repulsive conformation (expanded and strongly ionized) to a compact state primarily stabilized by strong charge-dipole attraction. With increasing l_B (decreasing temperature) and consequently reduced ionization and entropy, this attractive interaction induces compaction when enough dipoles are formed while also retaining enough ionized charges. Charge-dipole induced stable conformations have also been seen experimentally in coacervation of polyzwitterion and polyelectrolytes (45). Our theory suggests that both charge-charge and charge-dipole

interactions are required for such a transition to occur (see Fig. S7 for quantitative support). The lowest SCD variant (V792, orange curve), however, maintains a single free energy minimum due to its reduced charge-charge repulsion in the strongly ionized expanded state. Consequently, it is unable to realize sudden enthalpic decrease upon compaction and its dimension changes smoothly with l_B (or temperature).

For the remaining 790 variants, we determined ionization-conformation states across l_B , checked if two coexisting states appeared, and recorded the transition point l_B (or temperature) along with the corresponding chain dimension (R_{ee}) and ionization (α_-) of the two states. Two such examples exhibiting first-order transitions (V186, V544) are depicted in Fig. 5. A schematic locus (black dotted curve) illustrates the coexistent phases in dimension and ionization across all variants (see Fig. S8 for the actual locus points). In general, variants with higher SCD exhibit transitions with a wider jump in dimension and ionization taking place at a higher l_B (lower temperature). Thus, SCD can be used to control the separation of coexisting phases. In summary, the appearance of coexisting phases and coil-globule transition depends on sequence details as quantified in terms of SCD and a new sequence patterning due to charge-dipole interaction embedded in SCDD purely due to ion condensation. The drastic difference in responses among phospho-mimic variants with identical charge composition highlights how phosphorylation can be used to create sequences with distinct conformational and charge states.

We further tested the influence of other parameters on the nature of such transitions. Specifically, we altered the nonelectrostatic two-body interaction parameter ω_2 (Figs. S9 and S10 with $\omega_2 = -0.4$, motivated by our earlier estimate from simulation (30)) that can be altered inside a cell due to changes in local environment. This choice of ω_2 enhances the separation in chain dimension between the two phases, and decreases the transition l_B (increases the temperature). This effect is akin to hydrophobicity induced cooperativity reported earlier (46).

Finally, we created 12 variants of the wild type POA8H9 where the degree of phosphorylation was gradually changed from 1 to 12 by adding phosphorylatable sites one at a time in the descending order of their probability to be phosphorylated. For example, variant n has the n most likely to be phosphorylated sites phosphorylated and the remaining $12 - n$ sites are unphosphorylated. Phosphorylation likelihood scores were predicted by a server (47, 48). We notice,

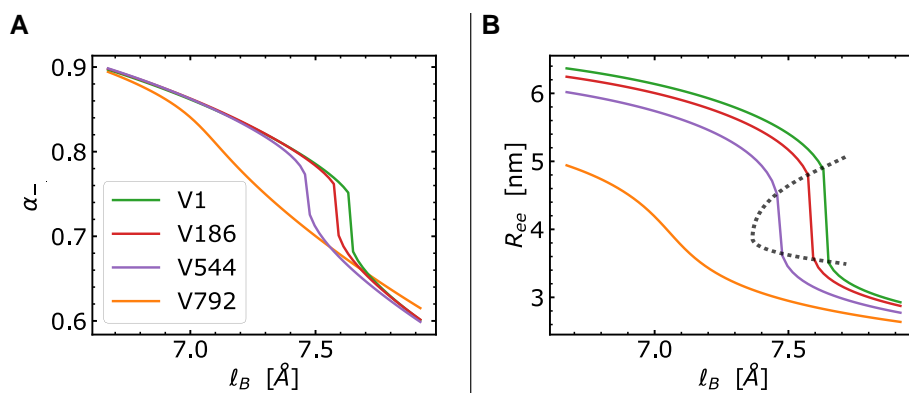


Fig. 5. Coexisting phases in conformation and charge appear for a subset of five site phosphovariants of POA8H9, marked by a first-order transition. Appearance of coexistence depends on specific sequence and SCD, illustrated by two extreme phosphovariants with SCD maximized (V1 in green with $\text{SCD} = 3.143$) and minimized (V792 in orange with $\text{SCD} = 1.141$), and two others with intermediate SCD (V186 and V544 in red and violet with $\text{SCD} = 2.878$ and 2.484 , respectively). A) Degree of ionization (α_-) for negative charges as a function of l_B show strong sequence dependence compared to that of the positive charges (not shown since it varies little across sequence and temperature). B) End-to-end distances as a function of l_B show strong dependence on charge patterning. The first-order transition occurs in both charge and conformation for the maximum and intermediate SCD variants. Dotted black curves sketch the region where this transition occurs across all five site variants. Parameters are $\tilde{p} = 0.6$, $\delta = 1.8$, with others kept as in Fig. 3.

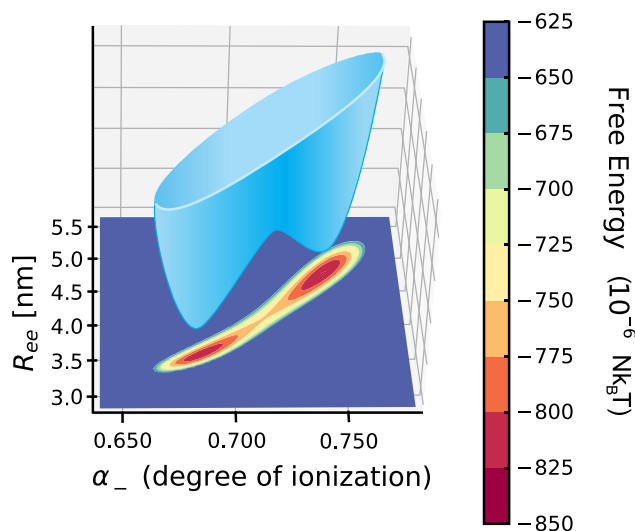


Fig. 6. Free energy landscape in ionization–conformation space reveals coexistence of two phases for the maximum SCD five site phosphovariant of POA8H9 (V1, green curve in Fig. 5). The 3D landscape in blue schematically illustrates the two basins, while the planar projection onto charge (α_-) and size (R_{ee}) shows the free energy quantitatively (in colored heat map) very near $T = 0^\circ\text{C}$ ($l_B = 7.64 \text{ \AA}$). The degree of ionization of the less prevalent charge is held fixed at $\alpha_+ = 0.961$ because it does not vary appreciably near the coexistence point. A value of 8.272 has been added to shift the reference free energy for presentation.

bistability appears beyond a critical degree (seven sites) of phosphorylation and the most likely phospho-variant (with nine sites phosphorylated) is predicted to exhibit bistability (see [Supplementary Material Fig. S11, Table S4](#)). We conclude that the appearance of two coexisting phases depends on sequence regulated by chemical modification such as phosphorylation and environment, offering a rich set of tools at the disposal of biology to tune ionization and conformation accordingly.

Salt induced conformational reentrance

Salt being a well-known regulator of IDP dimensions ([2, 10, 31, 49–51](#)), we calculate its effect on the five site phospho-mimic variants of POA8H9 (Fig. 7). Both the maximum and minimum SCD variants

(V1, V792) exhibit a reentrant behavior with a nonmonotonic dependence on c_s . The size first sharply decreases with c_s at low c_s , and then modestly increases at higher c_s , displaying the re-entrance phenomenon.

This reentrance is not due to salting out effect ([50](#)) but rather due to dipolar interactions giving rise to two distinct mechanisms. For both sequences, ionization is very high at very low salt concentration. In this regime, the dominant contribution to electrostatics is the repulsive charge–charge interaction. Screening of this repulsive interaction leads to chain compaction as salt is increased in the vicinity of zero salt. The addition of salt in general decreases ionization (increases condensation), which can have two outcomes. First, it can reduce the magnitude of the repulsive charge–charge interaction (green curve left panel in Fig. 7). Second, the charge–charge interaction may even become attractive due to asymmetric condensation of opposite charges (orange curve in Fig. 7). Addition of salt near the low salt limit also gives rise to dipolar attraction. For the high SCD variant (V1; green curve), this dipolar attraction eventually overcomes the antagonistic charge–charge interaction (repulsive) at moderate salt concentration. Then with further addition of salt, this dominant attractive dipolar interaction is screened and causes chain expansion. (See Fig. [S12](#), for quantitative support for varying contribution of charge–charge and dipolar interactions for these two variants.)

On the other hand, for the low SCD variant (V792; orange curve), charge–charge interaction becomes attractive at moderate salt. The synergistic effect of all three attractive interactions (charge–charge, charge–dipole, dipole–dipole) at the moderate salt concentration makes the chain more compact than V1. These attractive interactions will also get screened with further addition of salt, again leading to chain expansion. Note that the steep drop in conformation seen in Fig. 7 is not indicative of phase coexistence or first-order coil-globule transition. However, coexisting phases and corresponding coil-globule transition can appear as salt is varied under other conditions (see Figs. [S13](#) and [S14](#)).

Discussion

We present a sequence-dependent self-consistent theory of counterion condensation and chain dimension of IDPs and polyampholytes carrying both positive and negative charges, by

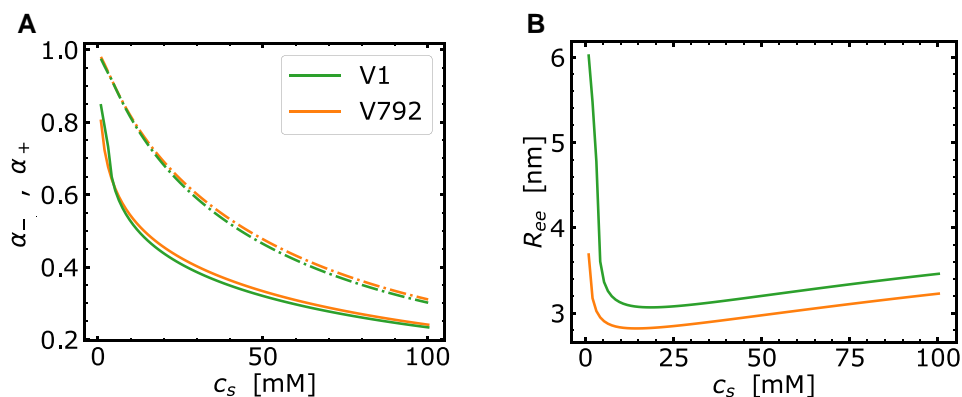


Fig. 7. Condensation and dipoles drive compaction to swelling transition with salt for phospho-mimic variants of POA8H9. A) Degrees of ionization for negative (solid lines) and positive (dashed lines) charges and B) end-to-end distance, as functions of salt for the two variants with five phosphosites (high SCD variant V1 in green, low SCD variant V792 in orange). Parameters are $\bar{p} = 0.6$, $\delta = 1.8$, $\omega_2 = 0$, and $\ell_B = 7.12$ Å ($T = 20^\circ\text{C}$); other parameters identical to those in Fig. 5. Sequence dependence and charge asymmetry are apparent, yet all degrees of ionization decrease monotonically with increasing salt. Conformational reentrance is due to electrostatic interactions changing from repulsive to attractive.

accounting for dipolar interactions going beyond monopole-charge effects. Theory shows that sequence charge patterning significantly influences both ionization and chain dimension. Furthermore, we show that some sequences exhibit two coexisting states in the ionization-size landscape. The calculated free energy barriers between these states in our model are low and may be blurred due to shot noise in the FRET histogram, at the same time reconfiguration times can be faster than the time scales measurable in single-molecule experiments. However, for certain designable sequences, the protein can readily undergo back-and-forth exchange between two drastically different conformations and charges, stimulating future experiments.

The ability to drastically modulate ionization and conformation by altering sequence, keeping same charge composition, may provide insights to the role of chemical modifications (such as phosphorylation), at specific sites and in specific order, acting as a regulatory switch (52). Ability to transition between two conformations is also relevant for intrinsically disordered regions (IDR), serving as a linker between two folded domains. Conformation fluctuations of IDRs can determine accessibility of these folded domains interacting with other domains or self-interacting with parts of the IDR (36, 51). Such interactions can synergistically combine with other switch-like mechanisms, for example multisite phosphorylation or gene expression.

One caveat of the present model is, it assumes charges of same type are equally ionized, independent of their position in the sequence. A fine-grained ionization landscape, considering all ionizable sites separately, will be high dimensional and may even exhibit multiphase coexistence. Recently, Pappu group reported distinct stable states in conformation by accounting for solvation differences between similarly charged residues, e.g. Lysine vs Arginine in atomic simulations of IDPs (53). Such differences in hydration would also be present between cations and anions, and between ionizable residues and solution ions (54). Incorporating fluctuations in ionization and residue/ion dependent solvation effects in the current model may provide further insights to conformational switching. Our model for an IDP (in infinitely dilute solution) will be relevant to understand solution behavior of IDPs undergoing liquid-liquid phase separation (LLPS), since single-chain behavior often provides insights to multichain solution physics (55–57). Ion condensation, currently neglected in most modeling approaches (58–60), will renormalize long-range electrostatics used to model LLPS. Concurrently,

formation of dipoles—akin to stickers—will contribute to the free energy of condensates and modify the overall phase diagram. Ion condensation will also impact complexation (33, 61–65) and phase separation (66–72) between oppositely charged macromolecules. Dipole driven salt induced reentrance predicted by our model may also give rise to salt-dependent reentrance, seen in LLPS of IDPs (73). Dipolar interactions give rise to new sequence dependent patterning metrics such as SCDD, SDD that will add to exciting ongoing efforts to relate IDP sequence and function (37, 39, 40, 74–76) with implications in protein design and evolution.

Acknowledgments

We thank Prof. Ben Schuler and Dr Aritra Chowdhury for providing valuable feedback on the manuscript and many stimulating discussion over the years.

Supplementary Material

Supplementary material is available at PNAS Nexus online.

Funding

This work was supported by National Science Foundation Grant No. DMR 2213103 (to M.P.), Air Force Office of Scientific Research Grant No. FA9550-23-1-0584 (to M.M.), and National Institute of Health R01GM138901 (to K.G.).

Author Contributions

M.P.: Methodology, Formal Analysis, Software, Investigation, Visualization, Validation, Data Curation, Writing - Original Draft, Writing - Review and Editing, Funding Acquisition. M.M.: Conceptualization, Supervision, Visualization, Writing - Original Draft, Writing - Review and Editing, Funding Acquisition. K.G.: Conceptualization, Supervision, Project Administration, Resources, Visualization, Writing - Original Draft, Writing - Review and Editing, Funding Acquisition.

Data Availability

All the generated data are in the article and Supplementary Material file. Code is available at the link, [charge regulation GitHub](#).

References

- Marsh J, Forman-Kay J. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys J.* 98(10):2383–2390.
- Muller-Spath S, et al. 2010. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci U S A.* 107(33):14609–14614.
- Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. 2010. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci U S A.* 107(18):8183–8188.
- Hofmann H, et al. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci U S A.* 109(40):16155–16160.
- Sizemore SM, Cope SM, Roy A, Ghirlanda G, Vaiana SM. 2015. Slow internal dynamics and charge expansion in the disordered protein CGRP: a comparison with amylin. *Biophys J.* 109(5):1038–1048.
- Srivastava D, Muthukumar M. 1996. Sequence dependence of conformations of polyampholytes. *Macromolecules.* 29(6):2324–2326.
- Das RK, Pappu RV. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci U S A.* 110(33):13392–13397.
- Sawle L, Ghosh K. 2015. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J Chem Phys.* 143(8):085101.
- Das RK, Ruff KM, Pappu RV. 2015. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol.* 32:102–112.
- Samanta HS, Chakraborty D, Thirumalai D. 2018. Charge fluctuation effects on the shape of flexible polyampholytes with applications to intrinsically disordered proteins. *J Chem Phys.* 149:163323.
- McCarty J, Delaney KT, Danielson SPO, Fredrickson GH, Shea JE. 2019. Complete phase diagram for liquid-liquid phase separation of intrinsically disordered proteins. *J Phys Chem Lett.* 10(8):1644–1652.
- Baul U, Chakraborty D, Mugnai ML, Straub JE, Thirumalai D. 2019. Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *J Phys Chem B.* 123(16):3462–3474.
- Zheng W, Dignon G, Brown M, Kim YC, Mittal J. 2020. Hydrophobic patterning complements charge patterning to describe conformational preferences of disordered proteins. *J Phys Chem Lett.* 11(9):3408–3415.
- Rumyantsev A, Johnner A, Pablo J. 2021. Sequence blockiness controls the structure of polyampholyte necklaces. *ACS Macro Lett.* 10(8):1048–1054.
- Dinic J, Schnorenberg M, Tirrell M. 2022. Sequence-controlled secondary structures and stimuli responsiveness of bioinspired polyampholytes. *Biomacromolecules.* 23(9):3798–3809.
- Muthukumar M. 2016. Ordinary-extraordinary transition in dynamics of solutions of charged macromolecules. *Proc Natl Acad Sci.* 113(45):12627–12632.
- Muthukumar M. 2017. 50th anniversary perspective: a perspective on polyelectrolyte solutions. *Macromolecules.* 50(24):9528–9560.
- Fossat M, Pappu R. 2019. q-Canonical Monte Carlo sampling for modeling the linkage between charge regulation and conformational equilibria of peptides. *J Phys Chem B.* 123(32):6952–6967.
- Ruggeri F, et al. 2017. Single-molecule electrometry. *Nat Nanotechnol.* 12(5):488–495.
- Fossat M, Posey A, Pappu R. 2021. Quantifying charge state heterogeneity for proteins with multiple ionizable residues. *Biophys J.* 120(24):5438–5453.
- Fossat M, Posey A, Pappu R. 2023. Uncovering the contributions of charge regulation to the stability of single alpha helices. *Chem Phys Chem.* 24(7):e202200746.
- King M, et al. 2024. Macromolecular condensation organizes nucleolar sub-phases to set up a pH gradient. *Cell.* 187(8):1889–1906.
- Fuoss R, Katchalsky A, Lifson S. 1951. The potential of an infinite rod-like molecule and the distribution of the counter ions. *Proc Natl Acad Sci U S A.* 37(9):579–589.
- Manning G. 1969. Limiting laws and counterion condensation in polyelectrolyte solutions. I. Colligative properties. *J Chem Phys.* 51(3):924–933.
- Oosawa F. 1971. *Polyelectrolytes.* New York: DEKKER.
- Netz R, Orland H. 2003. Variational charge renormalization in charged systems. *Eur Phys J E.* 11(3):301–311.
- Muthukumar M. 2004. Theory of counter-ion condensation on flexible polyelectrolytes: adsorption mechanism. *J Chem Phys.* 120(19):9343–9350.
- Kundagrami A, Muthukumar M. 2008. Theory of competitive adsorption of flexible polyelectrolytes: divalent salts. *J Chem Phys.* 128(24):244901.
- Muthukumar M. 2023. *Physics of charged macromolecules: synthetic and biological systems.* Cambridge University Press.
- Firman T, Ghosh K. 2018. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *J Chem Phys.* 148(12):123305.
- Huihui J, Firman T, Ghosh K. 2018. Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. *J Chem Phys.* 149(8):085101.
- Huihui J, Ghosh K. 2020. An analytical theory to describe sequence-specific inter-residue distance profiles for polyampholytes and intrinsically disordered proteins. *J Chem Phys.* 152(16):161102.
- Chowdhury A, et al. 2023. Driving forces of the complex formation between highly charged disordered proteins. *Proc Natl Acad Sci U S A.* 120(41):e2304036120.
- Nott TJ, et al. 2015. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol Cell.* 57(5):936–947.
- Zarin T, Tsai CN, AM Moses ANNBa. 2017. Selection maintains signaling function of a highly diverged intrinsically disordered region. *Proc Natl Acad Sci.* 114(8):E1450–E1459.
- Sherry KP, Das RK, Pappu RV, Barrick D. 2017. Control of transcriptional activity by design of charge patterning in the intrinsically disordered ram region of the notch receptor. *Proc Natl Acad Sci U S A.* 114(44):E9243–9252.
- Zarin T, et al. 2019. Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *Elife.* 8:46883.
- Huihui J, Ghosh K. 2021. Intrachain interaction topology can identify functionally similar intrinsically disordered proteins. *Biophysical J.* 120(10):1860–1868.
- Cohan M, Shinn M, Lalmansingh J, Pappu R. 2022. Uncovering non-random binary patterns within sequences of intrinsically disordered proteins. *J Mol Biol.* 434:167373.
- Tesei G, et al. 2024. Conformational ensembles of the human intrinsically disordered proteome. *Nature.* 626:897–904. <https://doi.org/10.1038/s41586-023-07004-5>.
- Ghosh K, Huihui J, Phillips M, Haider A. 2022. Rules of physical mathematics govern intrinsically disordered proteins. *Annu Rev Biophys.* 51:355–376.

- 42 Peng B, Muthukumar M. 2015. Modeling competitive substitution in a polyelectrolyte complex. *J Chem Phys.* 143:243133.
- 43 Vitalis A, Pappu RV. 2009. Absinth: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J Comput Chem.* 30(5):673–699.
- 44 Vitalis A, Pappu R. 2009. Methods for monte carlo simulations of biomacromolecules. *Annu Rep Comput Chem.* 5:49–76.
- 45 Margossian K, Brown M, Emerick T, Muthukumar M. 2022. Coacervation in polyzwitterion-polyelectrolyte systems and their potential applications for gastrointestinal drug delivery platforms. *Nat Commun.* 13(1):2250.
- 46 Kundagrami A, Muthukumar M. 2010. Effective charge and coil-globule transition of a polyelectrolyte chain. *Macromolecules.* 43(5):2574–2581.
- 47 Blom N, Gammeltoft S, Brunak S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol.* 294(5):1351–1362.
- 48 Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics.* 4(6):1633–1649.
- 49 Vancraenenbroeck R, Harel Y, Zheng W, Hofmann H. 2019. Polymer effects modulate binding in disordered proteins. *Proc Natl Acad Sci U S A.* 116(39):19506–19512.
- 50 Wohl S, Jakubowski M, Zheng W. 2021. Salt-dependent conformational changes of intrinsically disordered proteins. *J Phys Chem Lett.* 12(28):6684–6691.
- 51 Wiggers F, et al. 2021. Diffusion of a disordered protein on its folded ligand. *Proc Natl Acad Sci U S A.* 118(37):e2106690118.
- 52 Bah A, et al. 2015. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature.* 519:106–109.
- 53 Zeng X, Ruff K, Pappu R. 2022. Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins. *Proc Natl Acad Sci U S A.* 119: e2200559119.
- 54 Fossat M, Zeng X, Pappu R. 2021. Uncovering differences in hydration free energies and structures for model compound mimics of charged side chains of amino acids. *J Chem Phys B.* 125:4148–4161.
- 55 Lin YH, Chan HS. 2017. Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. *Biophys J.* 112:2043–2046.
- 56 Dignon GL, Zheng W, Best RB, Kim YC, Mittal J. 2018. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc Natl Acad Sci U S A.* 115:9929–9934.
- 57 Zeng X, Holehouse A, Chilkoti A, Mittag T, Pappu R. 2020. Connecting coil-to-globule transitions to full phase diagrams for intrinsically disordered proteins. *Biophys J.* 119:402–418.
- 58 Lin YH, Forman-Kay JD, Chan HS. 2016. Sequence-specific polyampholyte phase separation in membranes organelles. *Phys Rev Lett.* 117:178101.
- 59 Danielson SPO, McCarty J, Shea JE, Delaney KT, Fredrickson GH. 2019. Molecular design of self-coacervation phenomena in block polyampholytes. *Proc Natl Acad Sci U S A.* 116(17):8224–8232.
- 60 Lin Y, Brady J, Chan HS, Ghosh K. 2020. A unified analytical theory of heteropolymers for sequence specific phase behaviors of polyelectrolytes and polyampholytes. *J Chem Phys.* 152:045102.
- 61 Kudlay A, Ermoshkin A, delaCruz O. 2004. Complexation of oppositely charged polyelectrolytes. *Macromolecules.* 37(24):9231–9241.
- 62 Borgia A, et al. 2018. Extreme disordered in an ultrahigh-affinity protein complex. *Nature.* 555(7694):61–66.
- 63 Silva F, Derreumaux P, Pasquali S. 2018. Protein-RNA complexation driven by the charge regulation mechanism. *Biochem Biophys Res Commun.* 498:264–273.
- 64 Andreev M, Prabhu V, Douglas J, Tirrell M, Pablo JJD. 2018. Complex coacervation in polyelectrolytes from a coarse-grained model. *Macromolecules.* 51:6717–6723.
- 65 Amin AN, Lin YH, Das S, Chan H. 2020. Analytical theory for sequence-specific binary fuzzy complexes of charged intrinsically disordered proteins. *J Phys Chem B.* 124:6709–6720.
- 66 Pak CW, et al. 2016. Sequence determinants of intracellular phase separation by complex coacervation of a disordered proteins. *Mol Cell.* 63:72–85.
- 67 Shen K, Wang ZG. 2018. Polyelectrolyte chain structure and solution phase behavior. *Macromolecules.* 51:1706–1717.
- 68 Adhikari S, Leaf MA, Muthukumar M. 2018. Polyelectrolyte complex coacervation by electrostatic dipolar interactions. *J Chem Phys.* 149:163308.
- 69 Radhakrishna M, et al. 2017. Molecular connectivity and correlation effects on polymer coacervation. *Macromolecules.* 50: 3030–3037.
- 70 Lin YH, Brady JP, Forman-Kay JD, Chan HS. 2017. Charge pattern matching as a “fuzzy” mode of molecular recognition for the functional phase separations of intrinsically disordered proteins. *New J Phys.* 19:115003.
- 71 Li L, et al. 2018. Phase behavior and salt partitioning in polyelectrolyte complex coacervates. *Macromolecules.* 51:2988–2995.
- 72 Chang LW, et al. 2017. Sequence and entropy-based control of complex coacervates. *Nat Commun.* 8:1273.
- 73 Lin Y, et al. 2024. Electrostatics of salt-dependent reentrant phase behaviors highlights diverse roles of ATP in biomolecular condensates. arXiv:2401.04873, preprint: not peer reviewed. <https://doi.org/10.48550/arXiv.2401.04873>
- 74 Cohan M, Ruff K, Pappu R. 2019. Information theoretic measures for quantifying sequence-ensemble relationships of intrinsically disordered proteins. *Protein Eng Des Sel.* 32(4):191–202.
- 75 Griffith D, Holehouse A. 2021. Parrot is a flexible recurrent neural network framework for analysis of large protein datasets. *eLife.* 10:e70576.
- 76 Lu A, et al. 2022. Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning. *PLoS Comput Biol.* 18(6):e1010238.