ADVANCED SCIENCE Open Access
www.advancedscience.com

Mapping Charge Interactions in Intrinsically Disordered Proteins

Michael Phillips, Andrea Holla, Magdalena Wojtas, Aritra Chowdhury, Andrea Sottini, Sebastian L. B. König, Natalie Mutter, Nick Lamb, Jonathan Huihui, Monika Lopko, Andrea Soranno, Daniel Nettels, Andrzej Ożyhar, Benjamin Schuler, and Kingshuk Ghosh**

Intrinsically disordered proteins (IDPs) are often rich in charged residues, and electrostatic interactions have a pronounced effect on their conformational distributions, interactions and functions. However, attaining quantitative understanding of electrostatics is challenging because of the sequence-specific arrangement of charges in the chain, the long-range nature of electrostatic interactions, charge screening, and the condensation of counterions—effects that all need to be taken into account self-consistently. Here, analytically tractable quantitative models are developed to predict ensemble average distances between any pair of residues in IDPs as a function of sequence and salt concentration, explicitly considering charge patterning. These models are tested systematically against extensive single-molecule Förster resonance energy transfer (FRET) data mapping intrachain distances for a range of charged IDPs with different sequence compositions, as a function of salt concentration, and with different labeling positions and fluorophores. The resulting polymer model with a minimal set of adjustable parameters accounts for counterion condensation, the resulting effective charges, as well as dipolar interactions, and can be used to predict detailed intrachain distance maps between all residues. Analytical models of this kind offer a valuable complement to simulations and can provide fundamental insight into the interactions underlying the conformational distributions of IDPs.

1. Introduction

Intrinsically disordered proteins (IDPs) are enriched in charged amino acids,[1] and electrostatics thus play a key role in modulating their conformations and functions. Several aspects of electrostatics can affect IDP behavior. First, the number, type (positive or negative), and position of charges within the sequence influence the conformational ensembles of IDPs.[2–18] Second, changes in salt concentration can screen electrostatics (attractive or repulsive) in a sequence-dependent manner.[10,19] For example, screening the charge repulsion in polyelectrolytes leads to chain compaction, whereas screening charge attraction in polyampholytes can lead to chain expansion,[3,20–23] and two sequences with the same charge composition but different charge patterning can exhibit markedly different response under identical changes in ionic strength.[19,24] Third, the effective charges in the polypeptide chain can be variable, contrary to the typical assumption that the side chains are fully ionized at neutral pH. The origin of this variable

M. Phillips, N. Lamb, J. Huihui, K. Ghosh
Department of Physics and Astronomy
University of Denver
Denver, CO 80208, USA
E-mail: kingshuk.ghosh@du.edu

A. Holla, M. Wojtas, A. Chowdhury, A. Sottini, S. L. B. König, N. Mutter, A. Soranno, D. Nettels
Department of Biochemistry
University of Zurich
Zurich 8057, Switzerland

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/advs.202514056

DOI: 10.1002/advs.202514056

M. Wojtas, A. Chowdhury, M. Lopko, A. Ożyhar
Department of Biochemistry
Molecular Biology and Biotechnology
Wrocław University of Technology
Wrocław 50-373, Poland

A. Soranno
Department of Biochemistry and Molecular Biophysics, Center for Biomolecular Condensates
Washington University in St. Louis
St. Louis, MO 63110, USA

B. Schuler
Department of Biochemistry and Department of Physics
University of Zurich
Zurich 8057, Switzerland
E-mail: schuler@bioc.uzh.ch

charge state is that ionizable residues are in an equilibrium between two states: fully ionized and complexed with oppositely charged ions.[22,25–28] This phenomenon is often termed charge regulation for protonation equilibria, or charge renormalization for the interaction with other counterions from salt in solution. In both cases, the net result is that the effective charge of ionizable residues in an IDP can differ from the nominal charge commonly assigned to them.[27–32] The difference, often quantified by an effective degree of ionization, will depend on additional factors, such as the presence of other charges in the sequence, the dielectric constant in the vicinity of the chain, solution conditions (e.g., pH, ionic strength), and also on the conformation of the chain, which in turn depends on the sequence and degree of ionization. To answer the question of how these factors are coupled and how they influence IDP conformations, we need an approach integrating quantitative measurements and models.

On the experimental side (see Sections S1 and S2, Supporting Information), we used single-molecule Förster resonance energy transfer (FRET)[33–35] to measure end-to-end and intrachain distances across 18 different disordered protein sequences with different compositions and patterning that cover a diverse sequence space (**Figure 1**). A set of 16 naturally occurring intrinsically disordered regions (IDRs) of identical length but with large differences in amino acid composition, hydrophobicity, and charge content and patterning was previously selected from linkers found in RNA-binding proteins,[36] for which we probed the end-to-end distance (Table S1, Supporting Information). In a complementary set of samples, FRET dyes were placed at different positions in the highly negatively charged IDPs Starmaker (Stm)[37] and Prothymosin α (ProTα)[3,38] to probe intra-chain distances for different segments and segment lengths within those sequences (Figure 1; Table S1, Supporting Information). In addition to using this diverse set of sequences, we modulated the electrostatic interactions by probing the chain dimensions and intrachain distances across a broad range of concentrations of monovalent salt. The resulting data set thus provides rich information on electrostatic and non-electrostatic interactions in disordered proteins across a wide range of sequence compositions, chain lengths, and salt concentrations, making it ideally suited for developing and testing a quantitative framework that describes charge interactions in IDPs.

Toward this aim, we developed analytically tractable quantitative models based on the Edwards–Muthukumar Hamiltonian[39,40] to predict ensemble average distances between any pair of residues in IDPs as a function of sequence, explicitly considering charge patterning and ion condensation. Our theory includes sequence-specific electrostatic interactions between charges, as well as dipolar interactions (charge–dipole, and dipole–dipole) arising from ion-pairs formed between charges on the polypeptide chain and condensed counterions (Figure 1). These models were then optimized and tested against the large set of single-molecule FRET data. An advantage of analytical models is the ability to test hypotheses by varying the complexity of the models. Starting with the most complex model and subsequently reducing model complexity, we find that it is possible to describe the data by a minimal, near-predictive model. In this framework, sequence-specific non-electrostatic interactions are quantified from measurements at high salt concentration, where electrostatics are fully screened. Counterion interactions are described by two parameters: an effective separation (dipole length) between the side chain charge and condensed ion, and an effective dielectric in the vicinity of the chain. Both parameters can be learned from the data and used in a transferable manner to describe the majority of the sequences. In this formalism, effective ionization is due to counterion condensation and is highly dependent on the ionic strength and on the protein sequence. Taken together, our study provides a stringent test of models of IDPs with different charge content, patterning, and screening based on experimental data, which enable a quantitative description of the critical role charge interactions play in the conformations of IDPs.

2. Model

2.1. Models for Calculating the End-to-End Distance

We developed an analytical model of an IDP with N residues, and N_+ positive and N_- negative charges (upon full ionization). To model charge variation of residues, we assume that each charge type is ionized with a probability $\alpha_\pm \in [0, 1]$, yielding the effective charge $q_m = +\alpha_+$, $q_m = -\alpha_-$, or $q_m = 0$, depending on its classification (basic/acidic/neutral). We assume that charges of the same type (i.e., acidic or basic) have the same effective charge, irrespective of their location in the sequence. This mean-field assumption reduces model complexity and describes the data well, as will be seen in the Results section. We account for the charges of the N- and C-terminal amino and carboxyl groups, respectively, as well as the FRET dyes, by modifying the sequence appropriately; in one approach, dyes with charge -2 are incorporated by adopting the effective charge $q_m = -2\alpha_-$ (see Supporting Information Table S1 for specific sequences and Figure S1 and Section S3 for details of charge assignment for dyes and amino acids). The overall chain dimensions are given by a dimensionless swelling factor x defined as the ratio of the ensemble-averaged mean squared end-to-end distance, $\langle R_{ee}^2 \rangle$, and a Flory random coil reference state with mean squared end-to-end distance $Nb\ell$, where $b = 3.8$ Å is the bond length, and $\ell = 8.0$ Å is the Kuhn length (see References [13, 16] for details). The swelling factor $x = \langle R_{ee}^2 \rangle / Nb\ell$ and the effective charge weights (degrees of ionization) α_+, α_- depend on each other, so we require a self-consistent description. Thus, we construct a free energy F as a function of (x, α_+, α_-) with five contributions:[26] $F(x, \alpha_+, \alpha_-) = F_1(\alpha_+, \alpha_-) + F_2(\alpha_+, \alpha_-) + F_3(\alpha_+, \alpha_-) + F_4(\alpha_+, \alpha_-) + F_5(x, \alpha_+, \alpha_-)$. F_1, F_2 are the combinatorial and translational entropies of the ions bound to the chain and free in solution, respectively; F_3 is the fluctuation contribution of all ions; F_4 is the free energy (related to the equilibrium constant) of ion pair formation arising from each counter ion condensed with its oppositely charged partner on the chain, as in Ref. [26]; and F_5 gives the free energy of the chain arising from chain connectivity and intra-chain interactions. In Figure S2 (Supporting Information) and below we describe the details of these different free energy contributions.

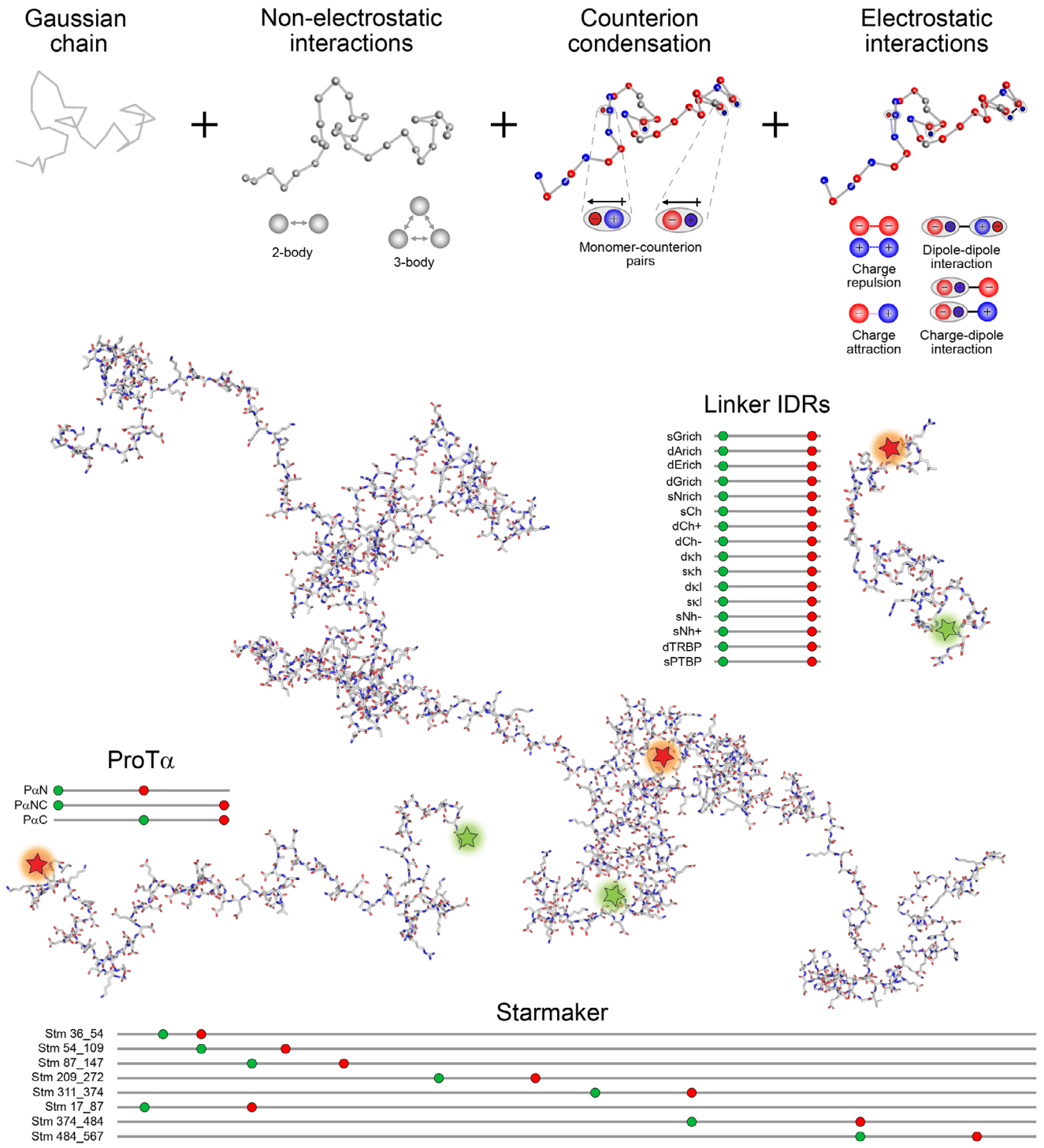

Figure 1. Illustration of the key components of the analytical theory used (top) and of the molecular systems investigated with single-molecule FRET (bottom) in an integrated approach to quantify ion condensation and its impact on the chain dimensions. In the theory, Gaussian chain statistics are combined with two-body and three-body excluded volume interactions. Condensed counterions on the charged side chains form dipoles (shaded ellipses) and give rise to charge-dipole, dipole-dipole, and charge-charge interactions. The theory accounts for these interactions for arbitrary sequences of positive (blue), negative (red) and neutral (grey) amino acids to predict ensemble-averaged distances ($\langle R_{ij}^2 \rangle$) between any residues i, j. A wide range of protein variants were used to parameterize the theory and test its predictions based on single-molecule FRET measurements probing intrachain distances. Structural representations of the different disordered proteins and schematics of the sequences and labeling positions (red and green) are shown. The structural representations were generated from the sequences using AlphaFold 3[41] and Afflecto.[42] The naming of the different linker IDRs reflects their sequence properties (see Table S1, Supporting Information, and Holla et al.[36] for details).

2.1.1. Details of Entropic and Fluctuation Contributions to the Free Energy

First, we describe the three contributions F_1, F_2 and F_3. The free energy F_1 corresponding to the combinatorial entropy of condensed counterions is given by

$$\frac{\beta F_1}{N} = f_+\left(\alpha_+ \ln(\alpha_+) + (1-\alpha_+)\ln(1-\alpha_+)\right) + f_-\left(\alpha_- \ln(\alpha_-) + (1-\alpha_-)\ln(1-\alpha_-)\right) \quad (1)$$

where we introduce the charge fractions $f_\pm = N_\pm/N$, and $\beta = 1/k_BT$. The free energy F_2 corresponding to the translational entropy from counterions free in solution is given by

$$\frac{\beta F_2}{N} = \left(f_+\alpha_+ + \tilde{c}_s/\tilde{\rho}\right)\ln(f_+\alpha_+\tilde{\rho} + \tilde{c}_s) + \left(f_-\alpha_- + \tilde{c}_s/\tilde{\rho}\right)\ln(f_-\alpha_-\tilde{\rho} + \tilde{c}_s) - \left(f_+\alpha_+ + f_-\alpha_- + 2\tilde{c}_s/\tilde{\rho}\right) \quad (2)$$

where $\tilde{\rho}$ and $\tilde{c}_s$ are the reduced concentrations of monomers and salt defined as $\tilde{\rho} \equiv \rho b^3$ and $\tilde{c}_s \equiv c_s b^3$, respectively. For example, given a concentration ρ in molar units (mol L^{-1}) and length b in Angstroms (Å), the reduced equivalent is obtained via $\tilde{\rho} = (6.023 \times 10^{-4}\, L\,\mathrm{mol}^{-1}/\,\text{Å}^3)\rho b^3$. We note that in our case, the contribution of the terms $f_+\alpha_+$ and $f_-\alpha_-$ is negligible for the counterion concentration, since the single-molecule experiments were performed at picomolar IDP concentrations, but we include the terms in the equations for generality. The electrostatic free energy F_3 from Debye–Hückel theory,[43] also known as the ion fluctuation contribution,[26] is

$$\frac{\beta F_3}{N} = -\frac{\kappa^3}{12\pi\rho} = -\frac{2\sqrt{\pi}}{3\tilde{\rho}}\,\tilde{\ell}_B^{3/2}\left[(f_+\alpha_+ + f_-\alpha_-)\tilde{\rho} + 2\tilde{c}_s\right]^{3/2} \quad (3)$$

where

$$\kappa^2 = 4\pi\ell_B\left[(f_+\alpha_+ + f_-\alpha_-)\rho + 2c_s\right] \quad (4)$$

with Bjerrum length $\ell_B = e^2/(4\pi\epsilon_0\epsilon k_BT)$, ϵ_0 is the vacuum permittivity, and $\epsilon = 80$ for water. We use the reduced Bjerrum length $\tilde{\ell}_B \equiv \ell_B/b$.

2.1.2. Free Energy of Ion Pair Formation

Next, we consider the free energy of ion pair formation along the chain given by F_4.[26,44] We write the energy of forming a single ion pair as an effective Coulomb interaction, $\beta\Delta E = -\ell_B\delta/d$. One may interpret the ratio δ/d as a single sequence-specific parameter; here, we keep the two parameters δ and d independent and emphasize the connection to the physical separation distance, which enables a consistent treatment for dipolar interactions (see ref. [28] for details). The effective separation distance, d, is defined by $1/d = [1/p + 1/(2\delta a_1) + 1/(2a_2)]$. Here, p is the physical separation distance, a_1 is the radius of the small counterion, and a_2 is the radius of the charged group of the side chain (see Chapter 5 of Ref. [45] for details). The charged side chain is much larger than the other length scales: $a_2 \gg a_1, p$. We expect that the physical separation is similar to the size of a counterion: $p \approx a_1$. Thus, we take the approximate form $1/d \approx (1/p)\,[1 + 1/(2\delta)]$. Upon counting the counterions condensed along the chain, and using the reduced dipole length $\tilde{p} = p/b$, and $\tilde{\ell}_B = \ell_B/b$, the total energy associated with ion pair formation is

$$\frac{\beta F_4}{N} = -\left[f_+(1-\alpha_+) + f_-(1-\alpha_-)\right]\frac{\tilde{\ell}_B}{\tilde{p}}\left(\delta + \frac{1}{2}\right) \quad (5)$$

2.1.3. Sequence-Dependent Free Energy of the Polymer Chain

Finally, the detailed form of the chain free energy F_5 is given by

$$\beta F_5 = \frac{3}{2}(x - \ln(x)) + \frac{\omega_3 B}{2}\left(\frac{3}{2\pi x}\right)^3 + 2\tilde{\ell}_B Q\left(\frac{3}{2\pi x}\right)^{1/2} + \Omega\left(\frac{3}{2\pi x}\right)^{3/2} \quad (6)$$

This free energy has four contributions: elastic entropy from chain connectivity (first term on the right hand side), three-body repulsive excluded volume with strength ω_3, two-body long-range electrostatics between two charges given by the Q term, and two-body non-electrostatic interactions (and other short-range interactions such as charge-dipole, dipole–dipole etc.) in Ω given by the last term. The three-body contribution is given by

$$B = \frac{1}{N}\sum_{l=3}^{N}\sum_{m=2}^{l-1}\sum_{n=1}^{m-1}\frac{(l-n)}{[(l-m)(m-n)]^{3/2}} \quad (7)$$

where l, m, n are the indices of residues participating in the three-body interaction. The three-body repulsive interaction is needed to prevent chain collapse in the case of strong intrachain attraction.[46,47] The sequence-dependent charge-charge electrostatic interactions with screening, described by Q, are defined as [19]

$$Q = \frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1} q_m q_n\,(m-n)^{1/2}\,A(\tilde{\kappa}^2 x(m-n)/6) \quad (8)$$

$$A(z) = 1 - \sqrt{\pi z}\,\exp(z)\,\mathrm{erfc}(\sqrt{z}) \quad (9)$$

where the effective charge on each residue is dictated by the charge weights, $q_m = \pm\alpha_\pm$. In the limit of zero screening and full ionization ($\alpha_\pm = 1$), Q reduces to the Sequence Charge Decoration (SCD) metric defined in prior work.[8,16] The inverse Debye screening length is given in reduced form by $\tilde{\kappa}^2 = 4\pi\tilde{\ell}_B\left[(f_+\alpha_+ + f_-\alpha_-)\tilde{\rho} + 2\tilde{c}_s\right]$, where densities of residues and salt ions are, in reduced form, $\tilde{\rho} \equiv \rho b^3$ and $\tilde{c}_s \equiv c_s b^3$.

The two-body interaction, given by Ω, has three distinct contributions. The first contribution is non-electrostatic, $\Omega_{\text{non-e}}$, given by

$$\Omega_{\text{non-e}} = \frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1}\omega_{m,n}(m-n)^{-1/2} \quad (10)$$

where $\omega_{m,n}$ is the interaction between amino acids m and n. Since the interaction parameters between individual residues are unknown, we use an effective interaction parameter $\omega_{2,ee}$ (to compute the end-to-end distance) defined as

$$\omega_{2,ee} = \left[\sum_{m=2}^{N}\sum_{n=1}^{m-1}\omega_{m,n}(m-n)^{-1/2}\right] / \left[\sum_{m=2}^{N}\sum_{n=1}^{m-1}(m-n)^{-1/2}\right] \quad (11)$$

Thus, $\omega_{2,ee}$ is a pairwise mean-field non-electrostatic interaction among all residues. Sequence specificity of $\omega_{2,ee}$ has been termed Sequence Hydropathy Decoration (SHD) and was used to model simulated chain dimensions.[13] More recently, it has been determined by a machine learning model using a large set of coarse-grained simulations.[48] With this definition, we get

$$\Omega_{\text{non-e}} = \omega_{2,ee}\frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1}(m-n)^{-1/2} \quad (12)$$

Two additional contributions to Ω are $\Omega_{\text{c-d}}$ and $\Omega_{\text{d-d}}$, resulting from directionally averaged charge–dipole and dipole–dipole interactions approximated as delta function potentials,[26,45] with

$$\Omega_{\text{c-d}} = \omega_{cd}\frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1}\left(c_m d_n + c_n d_m\right)(m-n)^{-1/2} \quad (13)$$

$$\Omega_{\text{d-d}} = \omega_{dd}\frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1} d_m d_n\,(m-n)^{-1/2} \quad (14)$$

where, for each residue, the charge state is determined by the unsigned weights (probabilities of ionization), $c_m = \alpha_{\pm}$, and the dipole state is dictated by the complementary weights (probabilities of condensation), $d_m = 1 - \alpha_{\pm}$. Non-ionizable residues do not contribute. The specific arrangement of charges in the sequence will make these contributions sequence-specific. Consequently, we get two new patterning metrics: Sequence Charge-Dipole Decoration ($SCDD = \Omega_{\text{c-d}}/\omega_{cd}$) and Sequence Dipole Decoration ($SDD = \Omega_{\text{d-d}}/\omega_{dd}$) as defined in ref. [28]. The magnitudes of the effective interaction (pseudopotential) arising from charge–dipole (ω_{cd}) and dipole–dipole (ω_{dd}) contributions are given by [26, 44, 45],

$$\omega_{cd} = -\frac{\pi}{3}\delta^2\tilde{\ell}_B^2\tilde{p}^2\exp(-2\tilde{\kappa})[2+\tilde{\kappa}] \quad (15)$$

$$\omega_{dd} = -\frac{\pi}{9}\delta^2\tilde{\ell}_B^2\tilde{p}^4\exp(-2\tilde{\kappa})[4+8\tilde{\kappa}+4\tilde{\kappa}^2+\tilde{\kappa}^3] \quad (16)$$

Thus, the salt concentration dependence of the chain dimensions is not only determined by screening of charge–charge interactions (given by Equation 8) but also by these dipole-induced patterning metrics.

2.1.4. Free Energy Minimization Scheme

The free energy thus depends on three adjustable parameters (Figure S2, Supporting Information): $\omega_{2,ee}$ accounting for sequence-specific non-electrostatic interactions that we cannot predict from first principles at present; δ, the dielectric mismatch between the bulk solvent and at a position in the vicinity of the chain; and p, the dipole length describing the distance between the charged group of the side chain and the neutralizing ion. All other parameters are known. By minimizing this free energy for given values of $\omega_{2,ee}$, p, and δ with respect to x, α_+, α_-, we self-consistently determine the optimal effective charge and ensemble-averaged end-to-end distance to compare against experimentally measured dye-dye distances. We determine $\omega_{2,ee}$ from the FRET measurements at high salt (0.4 M), where charge interactions are screened, and we use p and/or δ as adjustable parameters in the fit to the experimental data. We used this formalism recently for describing the salt-dependent end-to-end measurements of ProTα and its net charge.[28]

2.1.5. Different Model Variants Used

We use various forms of this model to identify the one that provides the best balance between describing the experimental FRET data and the simplicity of the model, especially the number of adjustable parameters. The first model is the simple case in which counter ion condensation is excluded; this case is recovered from the formulation above by setting $\alpha_{\pm} = 1$, and it does not require parameters p and δ. We call this the *full ionization model* and use the label M0. A related model is a mean-field polyampholyte model developed by Higgs and Joanny[49] (here referred to as "H-J") that does not adjust charge weights and is not sensitive to sequence charge patterning; we amend the original formulation with a three-body term as in the second term of Equation (6), with Equation (7), to prevent unphysical chain collapse (see Section S5 and Equations S1– S3, Supporting Information, for the details of the H-J model).

For the cases with counter ion condensation, the models differ by the method for determining p and δ. We find that a single fixed value of the dipole length, $p/b = 0.66$, is suitable for all sequences. However, δ is expected to depend on the local dielectric permittivity near the chain, which can depend on charge patterning and chain dimensions, and is therefore sequence-dependent. We thus define M1 as the case where δ is determined individually from chi-square fitting of the FRET-basefd end-to-end distances for each of the 16 linker IDRs previously selected for diversity in amino acid composition, hydrophobicity, and charge patterning[36] (Table S1, Supporting Information) as a function of salt concentration (while $\omega_{2,ee}$ was adjusted to match the high-salt data point). This procedure yields a linear trend relating δ to low-salt end-to-end distance (see Results). In a third model, M2, δ is determined from the low-salt end-to-end distances based on this linear trend (again adjusting $\omega_{2,ee}$), thus substantially reducing the number of fit parameters.

We do not explicitly model charge regulation, i.e., changes in the protonation state of the amino acids, since all experiments were performed near pH 7 and at salt concentrations above 50 mM. Under these conditions, the protonation states can be approximated based on the nominal pK_a values of the ionizable groups.[27,50] Thus, we neglect deviations that might occur at very low salt concentrations below the range we probe, and for histidine, which is present only in very low proportions in the

sequences we investigate. If charge reduction occurs due to both charge regulation and charge renormalization, our formalism, when fitted to the data, will yield a single set of mean-field parameter values that effectively account for both effects. More advanced variants of the theory that separate protonation effects and the complexation between charged residues and salt ions[51] could be developed based on suitable pH-dependent experimental data.

2.2. Model For Computing Intrachain Distance Maps

Based on the single-molecule FRET experiments, we have not only measured average end-to-end distances but also distances $\langle R_{ij}^2 \rangle$ between residues at positions i, j. Such experiments were performed for two IDPs, Stm[37] and ProTα,[3,38] which are much longer than the linker IDRs (see Table S1, Supporting Information, for details of the sequences and dye positions). To determine $\langle R_{ij}^2 \rangle$ in the context of our theory, we define residue pair (i, j)-specific swelling factors $x_{i,j}$ just like x defined above, and construct a free-energy function F that depends on x_{ij}, α_+, α_-. For this approach, F_1, F_2, F_3, and F_4 are defined as above, only the chain free energy F_5 is now constructed in the reaction coordinate x_{ij} as:

$$\beta F_5 = \frac{3}{2}(x_{ij} - \ln(x_{ij})) + \frac{\omega_3 T_{ij}}{2\,|i-j|}\left(\frac{3}{2\pi x_{ij}}\right)^3 + 2\tilde{\ell}_B Q_{ij}\left(\frac{3}{2\pi x_{ij}}\right)^{1/2} + \Omega_{ij}\left(\frac{3}{2\pi x_{ij}}\right)^{3/2} \quad (17)$$

where T_{ij} is the residue pair-specific three-body term defined in our earlier work,[16] and Q_{ij} is the patterning metric arising from charge–charge interactions specific to residue pair i, j, also known as the Sequence Charge Decoration Matrix (SCDM),[16] and defined as

$$Q_{i,j} = \frac{1}{(i-j)}\left[\sum_{m=j}^{i}\sum_{n=1}^{j-1} q_m q_n \frac{(m-j)^2}{(m-n)^{3/2}} A(\kappa^2 b^2 (m-n) x_{i,j}/6) + \sum_{m=j+1}^{i}\sum_{n=j}^{m-1} q_m q_n (m-n)^{1/2} A(\kappa^2 b^2 (m-n) x_{i,j}/6) + \sum_{m=i+1}^{N}\sum_{n=1}^{j-1} q_m q_n \frac{(i-j)^2}{(m-n)^{3/2}} A(\kappa^2 b^2 (m-n) x_{i,j}/6) + \sum_{m=i+1}^{N}\sum_{n=j}^{i} q_m q_n \frac{(i-n)^2}{(m-n)^{3/2}} A(\kappa^2 b^2 (m-n) x_{i,j}/6)\right] \quad (18)$$

with A defined in Equation (9) above.

The two-body interaction Ω_{ij} is also residue pair-specific and accounts for three contributions: 1) purely non-electrostatic interactions (also defined as Sequence Hydropathy Decoration Matrix or $SHDM$[16,52]); 2) sequence-dependent charge–dipole interactions (Sequence Charge–Dipole Decoration Matrix, $SCDDM$); and 3) sequence dipole-dipole interactions (sequence Dipole Decoration Matrix, $SDDM$). These patterning contributions are now dependent on the specific residue pair (i, j) and hence give rise to matrices generalizing the metrics defined for the end-to-end distance. In our previous work,[16] the $SHDM$ was derived as

$$SHDM_{i,j} = \frac{1}{(i-j)}\left[\sum_{m=j}^{i}\sum_{n=1}^{j-1}\omega_{m,n}\frac{(m-j)^2}{(m-n)^{5/2}} + \sum_{m=j+1}^{i}\sum_{n=j}^{m-1}\omega_{m,n}(m-n)^{-1/2} + \sum_{m=i+1}^{N}\sum_{n=j}^{i}\omega_{m,n}\frac{(i-n)^2}{(m-n)^{5/2}} + \sum_{m=i+1}^{N}\sum_{n=1}^{j-1}\omega_{m,n}\frac{(i-j)^2}{(m-n)^{5/2}}\right]. \quad (19)$$

As noted above, the $\omega_{m,n}$ values are not known a priori. However, as previously done, the entire contribution is written as a single parameter $\omega_{2,i,j}$, where $SHDM_{i,j}$ is normalized by the interaction-independent segmental part $\mathcal{N}_{i,j}$ (i.e., setting $\omega_{m,n} = 1$ in Equation 19). Specifically, we have

$$\omega_{2,i,j} = SHDM_{i,j}/\mathcal{N}_{i,j}. \quad (20)$$

and

$$\mathcal{N}_{i,j} = \frac{1}{(i-j)}\left[\sum_{m=j}^{i}\sum_{n=1}^{j-1}\frac{(m-j)^2}{(m-n)^{5/2}} + \sum_{m=j+1}^{i}\sum_{n=j}^{m-1}(m-n)^{-1/2} + \sum_{m=i+1}^{N}\sum_{n=j}^{i}\frac{(i-n)^2}{(m-n)^{5/2}} + \sum_{m=i+1}^{N}\sum_{n=1}^{j-1}\frac{(i-j)^2}{(m-n)^{5/2}}\right] \quad (21)$$

This definition is a generalization of $\omega_{2,ee}$ defined above (Equation 11) for the calculation of the end-to-end distance. In the limit of a homopolymer, where $\omega_{m,n}$ variation is zero, $\omega_{2,ee} = \omega_{2,i,j}$ will be a constant independent of i and j. Otherwise, $\omega_{2,i,j}$ will depend on i, j and will be sequence-specific. We determine $\omega_{2,i,j}$ by using the experimentally observed residue pair-specific chain dimensions $\langle R_{ij}^2 \rangle$ in the high-salt limit to determine the non-electrostatic interaction strength.

The sequence-dependent charge–dipole interaction written in matrix form (due to specific residue pairs), $SCDDM$, is defined as

$$SCDDM_{i,j} = \omega_{cd}\frac{1}{(i-j)}\left[\sum_{m=j}^{i}\sum_{n=1}^{j-1}(c_m d_n + c_n d_m)\frac{(m-j)^2}{(m-n)^{5/2}} + \sum_{m=j+1}^{i}\sum_{n=j}^{m-1}(c_m d_n + c_n d_m)(m-n)^{-1/2} + \sum_{m=i+1}^{N}\sum_{n=j}^{i}(c_m d_n + c_n d_m)\frac{(i-n)^2}{(m-n)^{5/2}} + \sum_{m=i+1}^{N}\sum_{n=1}^{j-1}(c_m d_n + c_n d_m)\frac{(i-j)^2}{(m-n)^{5/2}}\right]. \quad (22)$$

SDDM is defined as

$$SDDM_{i,j} = \omega_{dd}\frac{1}{(i-j)}\left[\sum_{m=j}^{i}\sum_{n=1}^{j-1} d_m d_n \frac{(m-j)^2}{(m-n)^{5/2}} + \sum_{m=j+1}^{i}\sum_{n=j}^{m-1} d_m d_n (m-n)^{-1/2} + \sum_{m=i+1}^{N}\sum_{n=j}^{i} d_m d_n \frac{(i-n)^2}{(m-n)^{5/2}} + \sum_{m=i+1}^{N}\sum_{n=1}^{j-1} d_m d_n \frac{(i-j)^2}{(m-n)^{5/2}}\right], \quad (23)$$

where ω_{cd}, ω_{dd}, c_m, d_m have the same meaning as defined above.

As before, this function is minimized to determine optimal swelling factors x_{ij} and degrees of ionization $\alpha_{\pm}$ for a given residue pair i, j. The swelling factors are related to distances by $x_{ij} = \langle R_{ij}^2\rangle/(|i-j|b\ell)$. We fitted $\omega_{2,i,j}$ specific to residue pairs based on the experimental data with FRET labels in different positions, while p and δ were assumed to be independent of the segment probed and kept invariant for a given protein. We used experimentally measured $C_\alpha - C_\alpha$ distances for these comparisons, and dyes were modeled as one residue with an effective charge of −2 (for Alexa 488 and 594), −1 (for CF6606R), or 0 (for Cy3B), without internal structure, in contrast to the analysis for the linker IDRs (see Section S3, Supporting Information). This formalism was also used to predict normalized distance maps $d^*_{i,j}$ (see Section S6, Supporting Information) for a given linker sequence with the assumption $\omega_{2,ee} = \omega_{2,i,j}$, and $\omega_{2,ee}$ determined by fitting the single-molecule FRET data for the end-to-end distances of the 16 sequence-diversified linker IDRs.[36]

3. Results

We performed single-molecule FRET experiments on 18 different IDPs and over a wide range of KCl concentrations to provide a data set that allows us to stringently test models that describe intrachain distances and their salt-dependence. By placing FRET donor and acceptor dyes at different positions in the sequence, we can probe not only end-to-end distances but also shorter segments within the IDPs. To cover a broad range of sequences, we use a combination of samples (Figure 1). With a focus on sequence composition, we measured the end-to-end distances in a recently established set of 16 sequences selected from intrinsically disordered regions (IDRs) connecting RNA-binding domains of identical length of 57 residues, but with large differences in amino acid composition, hydrophobicity, and charge patterning[36] (Table S1, Supporting Information). While the previous study[36] reported measurements only at a single salt concentration, here we performed measurements across different salt concentrations to rigorously test different electrostatic models. To complement these measurements and test the results also for segments within longer IDPs, we use Starmaker (Stm)[37] and Prothymosin α (ProTα),[3,38] two naturally occurring, highly negatively charged IDPs of 598 and 112–113 residues, respectively (Table S1, Supporting Information). We note that naturally evolved sequences have the advantage of higher success rates in recombinant expression for protein production and avoid issues such as translation arrest of long polylysine repeats.[53] Moreover, contributions other than charge interactions are important for a realistic parametrization of the theory, so the content of uncharged amino acid residues is an essential aspect of our sequence set. To further account for the contribution of the dyes, especially their charges,[3,36] we labeled the 16 IDRs with two different FRET pairs, on the one hand AlexaFluor 488 and 594, and on the other Cy3B and CF660R. The Alexa dyes carry a net charge of –2 each, whereas Cy3B is a net neutral zwitter ion, and CF660R carries a net charge of –1 (see Section S3, Supporting Information). For all of these different sequences and variants, we performed measurements across a broad range of salt concentrations to modulate the electrostatic interactions.

Based on the FRET data, we use the following strategy: We infer average inter-dye distances for the 16 IDRs from transfer efficiencies measured as a function of monovalent salt concentration using the SAW-ν distance distribution (see Section S2, Supporting Information),[54,55] which has previously been shown to yield monomodal distance distributions for the 16 IDRs that agree well with those from coarse-grained and atomistic implicit-solvent simulations;[36] we then use the resulting average distances for direct comparison with the values from the different theoretical models described above, where the charge interactions are assumed to be sufficiently screened at the highest salt concentrations used (>0.4 M) so that those distances can be used for estimating the non-electrostatic contribution to the free energy of the chains. By minimizing the deviation between experiment and theory for all 16 sequences and all salt concentrations globally, we obtain parameter sets that provide the best description of all sequences for each model. It is worth emphasizing that the sequences even exhibit qualitatively different trends with increasing salt concentration: some sequences compact, while others expand, reflecting the importance of the content and distribution of positive and negative charges within the sequence. This rich set of data collected for these diverse linker IDRs and two additional IDPs Stm and ProTα provides the first rigorous test of a unified analytical theory of electrostatics that self-consistently accounts for the effects of sequence, variable charge states, dipolar interactions, ionic strength and their impact on the chain conformation.

3.1. Full Ionization Is Insufficient for Describing Chain Dimensions Quantitatively

We begin with the simplest model, M0, to test its ability to describe the data. Model M0 only minimizes x and assumes $\alpha_{\pm} = 1$, implying that all charged residues are fully ionized. While M0 correctly captures the qualitative trends of salt-induced compaction/expansion of the majority of the sequences, it tends to overestimate the chain dimension in the low-salt regime (**Figure 2**). For the sequence sCh, M0 even fails to describe the qualitative trend, predicting a decrease in chain dimensions with increasing salt concentration, in contradiction to the data. For comparison, we also include the mean-field polyampholyte model of Higgs and Joanny,[3,49] which assumes the positive and negative charges to be distributed uniformly along the sequence

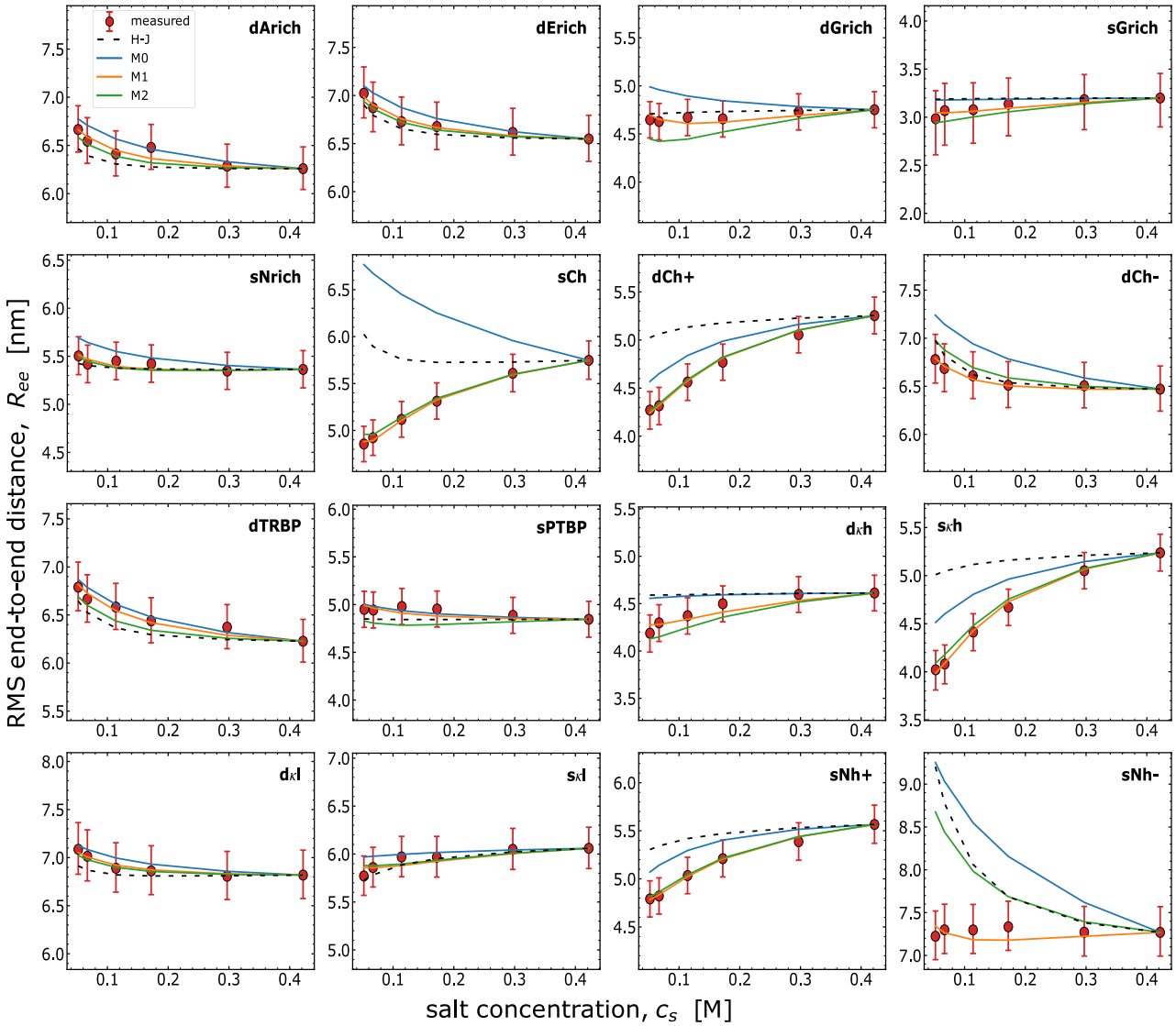

Figure 2. Comparison of the different theoretical models with single-molecule FRET results for 16 sequence-diversified IDRs labeled with Alexa 488/594 ("measured"). Full ionization models, purely compositional (Higgs-Joanny; H-J, black dashed line), or with sequence dependence (M0, blue line), tend to overestimate conformations compared to FRET-based RMS end-to-end distances across ionic strengths. Counterion models, M1 (orange line) with sequence-dependent δ, and the simpler near-predictive M2 (green line) with sequence-dependent δ given by the linear relation $\delta = -0.123\sqrt{\langle R_{ee}^2\rangle} + 2.146$ from low-salt $\sqrt{\langle R_{ee}^2\rangle}$ (in nm) (see Figure 3), both with $\tilde{p} = 0.66$, describe the data better (see Table 1). For all models, the two-body non-electrostatic interaction parameter $\omega_{2,\,ee}$ is determined by matching the data points at the highest salt concentration. Models M0, M1, M2 used a three-body non-electrostatic interaction strength of $\omega_3 = 0.1$. Higgs–Joanny (H–J) additionally required a strongly amplified three-body term, $\omega_3 = 10$, to prevent unphysical collapse of polyampholytic sequences at low salt. See Table S2 (Supporting Information) for parameter values.

and does therefore not take charge patterning into account. The resulting trends and quantitative agreement are generally similar to those of model M0. However, for a substantial fraction of the sequences, neither of these simple models provide a quantitative account of the experimental findings.

3.2. Model With Ion Condensation Describes Data Well Across IDR Sequences

To explain the ionic-strength-dependent data of all 16 linker IDR sequences, we thus test the model with ion condensation. An important aspect of model selection is to avoid overfitting. For this purpose, we apply a model with the minimal number of parameters. Although all three adjustable parameters are expected to be sequence-dependent, we expect the average dipole length p to be least sensitive to the sequence because it is not explicitly dependent on the local chain environment, while δ is dependent on local permittivity. Thus, we fit the data with model M1 where p is held fixed but δ is allowed to vary from sequence to sequence (see Section S8, Supporting Information, for details on our χ^2 fitting procedure).) We also determine the non-electrostatic parameter $\omega_{2,ee}$ for each sequence by matching the high-salt data, where

electrostatics is expected to be largely screened.[3] The best fit value of the dimensionless dipole length $\tilde{p} = 0.66$ and the values of δ (reported in Table S2, Supporting Information) for each sequence are within the expected range (see Section S7, Supporting Information). Model M1 fits the salt-dependent chain dimensions very well for virtually all linker sequences (Figure 2). We also note that the effective charge reduction due to counterion condensation can be pronounced (Figure S3, Supporting Information). Furthermore, the effective charges can differ substantially from sequence to sequence, and although they always decrease with increasing salt concentration, the amplitude of the change differs between sequences and between anionic and cationic groups.

The above analysis with M1 yields fitted δ values with a substantial spread from roughly 1.0–1.9 (Table S2, Supporting Information), precluding the use of a global value. Understanding the sequence-dependent variation of δ is important to reduce model complexity and build a predictive model. We hypothesize that sequences with more compact dimensions would tend to have a greater mismatch between bulk and local dielectric (larger δ) due to lower solvent accessibility. To test this hypothesis, we plot δ against $\sqrt{\langle R_{ee}^2 \rangle}$ at low salt concentration (**Figure 3**) and notice a modest anticorrelation, with sNh- and sPTBP being two outliers. The additional scatter indicates contributions of other local effects (e.g., polar residues, spatial variation of dielectric mismatch) on δ beyond just the overall chain compactness. Nevertheless, to reduce the number of adjustable parameters, we leverage the linear correlation to find the simplest relation that describes the salt concentration-dependent data for all sequences with a global fit. In this global fitting procedure, $\omega_{2,ee}$ is determined for each sequence by matching the high-salt $\langle R_{ee}^2 \rangle$ measurement. The linear relation between δ and $\sqrt{\langle R_{ee}^2 \rangle}$ at low salt is adjusted until the total χ^2 for 15 IDRs (all except sNh-) is minimized. Note that this is different from directly fitting sequence-specific δ values (blue points in Figure 3) obtained from M1 by fitting the salt concentration-dependent data for each sequence separately. The sequence-dependent value of δ obtained from the global fit can be estimated from the low-salt (52 mM) mean squared end-to-end distance $\sqrt{\langle R_{ee}^2 \rangle}$ (in nanometers) from the equation

$$\delta = -0.123 \sqrt{\langle R_{ee}^2 \rangle} + 2.146 \tag{24}$$

where the minimum allowed value of δ is unity since the dielectric in the volume occupied by the chain cannot be higher than that of water. We find that the resulting model M2, in which δ values are used based on this optimized line (dashed black line in Figure 3), describes the data well for most sequences (Figure 2), including sPTBP, which was an outlier in Figure 3, implying that sPTBP is not very sensitive to the dielectric mismatch parameter.

The only sequence that M2 fails to describe is sNh-, which is also an outlier in the dependence of δ on $\sqrt{\langle R_{ee}^2 \rangle}$ and was thus not included in the linear fit. sNh- is high in negative charge, with the highest net charge per residue among the IDRs of −0.35, and low in positive charge. Polyelectrolyte theory would thus predict a decrease in chain dimensions with the addition of salt, although there are contiguous stretches of 12 glutamates at the C-terminus and 3 lysines at the N-terminus, whose attractive interaction could lead to polyampholyte-like expansion with increasing salt concentration.[21,49] However, the single-molecule FRET measurements show that the dimensions of sNh- are largely insensitive to salt concentration. Within our model, this trend could be explained by a low degree of ionization or equivalently a high propensity to form dipoles by counterion condensation. A high value of dielectric mismatch promotes such dipole formation and a resulting low degree of ionization. The specific cause of such high dielectric mismatch is, however, unclear. On the other hand, it is possible that sNh- exhibits salt insensitivity due to other factors not included in the model. For example, it is conceivable that the dye interacts with the three lysine residues close to the N-terminus and forms local contacts not accounted for in the model.[36] Indeed, an influence of dye interactions is supported by the difference between the salt concentration dependence observed for the Alexa 488/594- and Cy3B/CF660R-labeled variants (**Figure 4**). There might thus be a competition between interactions of the lysine-rich region with the glutamate-rich region on the one hand and the N-terminal dye on the other; the latter is expected to be less pronounced for Cy3B and CF660R owing to their lower net charge compared to the Alexa dyes.[36] The salt insensitivity could also arise from specific non-electrostatic interactions that are only described at a mean-field level in our model.

Nevertheless, model M2 accounting for partial charge neutralization describes the data fairly well for all proteins except sNh-. **Table 1** provides a comparison between the four models (H–J, M0, M1, and M2). The models M1 and M2 taking into account counterion condensation perform better than the models without condensation (HJ and M0). The near-predictive model M2

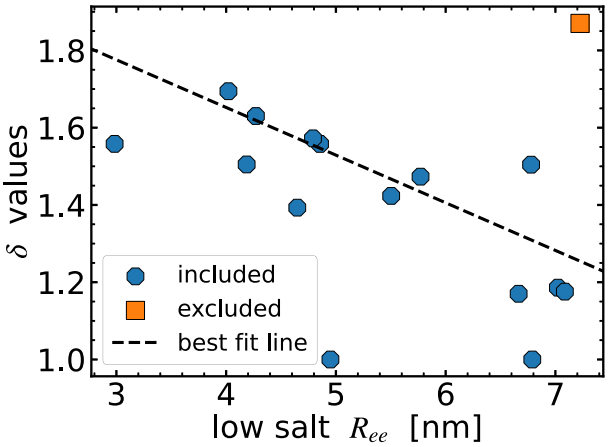

Figure 3. The dielectric mismatch parameter δ of all linker IDRs from model M1 (points) labeled with Alexa 488/594 is linearly correlated with the FRET-based distances $\sqrt{\langle R_{ee}^2 \rangle}$ (at 52 mM salt concentration), with a Pearson correlation coefficient of $r = -0.64$. Assuming a linear dependence of δ on $\sqrt{\langle R_{ee}^2 \rangle}$ at low salt (52 mM), we find the optimized dependence (black dashed line) by fitting $\langle R_{ee}^2 \rangle$ as a function of salt concentration (from FRET) for all 15 sequences simultaneously (all except the outlier sNh-, orange). For model M2, δ is set for each sequence according to this optimal line (black dashed line) given by $\delta = -0.123 \sqrt{\langle R_{ee}^2 \rangle} + 2.146$, thus removing all parameters adjustable individually for each chain except the non-electrostatic interaction strength $\omega_{2,ee}$.

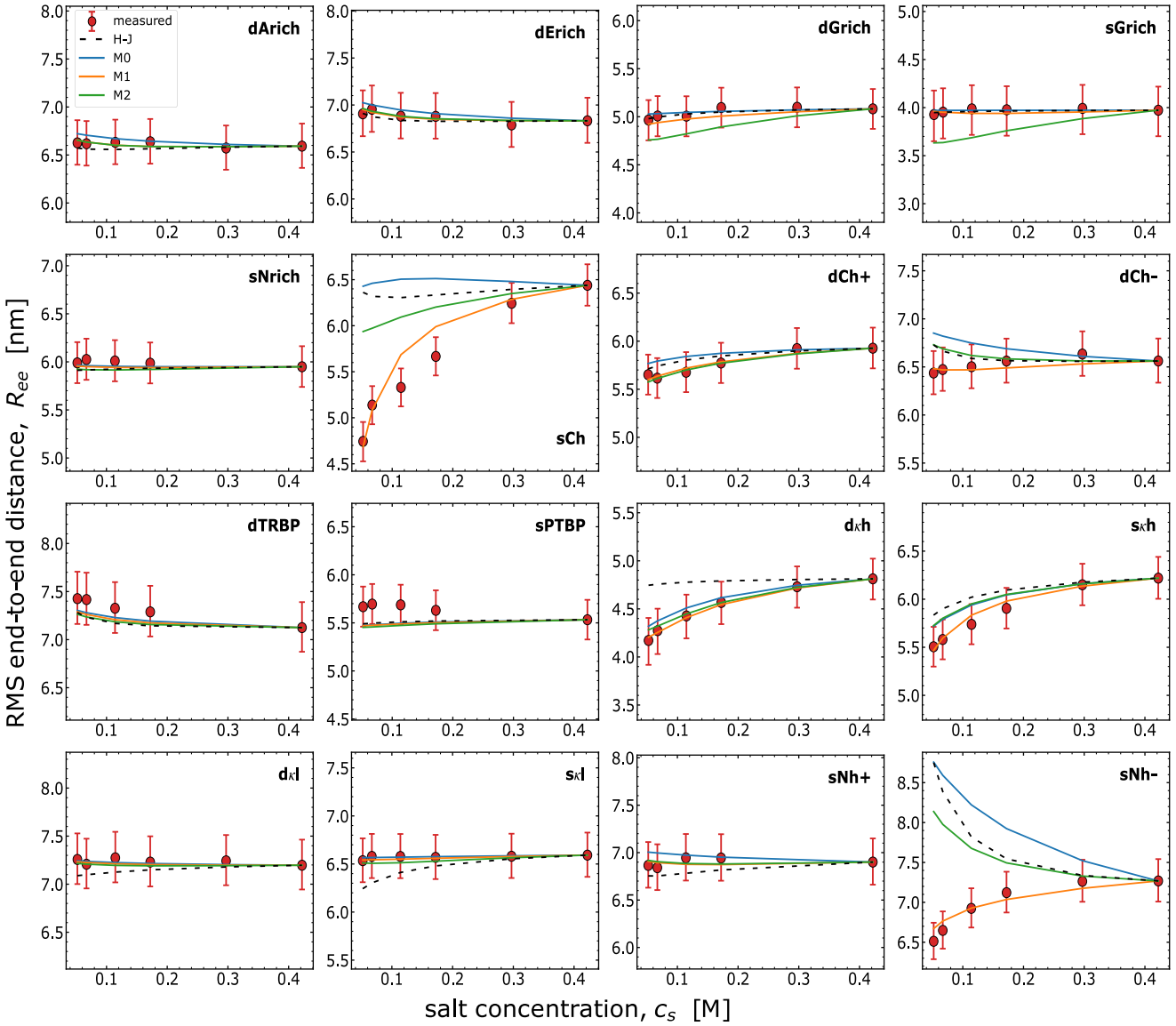

Figure 4. Comparison of the different theoretical models with single-molecule FRET results for the 16 sequence-diversified IDRs labeled with Cy3B/CF660R ("measured") (analogous to Figure 2): compositional Higgs–Joanny (H–J, black dashes), sequence-dependent model with full ionization (M0, blue line), counterion model with fitting δ individually (M1, orange line) and applying the near-predictive linear relation for δ with $\sqrt{\langle R_{ee}^2 \rangle}$ (M2, green line), all with $\tilde{p} = 0.66$. The near-predictive model M2 presented here is based on the same linear relation for predicting δ based upon the Alexa dye results shown in Figure 3), and still compares reasonably well with the FRET-based RMS end-to-end distances across ionic strengths. Parameter values are shown in Table S3 (Supporting Information).

does only slightly worse than model M1 with δ adjusted individually for each sequence. Thus, model M2 reduces the number of adjustable parameters without significantly compromising its ability to describe the data, and thus increases the predictive power of our formalism. To further facilitate the comparison of the models, we present the deviations of all four models from the experimental data in Figure S4 (Supporting Information). Positive differences ($\Delta R_{ee} \equiv \Delta\sqrt{\langle R_{ee}^2 \rangle}$) are where theory predicts more expanded conformations than measured, while negative differences are where theory predicts more compact conformations. We also provide comparison of χ^2 errors across models for each sequence (Figure S5, Supporting Information).

The inferred parameter values can be further used to quantitatively determine the different contributions to the total free energy (entropy of counterions, free energy of ion condensation, chain free energy, etc). For this purpose, we selected two representative sequences, sκh and dCh- due to their differences in charge composition and salt response. sκh is a strong polyampholyte with 11 positive and 11 negative charges, and expands with addition of salt. On the other hand, dCh- is polyelectrolyte-like, with 13 negative charges and 4 positive charges and compacts with addition of salt. Figure S6 (Supporting Information)

Table 1. Summary of model performance for the 16 IDRs with Alexa dyes for Higgs–Joanny (H–J), full ionization (model M0), counterion model with fitted δ (model M1), counterion model with δ from best linear dependence on low-salt chain dimensions $\sqrt{\langle R_{ee}^2 \rangle}$ (model M2). Mean errors (chi-squared) of $\sqrt{\langle R_{ee}^2 \rangle}$ across salt concentrations are determined for each sequence, then we report two means: i) the mean across all sequences (overall mean) and ii) the mean across all except sNh- (excluding sNh-). A comparison of χ^2 across models for each sequence is shown in Figure S5 (Supporting Information).

	H–J	M0	M1	M2
Overall mean	3.297	4.702	0.0397	0.700
excluding sNh-	2.586	3.725	0.0348	0.166

Table 2. Summary of model performance for the 16 IDRs with Cy3B/CF660R dyes using Higgs–Joanny (H–J), full ionization model (M0), counterion model with individually fitted δ (model M1), counterion model with δ from best linear dependence on low-salt $\sqrt{\langle R_{ee}^2 \rangle}$ (model M2). Mean errors (chi-square) of $\sqrt{\langle R_{ee}^2 \rangle}$ across salt concentrations are determined for each sequence, then we report two means: i) the mean across all sequences (overall mean) and ii) the mean across all except sNh- (excluding sNh-). A comparison of χ^2 across models for each sequence is shown in Figure S10 (Supporting Information).

	H–J	M0	M1	M2
Overall mean	3.413	3.941	0.1475	1.949
excluding sNh-	1.776	1.929	0.1466	1.026

shows the contributions of the different free energy terms for these two sequences using the parameter values obtained from model M1.

As detailed above, model M2 reduces the number of parameters by estimating δ from the measured end-to-end distance at low salt and ignores its variation with salt. This is a reasonable approximation given the greater variation in the end-to-end distance at low salt across different sequences than as a function of salt concentration for a given sequence. However, the formalism can be extended to model M3, where δ is allowed to vary with salt for a given sequence. The salt dependence of δ can be approximated using equation 24, where $\sqrt{\langle R_{ee}^2 \rangle}$ is the end-to-end distance measured at different salt concentrations. Model M3 compares well with the data (Figure S7, Supporting Information). We use this analysis primarily as a consistency check, since it is difficult to use M3 for predictions if salt-dependent measurements are not available. Variation in δ can also arise from the decrease in dielectric constant with increasing salt concentration, known as dielectric decrement.[56] However, for KCl and the highest salt concentration of 420 mM used here, the change in dielectric will be relatively small (from 80 to 75) with minimal impact on δ values. Thus, we conclude that, overall, M2 is the preferred model for its simplicity, practical use, and its ability to describe a wide range of data.

3.3. Testing the Model With a Different Dye Pair

To address the question of how much the charged dyes impact chain dimensions and whether our models are general enough to describe the effect, we collected ionic strength-dependent single-molecule FRET data using a different pair of dyes, Cy3B and CF660R, which carry a lower net charge than the AlexaFlours and have been found to exhibit less attractive interactions with positively charged residues.[36] We then used models M0, M1, and M2 to fit the data, with the same value of p and the same linear relation between δ and $\sqrt{\langle R_{ee}^2 \rangle}$ used for the Alexa dyes (Figure 3 and Equation 24) to estimate the dielectric mismatch for M2. The only independently adjustable parameter is the non-electrostatic interaction $\omega_{2,ee}$ determined from the high-salt data for each sequence. The resulting model parameters are listed in Table S3 (Supporting Information). As before, we find that both M1 and M2 describe the data well, with the exception of sNh- and sCh, for which M2 does not provide a good fit (Figure 4). Both sNh- and sCh show a strong salt-induced expansion in the experimental data, indicative of a pronounced polyampholyte effect. Indeed, sNh- and sCh are the two sequences with the combination of the highest total charge content and highest degree of charge segregation in the set.[36] The accumulation of opposite net charges near the termini may thus have a pronounced influence on chain dimensions, potentially modulated by charge-mediated dye interactions. We note that our model ignores the branched nature of the dye moieties attached via flexible linkers and any salt dependence of chain dimensions arising from mechanisms other than electrostatic effects. For example, the solubility of polar amino acids (and fluorophores) can vary with salt concentration, potentially modulating the salt-dependent variation of the chain dimensions.[57] We also ignored the spatial variation in δ for a given sequence even at a fixed salt concentration. However, differences in local compactness and the nature of chemical groups may cause fluctuations in the δ values. It is important to note that the data for sCh can be well described by M1, which optimizes the dielectric mismatch parameter within a range for each sequence individually, yielding a value of $\delta = 2.03$, whereas for M2 the value is $\delta = 1.56$.

As before, we compare all models (H–J, M0, M1, M2) for the 16 IDRs with Cy3B/CF660R in **Table 2** and Figure 4. Overall, we find that M1 is the best model in terms of absolute agreement with the data, but at the price of individually adjusted values of δ. The near-predictive model M2 performs better than both the full ionization (M0) and H-J model, and given the fixed relation between δ and $\sqrt{\langle R_{ee}^2 \rangle}$, it provides a good compromise between fit quality and the number of adjustable parameters.

We note that counterion condensation can lead to a substantial reduction in effective charge and varies across sequence and ionic strength (Figure S8, Supporting Information), consistent with our earlier observation with the Alexa dyes. As before, we also provide a comparison of the deviations of all four models from the measured data to highlight differences between models (Figure S9, Supporting Information). Moreover, Figure S10 (Supporting Information) compares the χ^2 values of the different models for each sequence. Finally, we provide a comparison of data against model M3, where δ is allowed to vary with ionic strength, for a given sequence (Figure S11, Supporting Information). M3 performs even slightly better than M2, confirming that our analysis is internally consistent. In fact, M3 improves the pre-

diction for sCh compared to M2 (as discussed earlier), indicating that the salt dependence of δ can be important for this sequence. We reiterate that model M2 (equation 24 derived from data collected using the Alexa dyes) can be used to describe the data collected with Cy3B/CF660R dye pairs. This result, along with the observation that δ values are in a reasonable range (see Section S7.3, Supporting Information, for estimates) and that M3 provides improved or comparable performance as M2, indicates that δ is likely to have physical significance, even though we are currently not able to determine it from first principles.

3.4. Inferring Inter-Residue Distance Maps

So far, we have used single-molecule FRET data to obtain mean square end-to-end distances $\langle R_{ee}^2 \rangle$ across salt concentrations (and sequences) to determine the adjustable parameters of the models. However, our resulting models can also be utilized to predict the average distances $\langle R_{ij}^2 \rangle$ between any two residues (i,j) in the chain, which are not available from the experimental data and would be difficult to obtain because only sequence separations greater than 30 or 40 residues can typically be probed with the Förster radii of the dye pairs available for single-molecule FRET.[33,34] We present the distance maps predicted by the models in the normalized form d_{ij}^* (see Section S6, Supporting Information, for details of calculation) using the parameters from M0 and compare them with the maps resulting from M1 (**Figure 5**), where parameters were obtained by fitting data collected using the Alexa dyes.

The resulting distance maps indicate clear deviations from the uniform scaling of interresidue distance with sequence separation expected for simple homopolymers,[20] and the regions of relative expansion (or compaction) are similar for the models M0 and M1. Moreover, features of distance maps (regions of expansion relative to homopolymer scaling in red and compaction in blue; see Section S6, Supporting Information, for details) for sequences dArich, sCh, dCh-, sPTBP, dκh, sκh, sNh+, sNh- compare well with those computed from the coarse-grained and atomistic simulations reported previously.[36] These simulated maps were predicted after reweighting the ensemble (for atomistic simulation using ABSINTH[58]) or by fitting non-electrostatic interaction parameters (for coarse-grained simulation with the HPS model[36]) to match FRET efficiency data. Modest agreement is found for dGrich, dκl, sκl. However, dErich, sNrich, dCh+, dTRBP, and sGrich show clear deviations between the distance maps based on the simulations and the predictions based on our models. Interestingly, the same sequences (except dCh+) also exhibit differences in distance maps between coarse-grained and atomistic simulations, indicating features in sequences that amplify differences in force fields and models. We note that theoretical predictions using M1 include dipolar interactions and partial ionization, which are not included in either simulation. On the other hand, non-electrostatic interactions in M0 and M1 are modeled with the mean-field assumption $\omega_{2,i,j} \approx \omega_{2,ee}$. Both factors can contribute to the observed discrepancies between theory and simulation. Additional experiments on the linker sequences would be needed to benchmark these predicted distance maps and thus assess the accuracy of theory, coarse-grained simulations (HPS model), and atomistic simulations.[36]

Figure S12 (Supporting Information) shows the predicted distance maps using parameters derived from data collected with the Cy3B/CF660R dye pair. We also compared distance maps predicted from parameters obtained by fitting data collected using the two different dye pairs (Alexa dyes against Cy3B/CF660R dyes). These distance maps (Figure S13, Supporting Information) for a given sequence are very similar, with two notable exceptions for sCh and sNh-. These differences most likely arise from the presence of the dyes, and are consistent with the effect of dyes seen in the measured chain dimensions for these two sequences (Figures 2 and 4).

3.5. Model With Ion Condensation Is Required to Describe Inter-Residue Distance Profiles

The preceding section highlights the ability of providing intrachain distances from our models. Motivated by this aspect, we tested our models' ability to describe inter-residue distances for segments within the chain for two longer IDPs, ProTα (112-113 residues) and Stm (598 residues). For Stm, we measured eight sets of distance pairs and for ProTα three distance pairs as a function of salt concentration. For this comparison, we use Cα-Cα inter-residue distances and model the dyes as one bead with an effective charge, neglecting the branched nature of the dyes not amenable to more detailed treatment in our current theoretical formalism (see Section S3, Supporting Information, for details of dye modeling). Since we have multiple sets of data for a given sequence, unlike for the linker sequences, we modify our parameter inference scheme from before. For a given sequence, we optimize for values of p and δ (similar to M1 for the 16 IDRs) that best describe the measured distances for all pairs. Moreover, residue pair (i,j)-specific distances allow us to determine interaction strength values ($\omega_{2,i,j}$, defined in equation 20) individually (and do not assume they are identical for all segments) by fitting the high-salt data for every segment probed. These parameters are expected to vary across residue pairs (i,j) due to variation in the non-electrostatic two-body interactions ($\omega_{m,n}$ in Equations 19 and 20) between different types of amino acids.

Both Stm and ProTα show a pronounced compaction with increasing salt concentration, as expected based on their large negative net charge and the low fraction of positively charged residues[3,37] (**Figure 6**). We find that the ion-condensation model M1 describes the salt-dependent distances for Stm and ProTα reasonably well (Figure 6). The fitted values of $\omega_{2,i,j}$, exhibit modest variations from segment to segment, with Stm segments spanning values from 0.3 to 1.7, and ProTα segments values from 0.9 to 1.7 (Tables S4 and S5, Supporting Information). Remaining discrepancies between theory and experiment could be due to effects such as a variation of non-electrostatic or dielectric mismatch parameters with ionic strength, which is neglected in the model. We also ignore local fluctuations in the dielectric mismatch. The full ionization model M0, in contrast, overestimates inter-residue distances for both proteins, similar to the trend observed for the 16 linker IDRs, highlighting the important role of counterion condensation and the resulting reduction in effective charge. The inferred values of p ($\tilde{p} = 0.45$) are within the range expected from our previous results (Section S7, Supporting Information). Furthermore, for both IDPs we obtain low

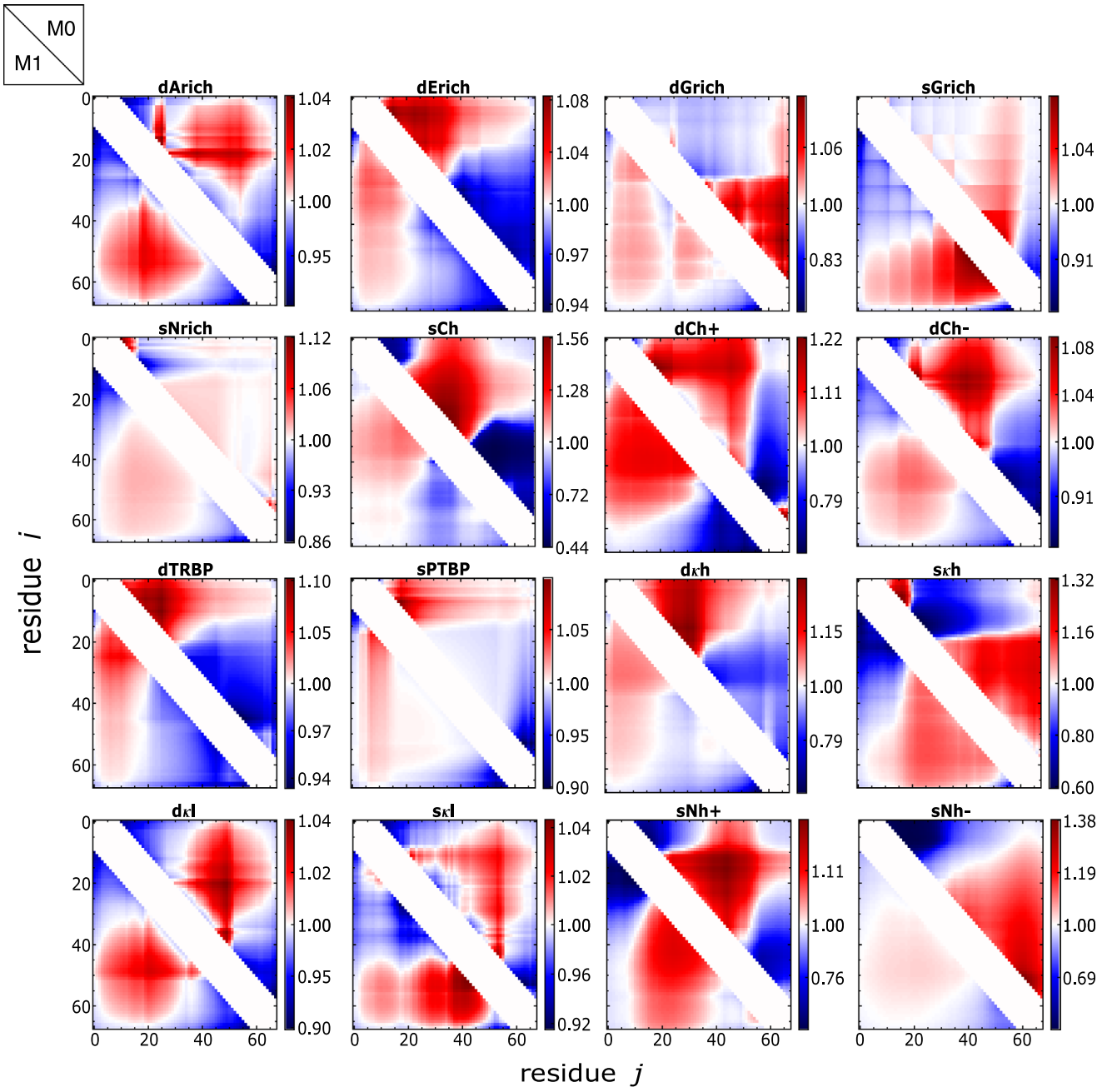

Figure 5. Inter-residue distance maps (normalized by fits to homopolymer scaling; see Section S6, Supporting Information) for the 16 IDRs at 150 mM salt ($T = 20\,°C$), predicted using parameters (determined by matching end-to-end distances collected using Alexa 488/594) of model M1 (bottom triangle) and M0 (upper triangle). The maps clearly show a dependence on sequence, and differences between models are qualitatively consistent with those of Ref. [36].

values of δ (1.3 for Stm and 1.1 for ProTα) and an expanded chain at low salt, consistent with the trend seen in Figure 3 for the linker IDRs. We also note that predicting the distance between specific residue pairs is associated with minimizing the free energy with respect to three variables: distance and two effective charges (for positive and negative groups). Consequently, predicted distances between specific residue pairs will have associated with them a prediction of the effective degrees of ionization. A self-consistent model must predict very similar values of these effective charges (degrees of ionization) that are global properties of the chain. As shown in Figure S14 (Supporting Information), these values are nearly identical, reflecting internal consistency of our model. As before, we also note that these effective charges gradually decrease with increasing salt concentration. As a further consistency check, we reanalyzed the 16 IDRs using Cα-Cα distances with the single-bead model for the dye. To ensure the robustness of our parameters, we kept the same values of p and δ from M2 and only fitted $\omega_{2,ee}$. Figures S15 and

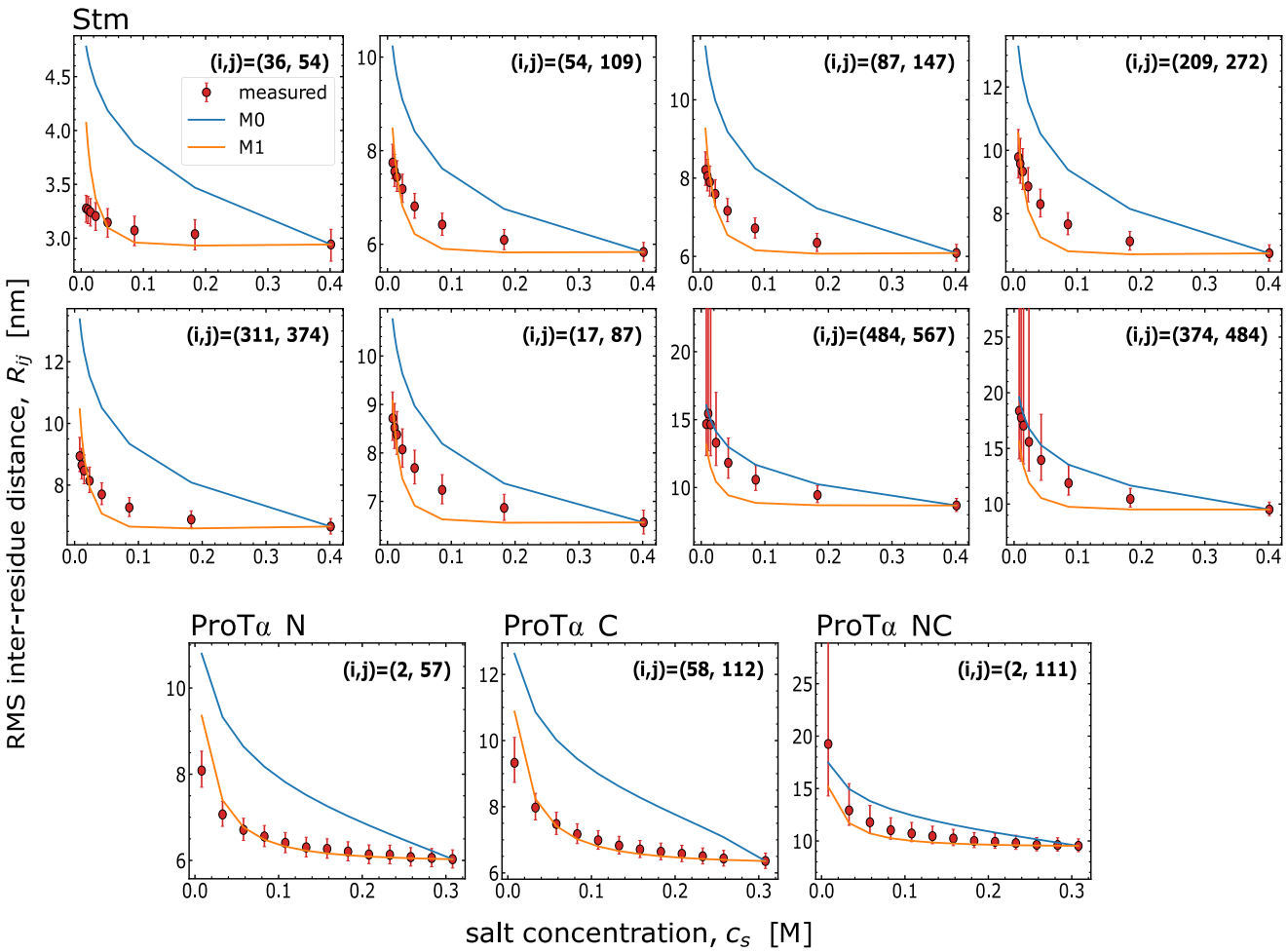

Figure 6. Inter-residue distances of Stm and ProTα as a function of salt concentration, as measured by single-molecule FRET (red), are described better by the sequence-dependent condensation model (M1, orange), compared to the full ionization model (M0, blue). The best parameter values for Stm used here are $\tilde{p} = 0.45$, $\delta = 1.30$, and for ProTα $\tilde{p} = 0.45$, $\delta = 1.10$. Each plot shows a different inter-residue pair as labeled by (i, j). Mean errors (chi-square) between theory and experiment averaged across residue pairs, for Stm, are 27.680 for M0 and 2.803 for M1; for ProTα, they are 18.787 for M0 and 0.590 for M1. See Tables S4 and S5 (Supporting Information) for additional parameters, and Table S1 (Supporting Information) for the sequences. See Figure S14 (Supporting Information) for effective charges (degrees of ionization).The large error bars for Stm (484, 567), (374,484), and ProTα NC are a consequence of the low transfer efficiencies observed for these variants (see Methods, Supporting Information).

S16 (Supporting Information) show good agreement between the model and the data, supporting the robustness of the inferred parameters irrespective of the two slightly different models. For the Alexa 488/594 fluorophores, mean errors (chi-squared) are 0.927 including all linker IDRs, and 0.147 when excluding sNh-. Similarly, for Cy3B/CF660R, mean errors (chi-square) are 2.244 including all sequences, and 0.956 excluding sNh-.

4. Discussion

A large set of single-molecule FRET experiments were performed to determine intrachain distances and global dimensions of IDPs with very different sequences and lengths; by varying the location of the FRET probes within the sequences; by using different FRET dye pairs; and by changing the salt concentration across a broad range to modulate the electrostatic interactions. These comprehensive sets of measurements were used to test different variants of analytically tractable polymer models that take into account the sequence-specific charge distributions and to quantify several factors that affect electrostatic interactions in IDPs and influence their conformational distributions. Specifically, we used the data to test two competing models: (i) a model (M0) where charged amino acids are fully ionized, and (ii) a model (M1) allowing partial ionization due to condensation of oppositely charged counterions on the side chains. Predicting the extent of ion condensation is non-trivial since it depends on the conformation which in turn depends on the condensation, requiring a self-consistent treatment. M1 is such a self-consistent analytical theory accounting for sequence effects (not just composition) to predict both chain dimensions and ionization. Although M0 also computes sequence-dependent electrostatics within an analytical theory, we find that it is insufficient to explain the data. Specifically, M0 tends to overestimate the chain dimensions at low salt concentrations. The ion condensation model M1, on the other

hand, can explain the data at different ionic strengths and outperforms M0. Different versions of ion condensation models were tested against the data to balance model complexity and agreement with experiment. It is possible to find a near-predictive model (M2) where the dipole length resulting from the distance between counterion and side chain charge is treated as a shared parameter across all sequences, and the dielectric mismatch parameter is described by an empirical relation based on its correlation with the dimensions of the chains.

The better performance of the models taking into account ion condensation can be traced to two factors. First, the reduction in effective charge significantly alters the charge-charge interactions (repulsive or attractive) compared to those in M0. Furthermore, new types of interactions, absent in M0, arise from the condensation of oppositely charged ions forming dipoles on the chain, which can contribute to intrachain interactions. Together, these two effects alter the sequence-dependent interactions and the chains' response to salt concentration and consequently their conformational distributions. All these effects, and their sequence and solution dependence, are self-consistently treated in model M1. This self-consistent model can also predict – although not seen in the current set of sequences – first order phase transitions in charge and chain dimensions space by varying the Bjerrum length (inversely related to temperature) and salt for some sequences, as seen in our previous work.[28]

At present, we ignore the patterning effects from uncharged amino acids and their response to salt concentration. Our analytically tractable model also ignores the branched nature of the dyes attached to the chain via flexible linkers.[36] These and other subtle conformational restrictions due to rotation and limited bond stretching, neglected in the theory, can be captured in simulations. Another assumption of the model is that charges of one type have the same degree of ionization, thus ignoring local differences in counterion condensation. Furthermore, we neglect spatial fluctuations of the dielectric constant and assume a mean-field value for a given sequence. Moreover, we neglect variation of the dielectric constant and hydrophobicity (affecting ω_2) with salt concentration. These effects may further modify the degree of condensation and may be responsible for the disagreement of the theory with one of the sequences we investigated experimentally.

The ability to describe a wide range of measurements on different disordered proteins with polymer models that take into account counterion condensation supports the need to further investigate the role of counterions for the interactions within and between IDPs, often ignored in computational studies of IDPs at present, especially in coarse-grained models, and its impact on IDP conformation. Further directions to explore would be the effect of divalent ions, particularly for poly-acidic IDPs, which would be particularly pertinent for understanding the role and regulation of IDPs in the cellular milieu. Accounting for this effect may also provide insights to the role of electrostatics in IDP function.[59] The model presented here will be of practical use in such predictions due to its analytical and high-throughput nature and the ability to generate new insights and hypotheses. For large IDPs, such as Stm, simulations remain challenging, especially at an atomistic level with explicit solvent, where explicit ion effects can be included.[60] An interesting aspect will be to extend the current theory to include an explicit treatment of protonation equilibria,[51,61] especially for solutions at low salt concentrations and in pH ranges close to the pK_a values of ionizable amino acid residues. Another interesting extension of the theory would be to include the contribution of salting-out effects or increased hydrophobicity at very high salt concentrations in the molar range.[23] Analytical models like the ones we present here will continue to make an important contribution to the multi-scale modeling of disordered proteins. The predicted degree of ionization might be used to modify charge states in coarse-grained simulations. Finally, our work addresses the long-standing question of counterion condensation in polymer physics mainly discussed for homopolymers. Our formalism extends the investigation to heteropolymers, highlighting the rich physics of the problem and stimulating future studies in natural and synthetic polymers.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors thank Lucia Franchini, Paweł Łukijańczuk, Daniela Oswald, Gessica Priola, Moa Hasler, Vincent Maximilian Münch, Steffen Winkler, and Ruijing Zhu for help with sample preparation and measurements; Edward Lemke for providing the pBAD-Int-CBD-12His plasmid; M. Muthukumar for many stimulating discussions over the years; and the Functional Genomics Center Zurich (FGCZ) of the University of Zurich and ETH Zurich, especially Dr. Serge Chesnov, for excellent mass spectrometry service. The authors acknowledge support from NIH (R01GM138901 to K.G.), NSF MPS Ascend Fellowship (DMR 2213103 to M.P.), the Swiss National Science Foundation (Grant number 310030_197776 to B.S.), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement ID 898228 to A.C., and Sciex (Sciex-NMS Fellowship 12.230 to M.W.). The authors also acknowledge support from high-performance computing (RDAC) at the University of Denver.

Open access publishing facilitated by University of Zurich, as part of the Wiley - University of Zurich agreement via the Consortium Of Swiss Academic Libraries.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

M.P., A.H., M.W., and A.C. contributed equally to this work. M.P., A.H., M.W., A.C., S.K., A.Sor., A.O., B.S., and K.G. conceptualized the project, with M.P., A.H., A. Sor., B.S., K.G. as lead; M.P, A.H., A.C., N.L., J.H., D.N., B.S., K.G. participated in formal analysis, with M.P., A.H., A.C. as lead; M.P., A.H., M.W., A.C., A.Sor., A. Sot., S.K., N.M., N.L., J.H., M.L., B.S., K.G. participated in investigation; M.P., A.H., M.W., A.C., J.H., D.N., B.S., K.G. participated in methodology development; M.P., N.L., D.N., participated in software development, with M.P. as lead; M.P., A.H., A.C., B.S., K.G. contributed to visualization, with M.P. and A.C. as lead; M.P., M.W., A.C., A.O., B.S., K.G. were responsible for funding acquisition, with B.S. and K.G. as lead; M.P., A.O., B.S., K.G. did project administration, with B.S. and K.G. as lead; M.P., A.H., A.Sor., A.O., B.S., K.G. contributed to supervision, with B.S. and K.G. as lead, M.P., B.S., K.G wrote the original draft, with K.G. as lead; M.P., A.H., M.W., A.C., S.K., N.M., A.O., B.S. and K.G. participated in reviewing and editing the final draft.

Data Availability Statement

The data that support the findings of this study are available from the corresponding authors upon reasonable request.

Keywords

intrinsically disordered proteins, single-molecule FRET, polymer theory, polyelectrolytes

Received: July 24, 2025
Revised: November 4, 2025
Published online:

[1] V. N. Uversky, J. R. Gillespie, A. L. Fink, *Proteins* **2000**, *41*, 415.
[2] J. Marsh, J. Forman-Kay, *Biophys. J.* **2010**, *98*, 2383.
[3] S. Müller-Späth, A. Soranno, H. V, H. Hofmann, S. Ruegger, L. Reymond, N. D, B. Schuler, *Proc. Natl. Acad. Sci.* **2010**, *107*, 14609.
[4] A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, R. V. Pappu, *Proc. Natl. Acad. Sci.* **2010**, *107*, 8183.
[5] H. Hofmann, A. Soranno, A. Borgia, K. Gast, D. Nettels, B. Schuler, *Proc. Natl. Acad. Sci.* **2012**, *109*, 16155.
[6] S. M. Sizemore, S. M. Cope, A. Roy, G. Ghirlanda, S. M. Vaiana, *Biophys. J.* **2015**, *109*, 1038.
[7] R. K. Das, R. V. Pappu, *Proc. Natl. Acad. Sci.* **2013**, *110*, 13392.
[8] L. Sawle, K. Ghosh, *J. Chem. Phys.* **2015**, *143*, 085101.
[9] R. K. Das, K. M. Ruff, R. V. Pappu, *Curr. Opin. Struct. Biol.* **2015**, *32*, 102.
[10] H. S. Samanta, D. Chakraborty, D. Thirumalai, *J. Chem. Phys.* **2018**, *149*, 163323.
[11] J. McCarty, K. T. Delaney, S. P. O. Danielson, G. H. Fredrickson, J. E. Shea, *J. Phys. Chem. Lett.* **2019**, *10*, 1644.
[12] U. Baul, D. Chakraborty, M. L. Mugnai, J. E. Straub, D. Thirumalai, *J. Phys. Chem. B.* **2019**, *123*, 3462.
[13] W. Zheng, G. Dignon, M. Brown, Y. C. Kim, J. Mittal, *J. Phys. Chem. Lett.* **2020**, *11*, 3408.
[14] A. Rumyantsev, A. Johner, J. Pablo, *ACS Macro Lett.* **2021**, *10*, 1048.
[15] J. Dinic, M. Schnorenberg, M. Tirrell, *Biomacromolecules* **2022**, *23*, 3798.
[16] K. Ghosh, J. Huihui, M. Phillips, A. Haider, *Annu. Rev. Biophys.* **2022**, *51*, 355.
[17] H. X. Zhou, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 3569.
[18] S. Das, Y. H. Lin, R. M. Vernon, J. D. Forman-Kay, H. S. Chan, *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 28795.
[19] J. Huihui, T. Firman, K. Ghosh, *J. Chem. Phys.* **2018**, *149*, 085101.
[20] M. Muthukumar, *Physics of Charged Macromolecules: Synthetic and Biological Systems*, Cambridge University Press, Cambridge; New York, NY **2023**.
[21] F. Oosawa, *Polyelectrolytes*, DEKKER, New York **1971**.
[22] M. Muthukumar, *Macromolecules* **2017**, *50*, 9528.
[23] R. Vancraenenbroeck, Y. S. Harel, W. Zheng, H. Hofmann, *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 19506.
[24] A. Firouzbakht, A. Haider, K. Gaalswyk, S. Alaeen, K. Ghosh, M. Gruebele, *Proc. Natl. Acad. Sci.* **2024**, *121*, e2316408121.
[25] G. Manning, *J. Chem. Phys.* **1969**, *51*, 924.
[26] M. Muthukumar, *J. Chem. Phys.* **2004**, *120*, 9343.
[27] A. Chowdhury, A. Borgia, S. Ghosh, A. Sottini, S. Mitra, R. Eapan, M. Borgia, T. Yang, N. Galvanetto, M. Ivanovic, P. Lukijanczuk, R. Zhu, D. Nettels, A. Kundagrami, B. Schuler, *Proc. Natl. Acad. Sci.* **2023**, *120*, e2304036120.
[28] M. Phillips, M. Muthukumar, K. Ghosh, *PNAS Nexus* **2024**, *3*, 367.
[29] F. Ruggeri, F. Zosel, N. Mutter, M. Rozycka, M. Wojtas, A. Ozyhar, B. Schuler, M. Krishnan, *Nat. Nanotechnol.* **2017**, *12*, 488.
[30] M. Fossat, A. Posey, R. Pappu, *Biophys. J.* **2021**, *120*, 5438.
[31] M. Fossat, X. Zeng, R. Pappu, *J. Chem. Phys. B* **2021**, *125*, 4148.
[32] M. Fossat, A. Posey, R. Pappu, *ChemPhysChem* **2023**, *24*, 202200746.
[33] B. Schuler, A. Soranno, H. Hofmann, D. Nettels, *Annu. Rev. Biophys.* **2016**, *45*, 207.
[34] L. A. Metskas, E. Rhoades, *Annu. Rev. Phys. Chem.* **2020**, *71*, 391.
[35] A. A. Deniz, S. Mukhopadhyay, E. A. Lemke, *J. R. Soc., Interface* **2008**, *5*, 15.
[36] A. Holla, E. W. Martin, T. Dannenhoffer-Lafage, K. M. Ruff, S. L. B. König, M. F. Nüesch, A. Chowdhury, J. M. Louis, A. Soranno, D. Nettels, R. V. Pappu, R. B. Best, T. Mittag, B. Schuler, *JACS Au* **2024**, *4*, 4729.
[37] T. M. Kapłon, G. Rymarczyk, M. Nocula-Ługowska, M. Jakób, M. Kochman, M. Lisowski, Z. Szewczuk, A. Ożyhar, *Biomacromolecules* **2008**, *9*, 2118.
[38] K. Gast, H. Damaschun, K. Eckert, K. Schulze-Forster, H. R. Maurer, M. Muller-Frohne, D. Zirwer, J. Czarnecki, G. Damaschun, *Biochemistry* **1995**, *34*, 13211.
[39] S. F. Edwards, P. Singh, *J. Chem. Soc., Faraday Trans. 2* **1979**, *75*, 1020.
[40] M. Muthukumar, *J. Chem. Phys.* **1987**, *86*, 7230.
[41] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, et al., *Nature* **2021**, *596*, 583.
[42] M. Pajkos, I. Clerc, C. Zanon, P. Bernado, J. Cortes, *J. Mol. Biol.* **2025**, *437*, 169003.
[43] D. McQuarrie, *Statistical Mechanics*, Harper & Row, **1975**.
[44] M. Muthukumar, *Macromolecules* **2017**, *50*, 9528.
[45] M. Muthukumar, *Physics of Charged Macromolecules: Synthetic and Biological Systems*, Cambridge University Press, **2023**.
[46] M. Muthukumar, *J. Chem. Phys* **1984**, *81*, 6272.
[47] T. Firman, K. Ghosh, *J. Chem. Phys.* **2018**, *148*, 123305.
[48] L. Houston, M. Phillips, A. Torres, K. Gaalswyk, K. Ghosh, *J. Chem. Theory Comput.* **2024**, *20*, 10266.
[49] P. Higgs, J. Joanny, *J. Chem. Phys.* **1991**, *94*, 1543.
[50] M. A. Hass, F. A. Mulder, *Annu. Rev. Biophys.* **2015**, *44*, 53.
[51] S. Ghosh, A. Chowdhury, D. T. Tomares, B. Schuler, A. Kundagrami, R. V. Pappu, J. Phys. Chem, in press.
[52] A. Haider, K. Gaalswyk, L. Houston, N. Ose, S. Ozkan, K. Ghosh, *J. Phys. Chem. B.* **2025**, *129*, 10285.
[53] L. N. Dimitrova, K. Kuroha, T. Tatematsu, T. Inada, *J. Biol. Chem.* **2009**, *284*, 10343.
[54] W. Zheng, G. H. Zerze, A. Borgia, J. Mittal, B. Schuler, R. B. Best, *J. Chem. Phys.* **2018**, *148*, 123329.
[55] E. D. Holmstrom, A. Holla, W. Zheng, D. Nettels, R. B. Best, B. Schuler, *Accurate Transfer Efficiencies, Distance Distributions, and Ensembles of Unfolded and Intrinsically Disordered Proteins From Single-Molecule FRET*, Elsevier, **2018**, pp. 287–325.
[56] *Y. Nakayama, J. Colloid and Interface Science* **2023**, *646*, 354.
[57] S. Wohl, M. Jakubowski, W. Zheng, *J. Phys. Chem. Lett.* **2021**, *12*, 6684.
[58] A. Vitalis, R. V. Pappu, *J. Comput. Chem.* **2009**, *30*, 673.
[59] J. Huihui, K. Ghosh, *Biophysical J.* **2021**, *120*, 1860.
[60] J.-E. Shea, R. B. Best, J. Mittal, *Curr. Opin. Struct. Biol.* **2021**, *67*, 219.
[61] M. Ghasemi, R. Larson, *Prog. Polym. Sci.* **2021**, *112*, 101322.

Supporting Information for
Mapping charge interactions in intrinsically disordered proteins

Michael Phillips[#] Andrea Holla[#] Magdalena Wojtas[#] Aritra Chowdhury[#] Andrea Sottini Sebastian L. B. König Natalie Mutter Nick Lamb Jonathan Huihui Monika Lopko Andrea Soranno Daniel Nettels Andrzej Ożyhar Benjamin Schuler Kingshuk Ghosh**

[#]These authors contributed equally to this work
Dr. M. Phillips
Department of Physics and Astronomy, University of Denver, Denver, CO, USA
Dr. A. Holla
Department of Biochemistry, University of Zurich, Zurich, Switzerland
Dr. M. Wojtas
Department of Biochemistry, University of Zurich, Zurich, Switzerland
Department of Biochemistry, Faculty of Chemistry, Wrocław University of Technology, Wrocław, Poland
Dr. A. Chowdhury
Department of Biochemistry, University of Zurich, Zurich, Switzerland
Dr. A. Sottini
Department of Biochemistry, University of Zurich, Zurich, Switzerland
Dr. S. L. B. König
Department of Biochemistry, University of Zurich, Zurich, Switzerland
N. Mutter
Department of Biochemistry, University of Zurich, Zurich, Switzerland
N. Lamb
Department of Physics and Astronomy, University of Denver, Denver, CO, USA
Dr. J. Huihui
Department of Physics and Astronomy, University of Denver, Denver, CO, USA
Dr. M. Lopko
Department of Biochemistry, Faculty of Chemistry, Wrocław University of Technology, Wrocław, Poland
Prof. Dr. A. Soranno
Department of Biochemistry, University of Zurich, Zurich, Switzerland Department of Biochemistry and Molecular Biophysics, Center for Biomolecular Condensates, Washington University in St. Louis, St. Louis, MO, United States
Dr. D. Nettels
Department of Biochemistry, University of Zurich, Zurich, Switzerland
Prof. Dr. A. Ożyhar
Department of Biochemistry, Faculty of Chemistry, Wrocław University of Technology, Wrocław, Poland
Prof. Dr. B. Schuler
Department of Biochemistry and Department of Physics, University of Zurich, Zurich, Switzerland
Email address: schuler@bioc.uzh.ch
Prof. Dr. K. Ghosh
Department of Physics and Astronomy, University of Denver, Denver, CO, USA
Email address: kingshuk.ghosh@du.edu

S1 Protein purification and labeling

The disordered linker sequences from RNA-binding proteins (see Table S1) encoded in pET-20b(+) expression vectors were expressed in *E. coli* and purified as described by Holla et al. [1]. Briefly, linker constructs were expressed as fusion constructs with N- and C-terminal fusions of a His_6-tag and a GB1-His_6-tag, respectively (both of which were separated from the IDR of interest by a thrombin cleavage site) in *E. coli* Rosetta (DE3) cells (Merck Biosciences) and purified under denaturing conditions (6M guanidinium chloride) using immobilized metal ion affinity chromatography (IMAC); dialysis to non-denaturing conditions followed by thrombin-induced tag cleavage, concentration using 3-kDa molecular weight cut-off centrifugal concentrators, reduction and a final purification using reversed-phase high performance liquid chromatography (RP-HPLC) on a C18 column (Reprosil Gold 200, Dr. Maisch GmbH) using a water/acetonitrile gradient yielded the linkers, which were lyophilized and used for labeling. The linkers were labeled with AlexaFluor 488/AlexaFluor 594 or Cy3B/CF660R as described previously using a sequential labeling approach in which the lyophilized reduced proteins were reacted with substoichiometric (molar ratio 0.7) Alexa 488 or Cy3B (donor dyes) maleimides, followed by separation of the donor-labeled species using RP-HPLC on a C18 column (Reprosil Gold 200, Dr. Maisch GmbH), which was subsequently labeled with a stoichiometric excess ($\tilde{3}$-fold molar excess) of Alexa 594 (for Alexa 488-labeled samples) or CF660R maleimide (for Cy3B-labeled samples); purification of the double-labeled species containing donor and acceptor and labeling permutations (where possible) was achieved with a subsequent final RP-HPLC step.

Stm variants containing pairs of cysteine residues instead of alanine were prepared. Alanine codons were replaced by cysteine codons by site-directed mutagenesis using a PCR-based method described by Ko and Ma[2]. In the first stage, the Stm gene was amplified into two separate PCR fragments using two pairs of anchor and mutagenic primers. The anchor primers contained restriction sites for subcloning the mutated gene into corresponding sites of a pQE-80L (Qiagen, Germany) vector derivative generated in our laboratory without His-tag. The forward and reverse anchor primers contained BamHI and HindIII restriction sites, respectively, and the two mutagenic primers contained the desired mutation near the recognition site of the BsmBI restriction enzyme. After digestion of the two PCR fragments with the appropriate enzymes, both fragments were ligated together and into the vector to generate the construct with the mutated gene. To introduce a second mutation, the entire procedure was repeated using the previously obtained mutant as the template for PCR. Alternatively, when possible, the second mutation was introduced during the first stage using a mutated anchor primer. The DNA sequences of all double mutants were confirmed by sequencing.

Double cysteine variants of Starmaker without the signal peptide (Stm variants; see Table S1 for sequences) were purified as wild-type Stm (WT Stm) following Kaplon et al. [3]. Untagged Stm variants, cloned into the pQE-80L vector (Qiagen, Germany), were overexpressed in BL21(DE3) pLysS *E. coli* cells (Novagen, Germany) and purified to homogeneity using a three-step protocol: fractionation with solid $(NH_4)_2SO_4$, gel filtration, and hydroxyapatite chromatography. The samples were then reduced, concentrated, and further purified by RP-HPLC on a Vydac 214 TPC4 column using a water/acetonitrile gradient, then lyophilized for subsequent labeling. For labeling, the Stm pellet was diluted in 200 µL of 100 mM phosphate buffer (pH 7.0). The amount of free sulfhydryl groups in Stm was estimated using 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB). To obtain pure donor- and acceptor-labeled Stm, a sequential labeling procedure was used. The first fluorophore (AlexaFluor 488 maleimide or AlexaFluor 594 maleimide), dissolved in DMSO, was added to the Stm solution at a 1 :1 molar ratio and allowed to react for 1 h at room temperature. The reaction was quenched by adding 10 mM DTT. The labeled Stm species were purified by RP-HPLC on a Vydac 214 TPC4 column. In most cases, the single-labeled Stm could be separated from unlabeled and double-labeled species; the fraction containing single-labeled Stm was frozen and lyophilized. For labeling with the second dye (AlexaFluor 594 or AlexaFluor 488 maleimide), the single-labeled Stm was diluted in 60 µL of 100 mM phosphate buffer (pH 7.0). The protein concentration was estimated from the maximum absorbance of the first dye, assuming single labeling. The second dye was then added in five- to ten-fold molar excess and allowed to react at room temperature for 1 h to 2 h. The reaction was terminated by adding 2 µL 2-mercaptoethanol. Finally, the double-labeled Stm was purified by RP-HPLC on a Vydac 214 TPC4 column.

ProTα C (isoform 1, see Table S1 for the sequence) was purified as an N-terminally His_6-tagged construct (the tag was removed by HRV 3C cleavage) and labeled as described previously[4][5]. ProTα NC and ProTα N (isoform 1, see Table S1) were cloned into a pBAD-IntCBD-12His vector and purified as described previously; briefly, the constructs were overexpressed in *E. coli* BL21 A1 cells carrying C-terminal fusions of the intein-chitin-binding domain-His_{12}, purified under native conditions with IMAC, followed by thiol-induced tag cleavage, dialysis, and reverse IMAC to remove the cleaved tag. The tag-free protein was concentrated with 3 kDa molecular-weight cut-off centrifugal filters and used for fluorescence labeling without further purification[5]. In brief, the reduced protein (treated with 10 mM TCEP followed by buffer exchange with centrifugal filters) was simultaneously labeled with ∼6.5 fold molar excess of Alexa 488 and Alexa 594 maleimide. After quenching the reaction and removing excess dye by sequential buffer exchange with centrifugal filtration, the double-labeled protein was purified by RP-HPLC on a Sunfire C18 column (Waters).

S2 Single-molecule measurements and analysis

Single-molecule experiments were performed on freely diffusing molecules with a MicroTime 200 confocal single-molecule instrument (PicoQuant, Berlin, Germany) or a custom-built setup as described previously[1]. Depending on the fluorophore under investigation, different excitation sources were used. Alexa 488 was excited using an LDH-D-C-485 diode laser (PicoQuant GmbH), whereas Alexa 594 and Cy3B were excited with a supercontinuum fiber laser (SC-450-4, Fianium Ltd.) that was filtered with a z582/15 or HC543.5/2 band-pass filter, respectively (Chroma Technology). For CF660R excitation, an LDH-D-C-640 diode laser (PicoQuant GmbH) was used. All lasers were operated at a pulse repetition rate of 20 MHz, which allowed for the pulsed interleaved excitation (PIE) of donor and acceptor molecules[6]. Fluorescence photons were collected using a UplanApo 60x/1.20W objective (Olympus) and passed through an appropriate multiband mirror and a 100-$\hat{\text{I}}\frac{1}{4}$m confocal pinhole. The photons were subsequently separated by polarization with a polarizing beam splitter and by wavelength with appropriate dichroic mirrors. Finally, after passing through optical band-pass filters, the photons were detected by avalanche photodiodes, and their arrival times were recorded using a HydraHarp 400 time-correlated single-photon counting system (PicoQuant) with a resolution of 16 ps.

Experiments were performed at a temperature of 22 °C using chambered cover slides (µ-Slide, ibidi). For the linker sequences, experiments employed approximately 100 pM labeled molecules in 20 mM KH_2PO_4/K_2HPO_4 buffer (pH 7.3) with different KCl concentrations, supplemented with 0.001 % Tween 20 and 10 mM DTT for the Cy3B/CF660R-labeled samples, or with 147 mM 2-mercaptoethanol for the Alexa dye-labeled samples. Measurements in ionic strength-matched Bis-Tris or Tris buffer with KCl showed minor differences in transfer efficiencies of less than 0.02 for sCh, sNh-, and dTRBP compared to the reported measurements in phosphate buffer. For Stm and ProTα (all Alexa 488- and Alexa 594-labeled samples), measurements were performed in 10 mM Tris-HCl buffer (pH 7.4) with different KCl concentrations, 0.01 % Tween 20, and 147 mM 2-mercaptoethanol.

Fluorescence signals were acquired from single molecules freely diffusing through the confocal volume, and transfer efficiency E and the stoichiometry ratio S were quantified from selected photon bursts, each burst representing a single molecule traversing the observation volume, according to

$$E = \frac{N_{AD}}{N_{DD} + N_{AD}}, \quad S = \frac{N_{DD} + N_{AD}}{N_{DD} + N_{AD} + N_{AA}},$$

where N_{DD} and N_{AD} are donor and acceptor photons after donor excitation,respectively, that are corrected for background signals, direct acceptor excitation, and differences in dye quantum yields or detector efficiencies. N_{AA} correspond to the acceptor photons after acceptor excitation, corrected for background and for relative donor/acceptor excitation intensities as described previously [7] [8]. For Stm and ProTα measurements, the correction factors were obtained from the PIE measurements as described previously[9][8]. Data analysis was performed with Fretica, a custom add-on for Mathematica (Wolfram Research), available at https://github.com/SchulerLab.

Transfer efficiency histograms obtained under different experimental conditions were fitted with a Gaus-

sian peak function (for linkers) or by Photon Distribution Analysis (PDA)[10][11] to extract the mean transfer efficiency $\langle E \rangle$. The RMS end-to-end inter-dye distances were then inferred by numerically solving

$$\langle E \rangle = \int_0^\infty P(r)\, E(r)\, \mathrm{d}r, \quad E(r) = \frac{R_0^6}{R_0^6 + r^6},$$

where $P(r)$ is the distance probability density function, and R_0 is the Förster radius of the chosen fluorophore pair (5.4 nm for Alexa 488/Alexa 594 and 6.0 nm for Cy3B/CF660R). An empirically modified self-avoiding random-walk model (SAW-ν) was employed for $P(r)$[12]. The length scaling exponent ν for the IDP segments probed was estimated assuming an equivalent of nine additional amino acids accounting for the fluorophores and their linkers[8, 12], and the RMS C_α distances were then calculated using that value of ν and the number of peptide bonds in the chain segment bracketed by the labeling positions[12, 13]. The origin of the large uncertainties in inter-dye distances for Stm and ProTα variants probing long and highly charged segments at low salt concentrations (Fig. 6) is a result of very low observed transfer efficiencies, E: If the value of E approaches or falls below the systematic uncertainty of 0.03, we can effectively only provide a lower bound on the distance; as a result, the upper error bars diverge.

Table S1: Sequences of all IDRs used.

Name	Sequence	Charge
dArich	GSGSCLPTGAEGRDSSKGEDSAEETEAKPAVVAPAPVVEAVSTPSAAFPSDATAEQGPILTKCTLGPR	−10/ + 4
dErich	GSGSCEGEMFGVMSEENSVVLSVEQPAELKEVADVSPPPTTRNHTIEMKPPLSAQQSESNPVCTLGPR	−10/ + 3
dGrich	GSGSCGPRTGLEGAGMAGGSGQQKRVFDGQSGPQDLGEAYRPLNHDGGDGGNRYSVIDRIQECTLGP R	−8/ + 6
sGrich	GSGSCSGGYGSERGGGYGSERGGGYGSERGGGYGSERGGGYGSQRSGGGYGGSQRSSYGSGSCTLGP R	−4/ + 6
sNrich	GSGSCLDQEDNNGPLLIKTANNLIQNNSNMLPLNALHNAPPMHLNEGGISNMRVNDSLPSNTCTLGPR	−5/ + 2
sCh	GSGSCKKAQPEMNDKDDNESGNEDAEENHDDEEDENEEEDRQVDQASKNKESKRKAQNKREDCTLG PR	−23/ + 11
dCh +	GSGSCQTPLKRIKVKTPGKSGAAARE GSVVSGTDGPTQTGKPERRKRLNPPKDKLIDMDDADCTLGP R	−8/ + 13
dCh −	GSGSCMGLPTGMEEKEEGTDESEQKPVVQTPAQPDDSAEVDSAALDQAESAKQQGPILTKHGCTLGP R	−13/ + 4
dTRBP	GSGSCLEPALEDSSSFSPLDSSLPEDIPVFTAAAAATPVPSVVLTRSPPMELQPPVSPQQSECTLGPR	−8/ + 1
sPTBP	GSGSCRPDLPSGDSQPSLDQTMAAAFGLSVPNVHGALAPLAIPSAAAAAAAAGRIAIPGLAGCTLGPR	−3/ + 2
dκh	GSGSCAMGGGPGPGTDFTSDQADFPDTLFQEFEPPAPRPGLAGGRPGDAALLSAAYGRRRLLCTLGP R	−7/ + 5
sκh	GSGSCKPRRLSKLRRSKKPADEENNAASQDPTVEATQERGQASEDPENAANNAKQAKPTSDDCTLGP R	−11/ + 11
dκl	GSGSCLPTGMDQKTTELMKVDEPVSTQETLPPVIKMEPEPVPITETPTDENARQQGPILTKHCTLGPR	−10/ + 5
sκl	GSGSCAPEGVFKLPAPPKEIKKSEKSEKSSDSINDKKEVESTTKTAATTTTTKKGTDNNTQFCTLGPR	−9/ + 11
sNh +	GSGSCPEEIETRKKDRKNRREMLKRTSALQPAAPKPTHKKPVPKRNVGAERKSTINEDLLPPCTLGPR	−8/ + 16
sNh −	GSGSCKKKLQDTPSEPMEKDPAEPETVPDGETPEDENPTEEGADNSSAKMEEEEEEEEEEEECTLGPR	−26/ + 5
Stm 36_54	GSAPVSNNNGTDNDESAADQRHILTVQFNVGTPAPCDGDSVTTDGKDSAEKNECPGDSSDTTEKPGT TDGKDSAEQHGVTTDGKDEAEQHGVTTEGQDSAEKRGETDGAPDKPDTQNGTDDTDSDKETDASHH KTGDSDENKDKPSAEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDED ANKSSTEADKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNS EADEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTADSKDSDEDKDKDHSEQKDSEDHEHKEK HTKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKEASKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDS HSDSDSDSHSDSDSDSDSHSDSDSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSE SSETAEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDD DDIDAHDDEAGVEHGTDEASKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	−15/ + 5
Stm 54_109	GSAPVSNNNGTDNDESAADQRHILTVQFNVGTPAPADGDSVTTDGKDSAEKNECPGDSSDTTEKPGTT DGKDSAEQHGVTTDGKDEAEQHGVTTEGQDSAEKRGETDGCPDKPDTQNGTDDTDSDKETDASHHK TGDSDENKDKPSAEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDEDA NKSSTEADKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNSE ADEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTADSKDSDEDKDKDHSEQKDSEDHEHKEKH TKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKEASKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDSH SDSDSDSHSDSHSDSDSDSHSDSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESS ETAEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDDD DIDAHDDEAGVEHGTDEASKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	
Stm 87_147	GSAPVSNNNGTDNDESAADQRHILTVQFNVGTPAPADGDSVTTDGKDSAEKNEAPGDSSDTTEKPGTT DGKDSAEQHGVTTDGKDECEQHGVTTEGQDSAEKRGETDGAPDKPDTQNGTDDTDSDKETDASHHK TGDSDENKDKPSCEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDEDA NKSSTEADKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNSE ADEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTADSKDSDEDKDKDHSEQKDSEDHEHKEKH TKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKEASKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDSH SDSDSDSHSDSHSDSDSDSHSDSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESS ETAEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDDD DIDAHDDEAGVEHGTDEASKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	+18/ + 7
Stm 209_272	GSAPVSNNNGTDNDESAADQRHILTVQFNVGTPAPADGDSVTTDGKDSAEKNEAPGDSSDTTEKPGTT DGKDSAEQHGVTTDGKDEAEQHGVTTEGQDSAEKRGETDGAPDKPDTQNGTDDTDSDKETDASHHK TGDSDENKDKPSAEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDEDA NKSSTECDKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNSE CDEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTADSKDSDEDKDKDHSEQKDSEDHEHKEKH TKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKEASKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDSH SDSDSDSHSDSHSDSDSDSHSDSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESS ETAEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDDD DIDAHDDEAGVEHGTDEASKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	−27/ + 12
Stm 311_374	GSAPVSNNNGTDNDESAADQRHILTVQFNVGTPAPADGDSVTTDGKDSAEKNEAPGDSSDTTEKPGTT DGKDSAEQHGVTTDGKDEAEQHGVTTEGQDSAEKRGETDGAPDKPDTQNGTDDTDSDKETDASHHK TGDSDENKDKPSAEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDEDA NKSSTEADKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNSE ADEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTCDSKDSDEDKDKDHSEQKDSEDHEHKEKH TKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKECSKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDSH SDSDSDSHSDSHSDSDSDSHSDSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESS ETAEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDDD DIDAHDDEAGVEHGTDEASKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	−30/ + 14
Stm 17_87	GSAPVSNNNGTDNDESCADQRHILTVQFNVGTPAPADGDSVTTDGKDSAEKNEAPGDSSDTTEKPGTT DGKDSAEQHGVTTDGKDECEQHGVTTEGQDSAEKRGETDGAPDKPDTQNGTDDTDSDKETDASHHK TGDSDENKDKPSAEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDEDA NKSSTEADKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNSE ADEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTADSKDSDEDKDKDHSEQKDSEDHEHKEKH TKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKEASKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDSH SDSDSDSHSDSHSDSDSDSHSDSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESS ETAEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDDD DIDAHDDEAGVEHGTDEASKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	

Continued on next page

Table S1 (continued)

Name	Sequence	Charge
Stm 484_567	GSAPVSNNNGTDNDESAADQRHILTVQFNVGTPAPADGDSVTTDGKDSAEKNEAPGDSSDTTEKPGTT DGKDSAEQHGVTTDGKDEAEQHGVTTEGQDSAEKRGETDGAPDKPDTQNGTDDTDSDKETDASHHK TGDSDENKDKPSAEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDEDA NKSSTEADKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNSE ADEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTADSKDSDEDKDKDHSEQKDSEDHEHKEKH TKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKEASKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDSH SDSDSDSHSDSHSDSDSDSHSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESS ETCEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDDD DIDAHDDEAGVEHGTDECSKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	−38/ + 5
Stm 374_484	GSAPVSNNNGTDNDESAADQRHILTVQFNVGTPAPADGDSVTTDGKDSAEKNEAPGDSSDTTEKPGTT DGKDSAEQHGVTTDGKDEAEQHGVTTEGQDSAEKRGETDGAPDKPDTQNGTDDTDSDKETDASHHK TGDSDENKDKPSAEDHTDGNHAGKDSTDSKESPDTTDKPEGPDSDSAPDGDSASAEKTDSDHSPDEDA NKSSTEADKDDTSDKDSSQTDEKHDSDASDKDEKHEDKDEKSDEKDSSKDSEDKSLEKSDKSDDGSNSE ADEQKESVESKDHDSDSQDSDSAEKKEKHDDKDQDSSDTADSKDSDEDKDKDHSEQKDSEDHEHKEKH TKDKEEHKDSDEGKDDEDKSKSDEHDKDESESKECSKSDESEQEEKKDDKSDSDNSSRDSHSDSDSDSH SDSDSDSHSDSHSDSDSDSHSDSDSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESS ETCEEDTNDDKDSSVEKDKTDSSDSASVEANDSDDEHDDDSKDATPSSEDHTAEKTDEDSHDVSDDDD DIDAHDDEAGVEHGTDEASKPHQEPDHHDDTTHGSDDGSKTSMPISKLN	−42/ + 12
ProTα N	GCDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAENEENGEQEADNEVDEECEEGGEEEEEE EEGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQKTDEDDGA	−20/ + 5
ProTα C	GPSDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAENEENGEQEADNEVDEECEEGGEEEEEE EEGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQKTDEDC	−32/ + 5
ProTα NC	GCDAAVDTSSEITTKDLKEKKEVVEEAENGRDAPANGNAENEENGEQEADNEVDEEEEEGGEEEEEEE EGDGEEEDGDEDEEAESATGKRAAEDDEDDDVDTKKQKTDEDCGA	−53/ + 10

S3 Modeling the FRET Dyes

The properties of the fluorophores, especially their charge and finite size, should be included in the model to account for their effect on the observed dye-to-dye distance. Moreover, the charges from the terminal α-amino and α-carboxyl groups need to be accounted for. For the IDRs from RNA-binding proteins, we modeled the dyes as short sequences of residues that approximate the size and charge of the fluorophores. Dyes are attached to Cysteine residues near the terminals of the sequences, with small tails at each end (Table S1); each sequence begins with `GSGS` and ends with `TLGPR`. To account for the terminal charges, we replaced the first residue `G` $\rightarrow$ `K`, and the last residue `R` $\rightarrow$ `A`. Short sequences of 4 and 5 residues approximate the size of the dyes[8] for both dye pairs (Alexa 488/594 or Cy3B/CF660R, see Figure S1). Since the donor/acceptor labeling permutations are not well defined in all cases[1], we modeled both labeling permutations and found negligible differences in inter-dye distances. For simplicity, all results are presented with dyes modeled with the donor on the N- and the acceptor at the C-terminus, and we represent the charge content of each dye as a linear sequence of suitably charged residues. Alexa 488 was thus modeled as `EEKE`, and Alexa 594 as `EAEKE`; in combination with the amino acid tails and the terminal charges, we ultimately replaced `GSGS`$\rightarrow$`AEKE` at the N-terminus and `TLGPR`$\rightarrow$`EAEKE` at the C-terminus. Simliarly, Cy3B was modeled as `EKAA`, and CF660R as `EEKKE`; again combining with the terminal charges, we ultimately replaced `GSGS`$\rightarrow$`EKAA` at the N-terminus and `TLGPR`$\rightarrow$`EEKKE` at the C-terminus.

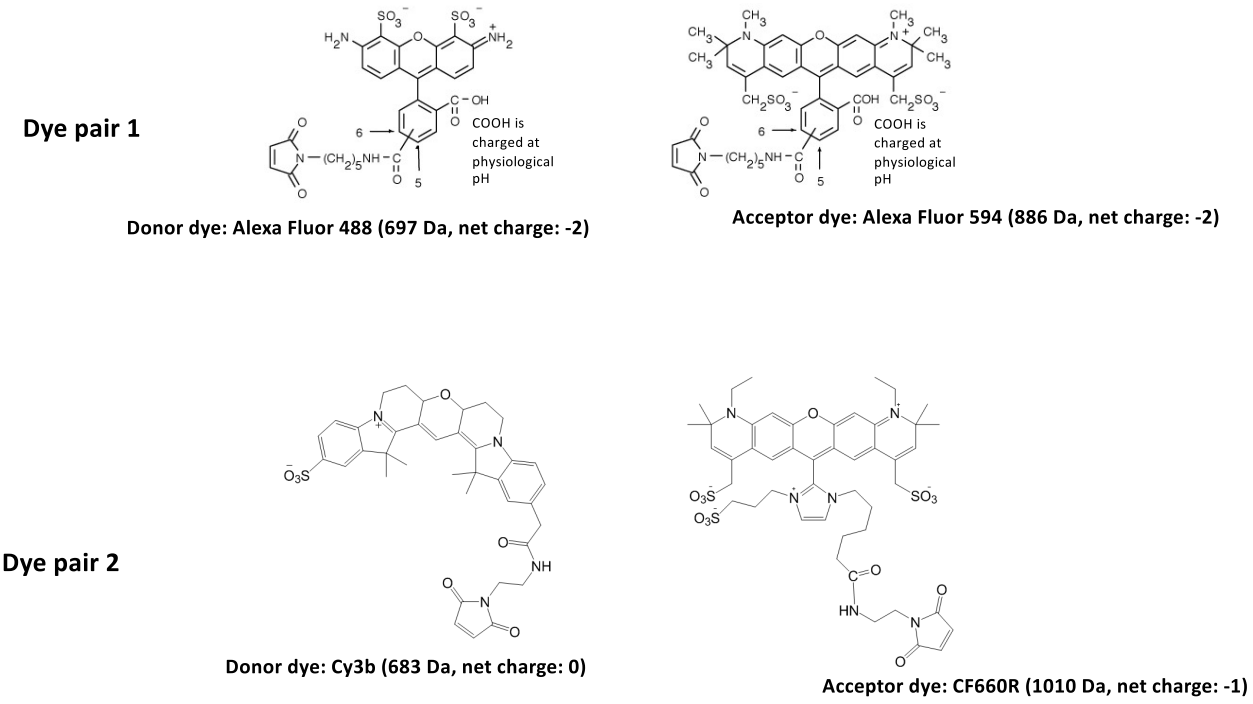

Figure S1: FRET dye pairs used for labeling of the linker sequences. For Stm and ProTα, only dye pair 1 (Alexa 488/594) was used. To approximate their size and charge, we modeled the first of each pair (left) as a 4-residue segment at the N-terminus, and the second of each pair (right) as a 5-residue segment at the C-terminus.

For long and highly charged sequences, such as Stm and ProTα, the dyes have a minor influence on chain properties. Different segments of these sequences were probed by placing Cys (`C`) for dye labeling at specific positions along the chain, which precludes the possibility of modeling the size of the fluorophores in the detail used for linker sequences; instead, we modeled the dyes by simply replacing Cys by single residues of appropriate net charge. Due to their large lengths and charge, we also omitted adjustments from terminal charges as their effects would be negligible. Only the dye pair Alexa 488/594 was used in

both Stm and ProTα, so we ultimately just replaced each `C`→`X` with charge -2. For the sake of comparison between the two approaches of representing the dyes, we also modeled the dyes in the linker sequences without including additional residues accounting for their size, using only the net charge of the dyes as single residues at the Cys labeling positions, but keeping the terminal charge adjustments as described above. For Alexa 488/594, we replaced each `C` by a residue `X` with charge -2. For Cy3B/CF660R, we left the first `C` unchanged since Cy3B is net neutral, and we replaced the second `C` by `D` since CF660R has net charge -1. Note that full sequences of Table S1 are used in the inter-residue theory for RMS distances $\sqrt{\langle R_{ij}^2 \rangle}$, not just the segments between dyes at locations (i, j), because the theory accounts for context provided by the parts of the sequence outside of the probed segments.

We also note that charge assignments of residues in all sequences are `K,R` $\rightarrow +1$ and `E,D` $\rightarrow -1$. In most sequences, we consider Histidine as neutral. This approximation is justified since the linker IDRs have at most 2 His residues, and ProTα has zero. However, Stm has 32 His residues so for those sequences we use the charge assignment `H` $\rightarrow +0.5$.

S4 Details of free energy

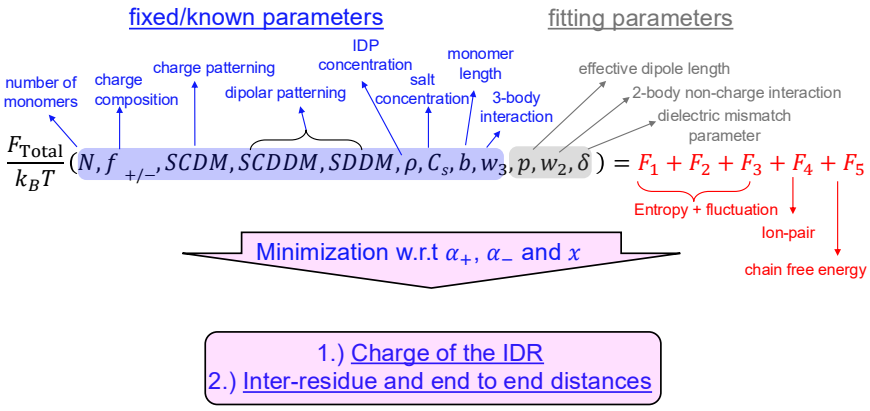

Figure S2: A schematic of the free energy and its optimization. The different contributions to the total free energy are depicted in red. The free energy is minimized with respect to effective ionization and swelling ratio to determine optimal charge and distance. Several parameters are known and fixed (in blue), and only the parameters in gray are inferred from fitting the experimental data.

S5 Higgs-Joanny Model

We describe here the specific form of the Higgs-Joanny model (H-J) used in main text Figure 2 and 4, as well as SI Figures S4 and S9. We present the H-J model using the same swelling/compaction variable x as employed in the sequence patterning models (M0, M1, M2), related to RMS end-to-end distance by $\langle R_{ee}^2 \rangle = N\ell b x$. This notation differs from the original formulation, which also defines a swelling/compaction parameter but in a different way[14].

For a sequence with N residues, N_+ with positive charge (K, R) and N_- with negative charge (E, D), and the rest neutral (either hydrophobic or polar), the fractions of each charge type are $f_+ = N_+/N$ and $f_- = N_-/N$. For the Debye length, κ^{-1} in Debye-Hückel screening, we use $\kappa^2 = 8\pi\ell_B c_s$, where ℓ_B is the Bjerrum length and c_s is the salt concentration (equivalent to ionic strength in our case of monovalent ions).

Essentially, the H-J model is equivalent to a self-avoiding walk model with the two-body interaction modulated by electrostatic effects. The bare two-body non-electrostatic interaction is ω_2. The two-body interaction is then combined with electrostatic effects due to the polyelectrolyte and polyampholyte properties of the sequence related to net charge and total charge, respectively (see Ref. [14] for details). The total two-body interaction strength is given by

$$v_2 = \omega_2 + \frac{4\pi\ell_B(f_+ - f_-)^2}{(\ell b)^{3/2}\,\kappa^2} - \frac{\pi\ell_B^2(f_+ + f_-)^2}{(\ell b)^{3/2}\,\kappa} \tag{S1}$$

Note that we use a geometric mean between bond length ($b = 3.8\,\text{Å}$) and Kuhn length ($\ell = 8\,\text{Å}$) as the single length scale in the H-J model, $\sqrt{\ell b} \approx 5.5\,\text{Å}$, to reconcile the original H-J formulation with the relation $\langle R_{ee}^2 \rangle = N\ell b x$. We note that the Higgs-Joanny model is strictly applicable only in the high-salt limit ($\kappa^2 \ell b \gg 1$) and formally diverges as the salt concentration approaches zero. This necessitates the introduction of a three-body repulsive interaction to maintain physical results at low salt (52 mM), as described below.

Although the original formulation of Higgs and Joanny is in terms of an equation of state, we integrate that to obtain the effective free energy. With the above two-body interaction strength, the H-J model adopts the same form as our model M0 apart from the sequence-dependent electrostatic contribution, i.e. it reads as the main text Eq. 6 with $Q = 0$. Two-body interactions are included in the term Ω in the same way as in the main text Eq. 12, replacing $\omega_{2,ee} \to v_2$. Based on that, we write the weight of two-body interactions here with the double summation

$$\Omega_0 = \frac{1}{N}\sum_{m=2}^{N}\sum_{n=1}^{m-1}(m-n)^{-1/2} \tag{S2}$$

The original H-J model does not include any three-body interaction, but we add it here because it is needed to prevent unphysical chain collapse in the strong polyampholyte limit, which applies for a subset of the linker IDRs. The three-body term has the same form as in main text Eq. 6, with weight B given by main text Eq. 7.

The effective free energy for the H-J model as used in this work is then

$$\beta F_{HJ} = \frac{3}{2}(x - \ln(x)) + v_2\Omega_0\left(\frac{3}{2\pi x}\right)^{3/2} + \frac{\omega_3 B}{2}\left(\frac{3}{2\pi x}\right)^3 \tag{S3}$$

As with our model M0, the above free energy for the H-J model is minimized to obtain x, and thus $\langle R_{ee}^2 \rangle$, at each salt concentration. Note that differentiating Eq. S3 and setting $\omega_3 = 0$ in the limit of large N recovers the original H-J formulation of Ref. [14], Eq. 4.12.

S6 Generating normalized distance maps

The overall theoretical approach for calculating distance maps is explained in main text Section 2.2. Briefly, we have a free energy that depends on sequence (charge patterning) and effective charge, which determines the inter-residue compaction or swelling through the chain factor x_{ij}, for each residue index pair (i, j). We then obtain the actual RMS distance from the relation $\langle R_{ij}^2 \rangle = x_{ij}\,|i-j|b\ell$, using the bond length $b = 3.8\,\text{Å}$ and Kuhn length $\ell = 8.0\,\text{Å}$.

It is possible to construct inter-residue profile maps using the RMS distance directly, $\sqrt{\langle R_{ij}^2 \rangle}$. However, depending on the sequence, the features of such a direct map are generally dominated by homopolymer scaling, i.e. $\sqrt{\langle R_{ij}^2 \rangle} \propto |i-j|^{\nu}$, where ν is a scaling exponent. In particular, differences between sequences of the same length can be very subtle in the direct mapping, even though the theory includes sequence dependence. Further, differences between models (e.g. for the linker IDRs, the full ionization model, M0, and the fitted effective charge model with counterion condensation, M1) can appear negligible based on direct maps.

For better comparison between sequences and between models, and to delineate subtle features of the maps, we therefore sought a way to normalize the distance maps, focusing on deviations from the homopolymer model as was done in the previous work [1, 15, 16]. We term the normalized distance map d_{ij}^*. This normalization divides the homopolymer contribution as

$$\begin{aligned} d_{ij}^* &= \frac{\sqrt{\langle R_{ij}^2 \rangle}}{A_0 |i-j|^{\nu}} \\ &= \frac{\sqrt{b\ell x_{ij}}\;|i-j|^{0.5-\nu}}{A_0} \end{aligned} \tag{S4}$$

where the prefactor A_0 and exponent ν must be determined from a fitting procedure. As eq. S4 makes clear, the normalization is not simply mapping x_{ij}.

The fitting procedure to determine A_0 and ν first requires the RMS distance $\sqrt{\langle R_{ij}^2 \rangle}$ averaged across residues with fixed index separation $\Delta = |i-j|$,

$$R_{\Delta} = \frac{1}{N-\Delta} \sum_{i=1+\Delta}^{N} \left[\sqrt{\langle R_{ij}^2 \rangle}\, \right]_{j=i-\Delta} \tag{S5}$$

We then use a least-squares algorithm to fit eq. S5 to the homopolymer form, $R_{\Delta} \approx A_0\,\Delta^{\nu}$. The best fit values of A_0 and ν (in the range $\nu \in [0.3, 0.7]$) are then used in eq. S4 to obtain d_{ij}^*.

This procedure was employed for all sequences and models of interest; the resulting normalized maps of d_{ij}^* are presented in main text Figure 5, and SI Figure S12 and S13.

S7 Estimates of key model parameters

S7.1 Dipole Length

The physical separation, p, of a condensed ion from the side chain ion it interacts with is also the length of dipoles in charge-dipole and dipole-dipole interactions. We estimate p based on properties of proteins and their constituent amino acids to apply our model and assess best fit parameters. Charged side chains contain either a carboxylate group (Glutamic acid and Aspartic acid), a protonated amino group (Lysine), or a guanidinium group (Arginine). We consider the ions in the system to be dominated by salt ions, i.e. K^+ and Cl^-, rather than the smaller H^+. The distance separating a negative side chain ion O^- from its condensed counterion K^+ is estimated by the sum of half bond lengths of C-O and K-Cl. Likewise, the distance separating a positive side chain ion from its counterion is the sum of half bond lengths from N-H and K-Cl. Using values from NIST[17] to two significant figures, the bond length of C-O is 1.2 Å, N-H is 1.0 Å, and K-Cl is 2.7 Å. Sums of half bond lengths are then 1.95 Å and 1.85 Å. Taking the mean, we arrive at $p \approx 1.9$ Å. We obtain non-dimensional values by dividing by the bond length between amino acids, $b = 3.8$ Å, giving $\tilde{p} = p/b \approx 0.5$. The range of these values gives approximately $0.48 < \tilde{p} < 0.51$. A second independent estimate of dipole length was established previously based on the Avogadro software (see SI of ref. [18]). The range of values from Avogadro gives approximately $0.48 < \tilde{p} < 0.65$. Based on these considerations and our previous estimates including NaCl as ionic contributions [18], we expect the separation to be approximately in the interval $\tilde{p} \in [0.45, 0.65]$.

S7.2 Protein Concentration

FRET experiments were performed on single molecules, corresponding to very low protein concentrations of approximately 1 pM. Theoretical models with ion condensation involve division by the concentration, which can present numerical difficulties when it is too low. We tested several values and found that results are insensitive to the specific concentration so long as it is significantly less than 1 mM. Ultimately, we used the concentration $\rho = 1\,\mu$M in all theoretical formulations, for all sequences; this is low enough to mimic experimental reality and high enough to support numerical minimization of the free energy. In reduced (non-dimensional) units, the value is $\tilde{\rho} = \rho b^3 = 3.304 \times 10^{-8}$.

S7.3 Dielectric mismatch

Another important parameter in our model is the dielectric mismatch, defined as the ratio of bulk permittivity of water to the local permittivity near a protein, $\delta = \epsilon/\epsilon_l$. To assess our fitted values, we refer to an estimate of local permittivity from a different model fit to data, $\epsilon_l = 45 \pm 13$ [19]. With the standard value for water, $\epsilon = 80$, this suggests an expected ratio $\delta \approx 1.8$. Accounting for reported uncertainty in ϵ_l leads to the estimated range $\delta \in [1.3, 2.5]$.

S8 Parameter values for linkers IDRs from various models

We report the final values of physical parameters for each model used.

The three-body interaction parameter ω_3 was assumed to be 0.1 throughout based on a previous estimate using all-atom simulations [20]. Recent work [21] based on coarse-grained simulations has also shown $\omega_3 = 0.1$ to be a reasonable choice and $\omega_3 = 0.2$ to be optimal. The full ionization model, M0, has only a single free parameter — the two-body interaction strength $\omega_{2,ee}$ that was determined by matching the high-salt data. This parameter can be obtained by a simple algebraic equation (see Appendix in Reference [22]) derived by setting the derivative of free energy to zero, the criterion for the minimum in free energy. Since derivatives can be zero for both the maximum and minimum of the free energy, it is important to ensure that the free energy profile with the inferred $\omega_{2,ee}$ indeed yields a free energy minimum at the chain dimensions that match the experimentally measured value. For all cases, this additional self-consistency check was performed.

The ion condensation models, M1 and M2, differ only in how the dielectric mismatch parameter, δ, is determined. In M1, δ is determined by fitting the end-to-end distance vs salt data, constrained to the interval $\delta \in [1.0, 2.7]$, with $\omega_{2,ee}$ still set to match the high-salt measurement, and the other parameters fixed. In M2, δ is determined from the best linear fit with low-salt end-to-end distance. This line itself is determined by adjusting slope and intercept at each iteration, setting the value of δ for each sequence based on that iteration's line (while ensuring $\delta \geq 1$), and calculating χ^2 across all sequences (except the outlier sNh−); iterations continued until the total χ^2 was minimized. The dipole length was kept fixed at $\tilde{p} = 0.66$ for both M1 and M2 — this is the optimal dipole length determined by global search, repeating the fitting process of M1 for many values of $\tilde{p}$ and selecting the one yielding minimum total χ^2 from all linker IDRs. We use a standard definition of χ^2, for measured data with asymmetric error bars. We write a given measurement of RMS end-to-end distance as R_m and the corresponding upper and lower uncertainties as σ_+ and σ_-, respectively. The corresponding theoretical prediction is then R_p. Note, all these values depend on salt concentration, c_s. For a given sequence S, we obtain a mean chi-squared error across salt concentrations as

$$\chi^2(S) = \frac{1}{N_s} \sum_{\{c_s\}} \frac{(R_p - R_m)^2}{\sigma_+ \sigma_-} \tag{S6}$$

where N_s is the number of different salt concentrations for that sequence. A total value is obtained by simply summing over all sequences under consideration: $\chi^2_{total} = \sum_{\{S\}} \chi^2(S)$. (A mean over sequences can be used alternatively, but it makes no difference for optimization.) This approach was also used for fitting the results from Stm and ProTα, simply using inter-residue distances and their uncertainties instead of end-to-end.

Below we provide parameter values for different models, specified as table columns for comparison. Sequences with different dye pairs, Alexa 488/594 and Cy3B/CF660R, are in separate tables.

Sequence	**M0** $\omega_{2,ee}$	**M1** $\omega_{2,ee}$	**M2** $\omega_{2,ee}$	**M1** δ	**M2** δ
dArich	0.458	0.694	0.747	1.170	1.324
dErich	0.658	0.947	0.983	1.186	1.279
dGrich	-0.105	0.022	0.059	1.393	1.572
sGrich	-0.426	-0.350	-0.321	1.558	1.778
sNrich	0.053	0.167	0.172	1.423	1.467
sCh	-0.362	0.857	0.846	1.558	1.547
dCh+	0.162	0.380	0.376	1.630	1.619
dCh−	0.409	1.024	0.925	1.504	1.310
dTRBP	0.349	0.551	0.653	1.000	1.308
sPTBP	-0.077	-0.052	-0.011	1.000	1.535
dκh	-0.137	-0.013	0.008	1.505	1.629
sκh	0.147	0.411	0.394	1.694	1.650
dκl	1.128	1.324	1.354	1.176	1.272
sκl	0.620	0.761	0.748	1.473	1.434
sNh+	0.230	0.563	0.555	1.572	1.555
sNh−	-0.437	2.660	1.731	1.870	1.255

Table S2: Parameters resulting from each model (M0, M1, M2). Ion condensation models (M1,M2) used a fixed dipole length, $\tilde{p} = 0.66$. The proteins here were labeled with Alexa 488/594.

Sequence	**M0** $\omega_{2,ee}$	**M1** $\omega_{2,ee}$	**M2** $\omega_{2,ee}$	**M1** δ	**M2** δ
dArich	0.955	1.105	1.106	1.325	1.328
dErich	1.193	1.375	1.364	1.339	1.294
dGrich	0.059	0.073	0.150	1.000	1.533
sGrich	-0.237	-0.221	-0.153	1.000	1.661
sNrich	0.468	0.474	0.497	1.000	1.407
sCh	0.395	1.811	1.398	2.034	1.561
dCh+	0.428	0.633	0.652	1.399	1.449
dCh−	0.742	1.213	1.078	1.740	1.352
dTRBP	1.610	1.707	1.751	1.000	1.230
sPTBP	0.234	0.228	0.242	1.000	1.447
dκh	-0.012	0.071	0.059	1.715	1.631
sκh	0.744	1.089	0.897	2.000	1.467
dκl	1.843	1.898	1.943	1.000	1.251
sκl	1.110	1.130	1.205	1.000	1.340
sNh+	1.245	1.595	1.578	1.330	1.299
sNh−	0.079	3.356	1.993	2.700	1.342

Table S3: Parameters resulting from each model (M0, M1, M2). Ion condensation models (M1,M2) used a fixed dipole length, $\tilde{p} = 0.66$. The proteins here were lableled with Cy3B/CF660R.

S9 Non-Electrostatic Interactions for Stm and ProTα

We report the values of two-body non-electrostatic interaction strengths, $\omega_{2,ij}$, for Stm and ProTα. They depend on probed residue indices (i, j) as described in main text Section 2.2 and equation 19, 20 and 21. They were obtained by matching the high-salt data point of each inter-residue segment, for each sequence. We include results from the full ionization model (M0), as well as the effective charge counterion condensation model (M1). In all cases, the three-body non-electrostatic interaction strength was fixed at $\omega_3 = 0.1$.

Residue Indices (i, j)	**M0** $\omega_{2,ij}$	**M1** $\omega_{2,ij}$
(36,54)	-0.812	0.322
(54,109)	-0.441	0.530
(87,147)	-0.581	0.490
(209,272)	-0.879	0.810
(311,374)	-0.870	0.771
(17,87)	-0.201	0.540
(484,567)	-1.172	1.718
(374,484)	-1.185	0.997

Table S4: Two-body interaction strengths for each segment (pair of residue indices shown in left column) as determined by matching the high-salt data, for Stm. These are parameter values for both models (M0 and M1) shown in Figure 6 of the main text. In M1, additional parameters are $\tilde{p} = 0.45$, $\delta = 1.30$

Residue Indices (i, j)	**M0** $\omega_{2,ij}$	**M1** $\omega_{2,ij}$
(2,57)	-1.987	0.935
(58,112)	-5.425	1.433
(2,111)	-3.159	1.741

Table S5: Two-body interaction strengths for each segment (pair of residue indices shown in left column) as determined by matching the high-salt data, for ProTα N, C, NC from top to bottom (see Table S1). These are parameter values for both models (M0 and M1) shown in Figure 6 of the main text. In M1, additional parameters are $\tilde{p} = 0.45$, $\delta = 1.10$

S10 Additional plots for linker IDRs

The figure below shows predicted degrees of ionization from models M1 and M2 using parameters inferred by fitting data collected using Alexa dyes.

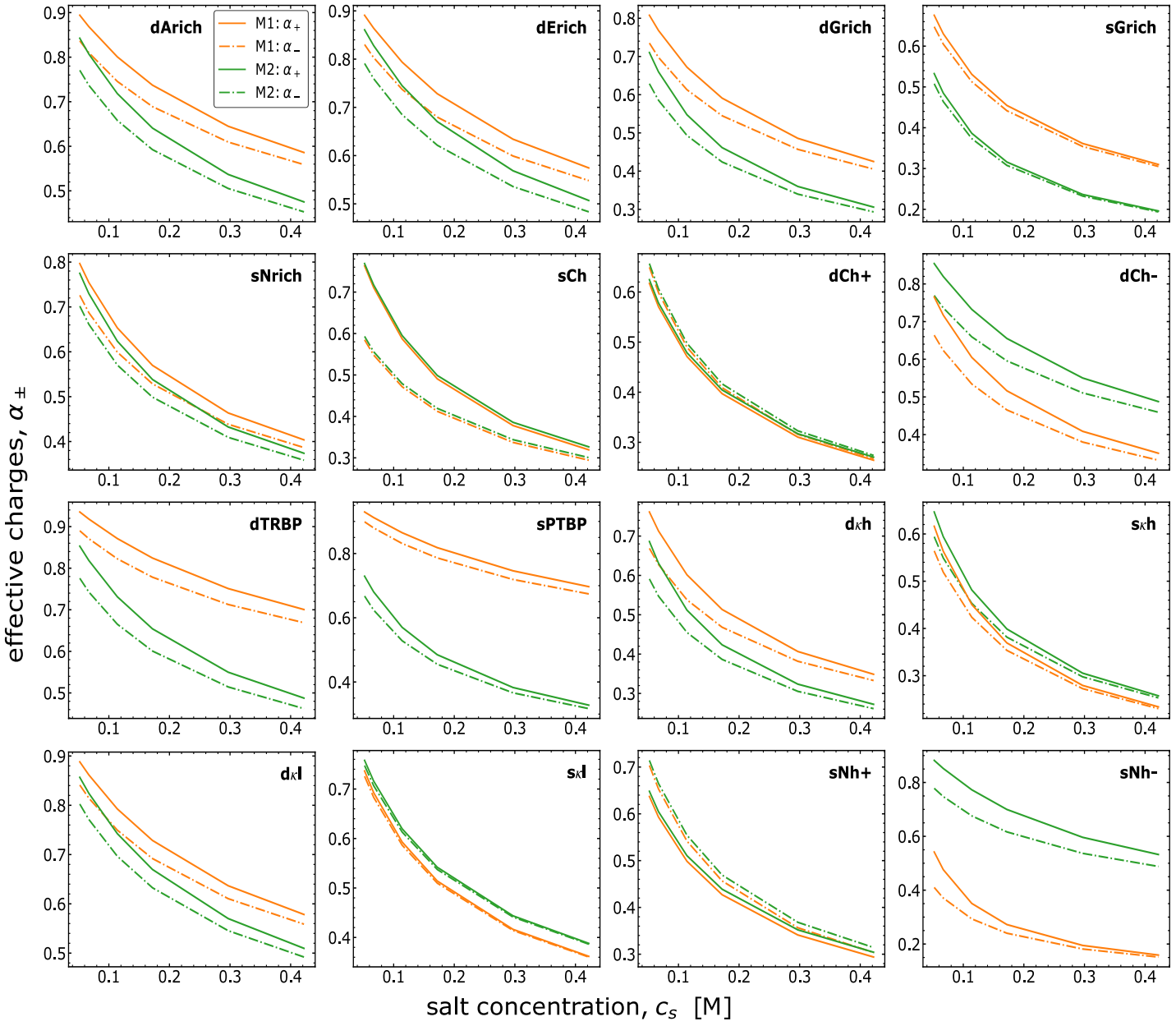

Figure S3: Effective charge (degrees of ionization) as a function of salt concentration (solid lines for positive charge, dashed lines for negative charge) for linker sequences predicted using parameters (of model M1, and near-predictive model M2) obtained by fitting data collected using Alexa 488/594 fluorophores. Charge can vary across a wide range, for each sequence and across sequences, but in all cases, ionization decreases as salt is increased because more ions are made available for condensation on charged groups in the IDR. The differences between the two models (M1 and M2), as seen for some sequences, are due to differences in the δ values (see Table S2)

The figure below shows the differences between fitted R_{ee} for all models (H-J, M0, M1, M2) and R_{ee} from the data collected using Alexa dyes.

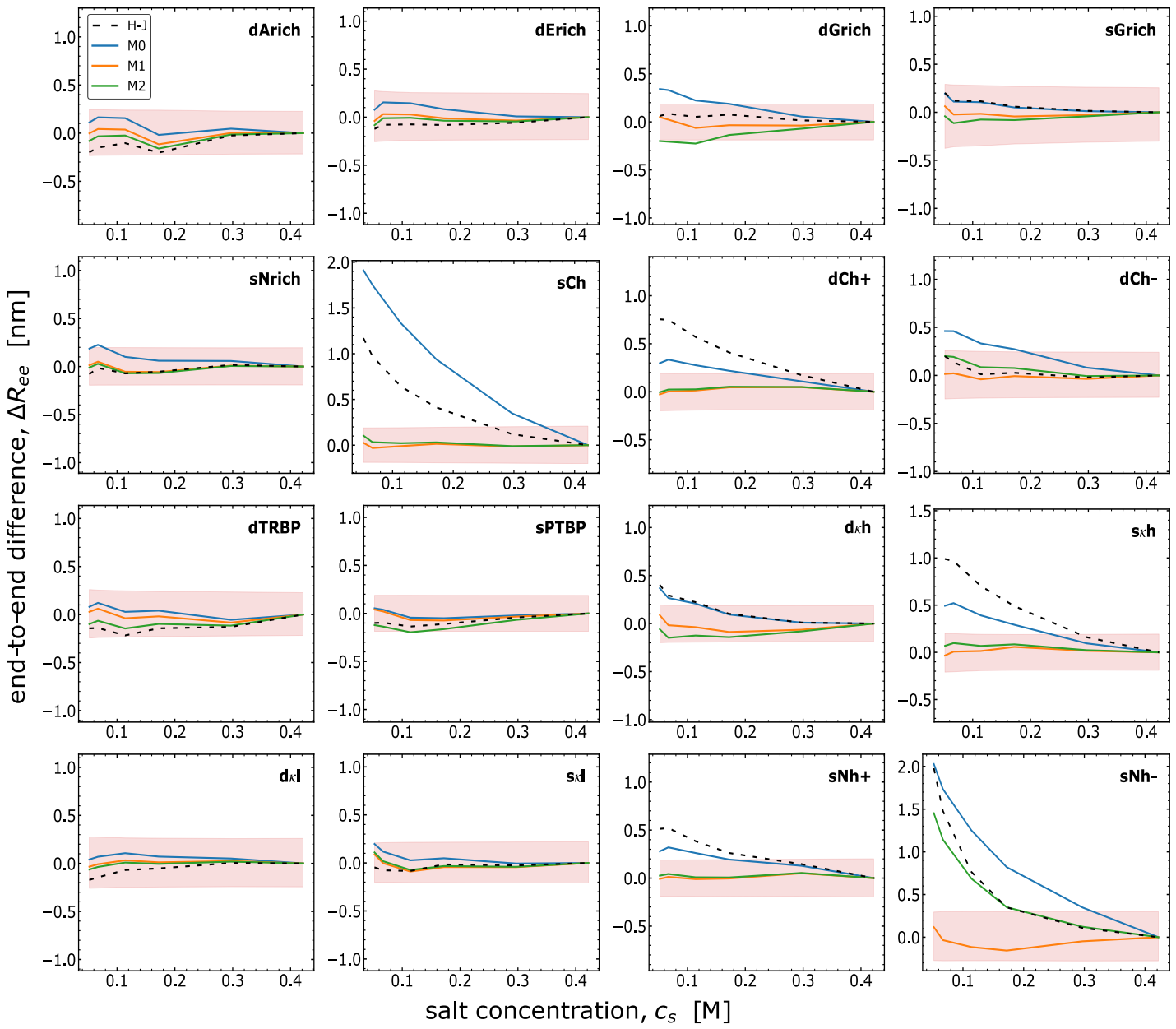

Figure S4: Differences between the end-to-end distances for linker sequences from all four models (H-J, M0, M1, M2) and end-to-end distances from the measurements show that counterion models (M1, M2) outperform simpler models (H-J, M0). These results are for Alexa 488/594 fluorophores. The shaded region indicates the systematic uncertainty in the single-molecule FRET measurements.

The figure below shows the mean errors (chi-squared), for each of the 16 linker IDRs labeled with the Alexa dyes, and the overall mean across all sequences.

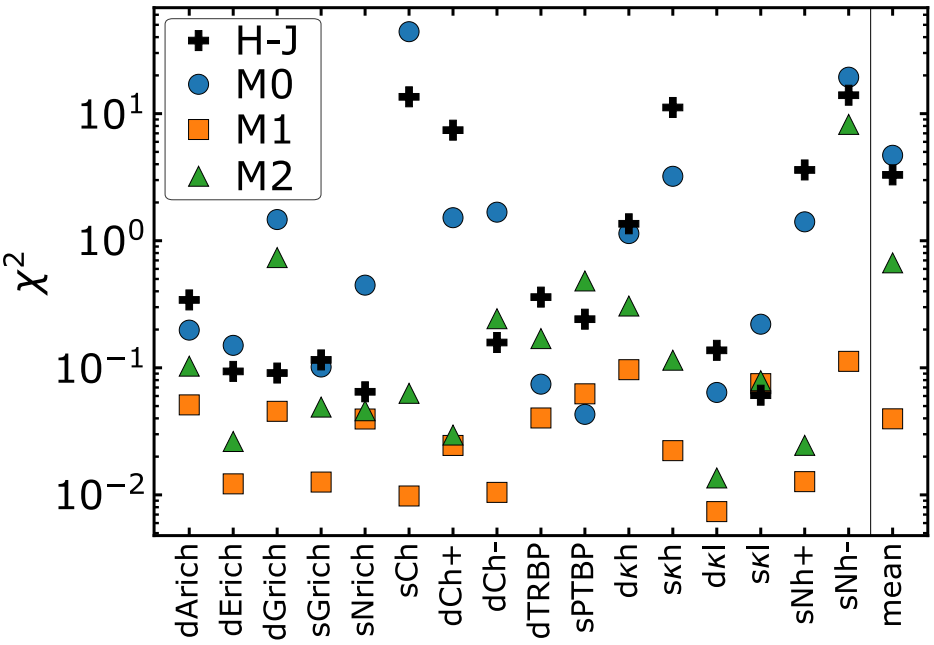

Figure S5: Chi-squared errors compared for the four models (H-J, M0, M1, M2) for each IDR with the Alexa dyes, with the overall mean across all sequences in the last entry, showing that counterion models (M1, M2) perform better than the simpler models (H-J, M0). A log scale is used to highlight the differences between models.

Here we compare the different components of the Free Energy contributions for two select sequences using parameters of model M1 inferred from data collected with the Alexa 488/594 fluorophores. For both sequences, fluctuation and entropic contributions ($F_1 + F_2 + F_3$) increase with salt as expected, while free energy of ion pair formation (F_4) becomes more attractive due to the enhanced propensity to condense with salt. However, contrasting trends between dCh- and sκh are seen when the salt dependence of individual terms of the chain free energy (F_5) are computed. For polyelectrolyte-like sequences such as dCh-, charge-charge interaction is repulsive and decreases with increasing salt concentration due to screening, while sκh exhibits polyampholytic features with charge-charge attraction that decreases (becomes less negative) with increasing salt concentration. The total chain free energy (F_5) for dCh- is monotonically decreasing with salt, primarily due to screening of charge-charge interactions, which dominate over dipolar interactions, while sκh shows the opposite trend, since both charge-charge and charge-dipole interactions are attractive and get progressively screened with increasing salt concentration.

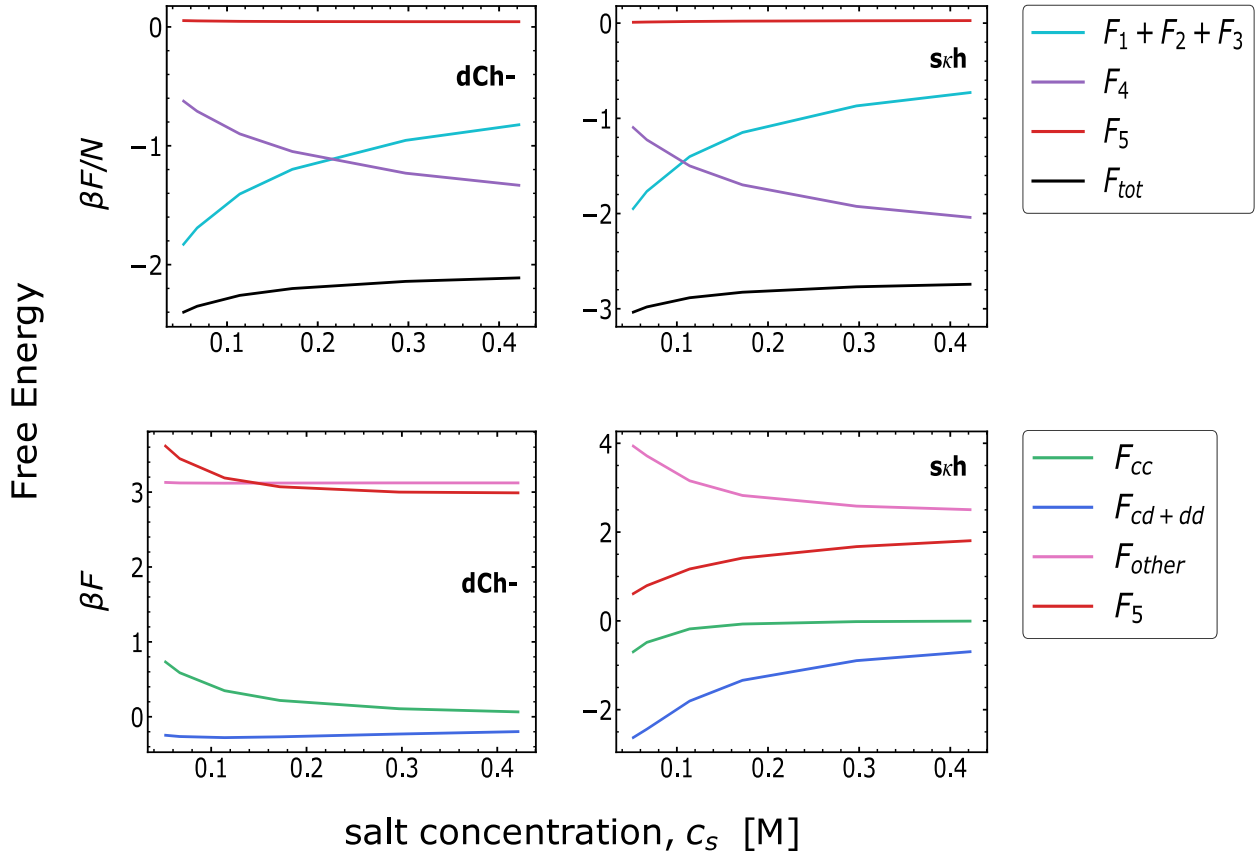

Figure S6: Free Energy contributions are compared as a function of salt concentration under model M1, for two select linker IDRs. The choice of sequences is based on composition: dCh− (left) is polyelectrolyte-like, while sκh (right) is polyampholyte-like. The top row shows different contributions to the overall Free Energy terms: entropy and fluctuation contributions together ($F_1 + F_2 + F_3$), ion pair formation (F_4), chain conformations and electrostatics (F_5), and the total (F_{tot}). These are evaluated in dimensionless units per monomer ($\beta F/N$). The bottom row shows the different contributions of F_5: charge-charge electrostatics (F_{cc}), charge-dipole and dipole-dipole electrostatics (F_{cd+dd}), remaining contributions including non-electrostatic interactions (F_{other}), and total F_5. These are also dimensionless but for the entire chain (βF). Note that entropic and fluctuation terms F_2 and F_3 shown here are relative to the zero ionization case, i.e. $F_2 \to F_2(\alpha_\pm, c_s) - F_2(\alpha_\pm = 0, c_s)$ [and similar for F_3], because the salt contributions dominate; subtracting that part exposes the chain contributions.

The figure below is a consistency check to ensure that the M3 model (salt-dependent dielectric constant) compares well against the data collected using the Alexa dyes. However, this model is not of practical use since it requires knowing the value of the chain dimensions a priori.

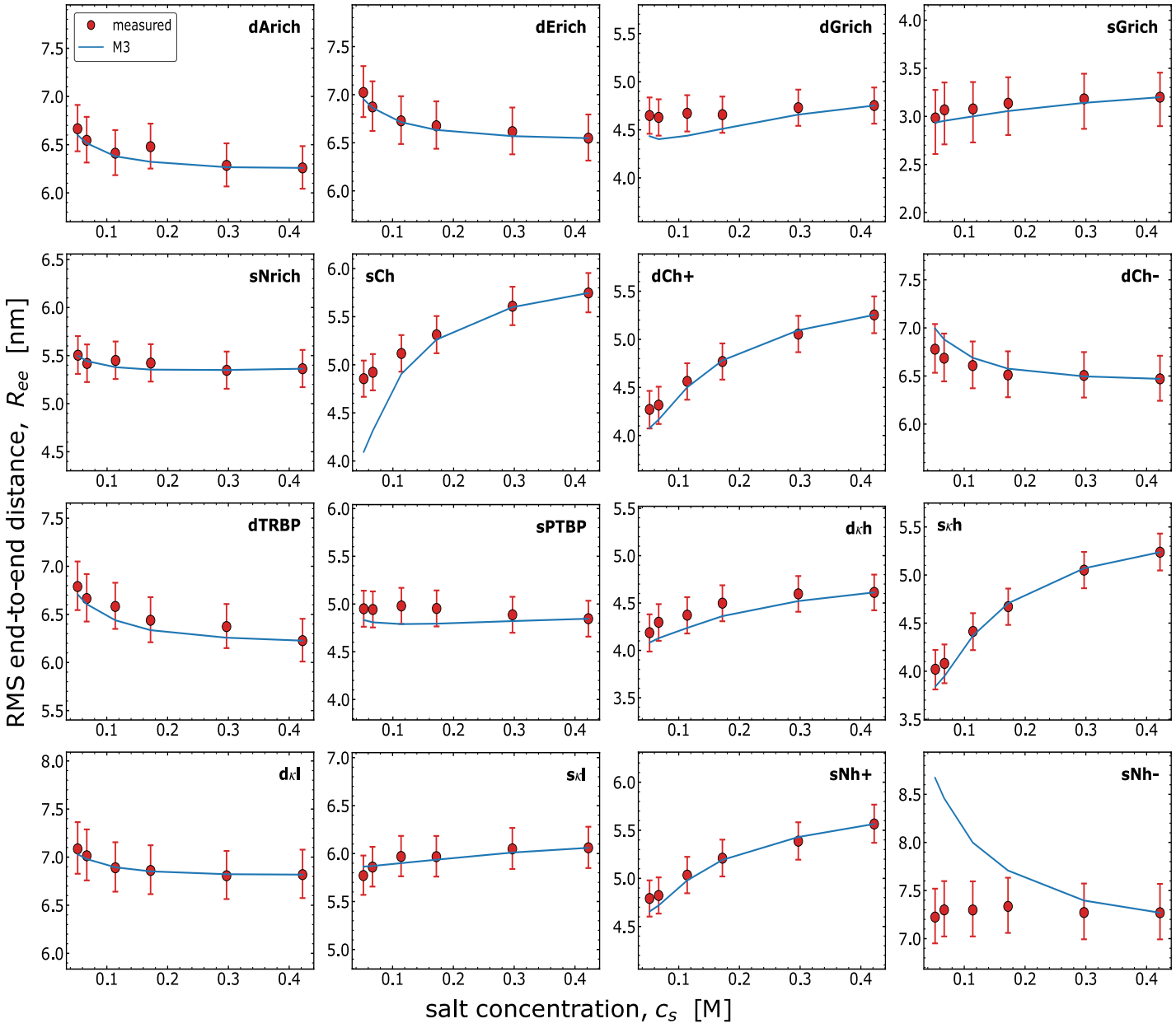

Figure S7: End-to-end distances as a function of salt concentration for all IDRs sequences labeled with Alexa 488/594, with the salt-dependent δ model (M3), based upon the linear fit (main text equation 24), compare well with FRET measurements. Model M3 performs slightly worse than M2, with χ^2 (excluding sNh−) of 0.513 for M3 and 0.166 for M2 (see main text Figure 2 and Table 1).

The figure below shows the predicted degrees of ionization from models M1 and M2 using parameters inferred by fitting the linker IDR data collected using the Cy3B/CF660R dye pair.

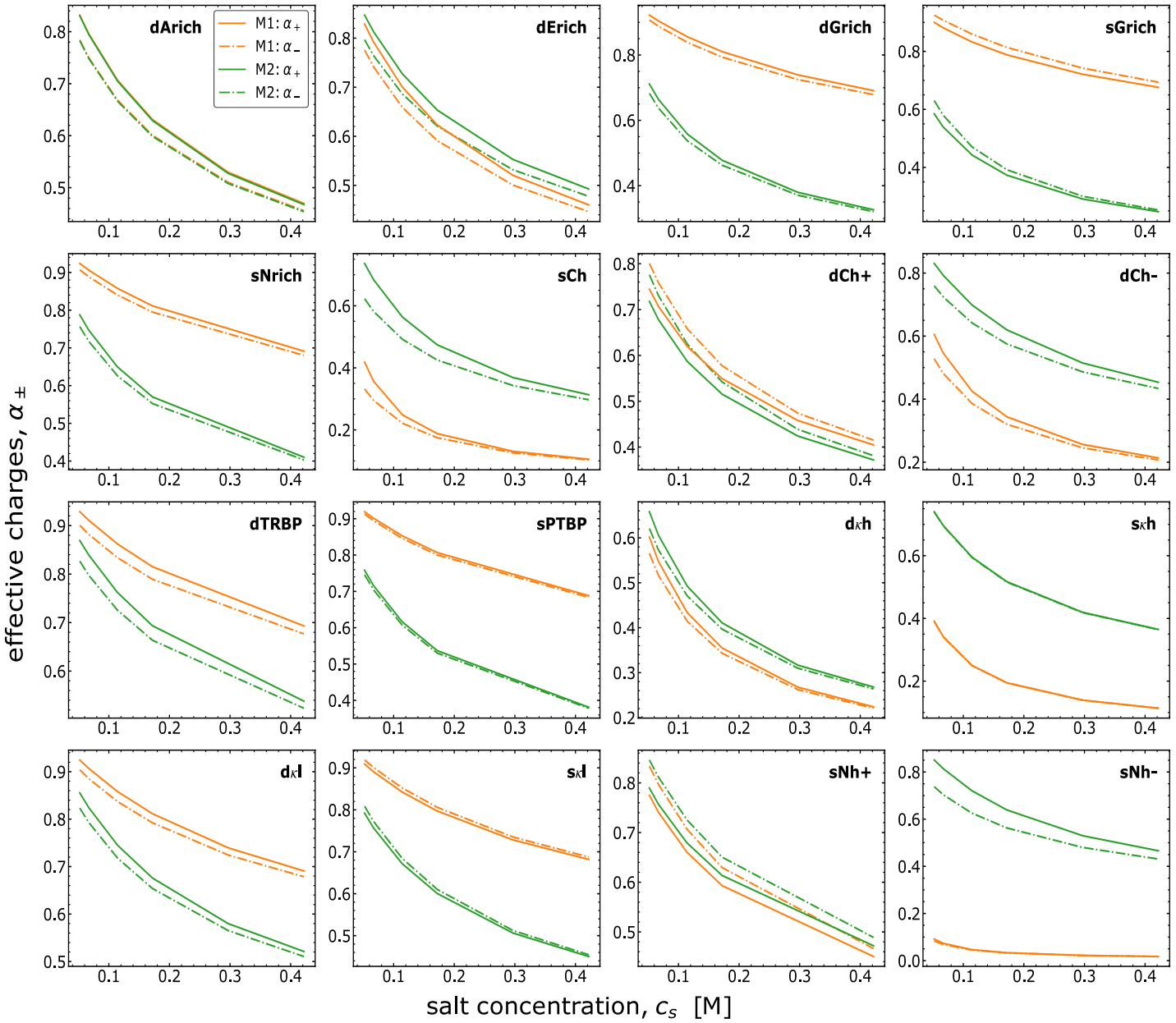

Figure S8: Effective charge (degree of ionization) as a function of salt concentration (solid lines for positive charge, dashed lines for negative charge) for linker IDRs predicted using parameters (of model M1, and near-predictive model M2) obtained by fitting data collected using the Cy3B/CF660R dye pair. As with the Alexa dyes (Figure S3), charge can vary across a wide range, for each sequence and across sequences, but in all cases, ionization decreases as salt is increased because more ions are made available for condensation on the charged groups of the IDR. The differences between two models (M1 and M2), as seen for some sequences, are due to differences in the δ values (see Table S3).

The figure below shows the differences between the fitted R_{ee} for all models (H-J, M0, M1, M2) and R_{ee} from linker IDR data collected using Cy3B/CF660R dye pair.

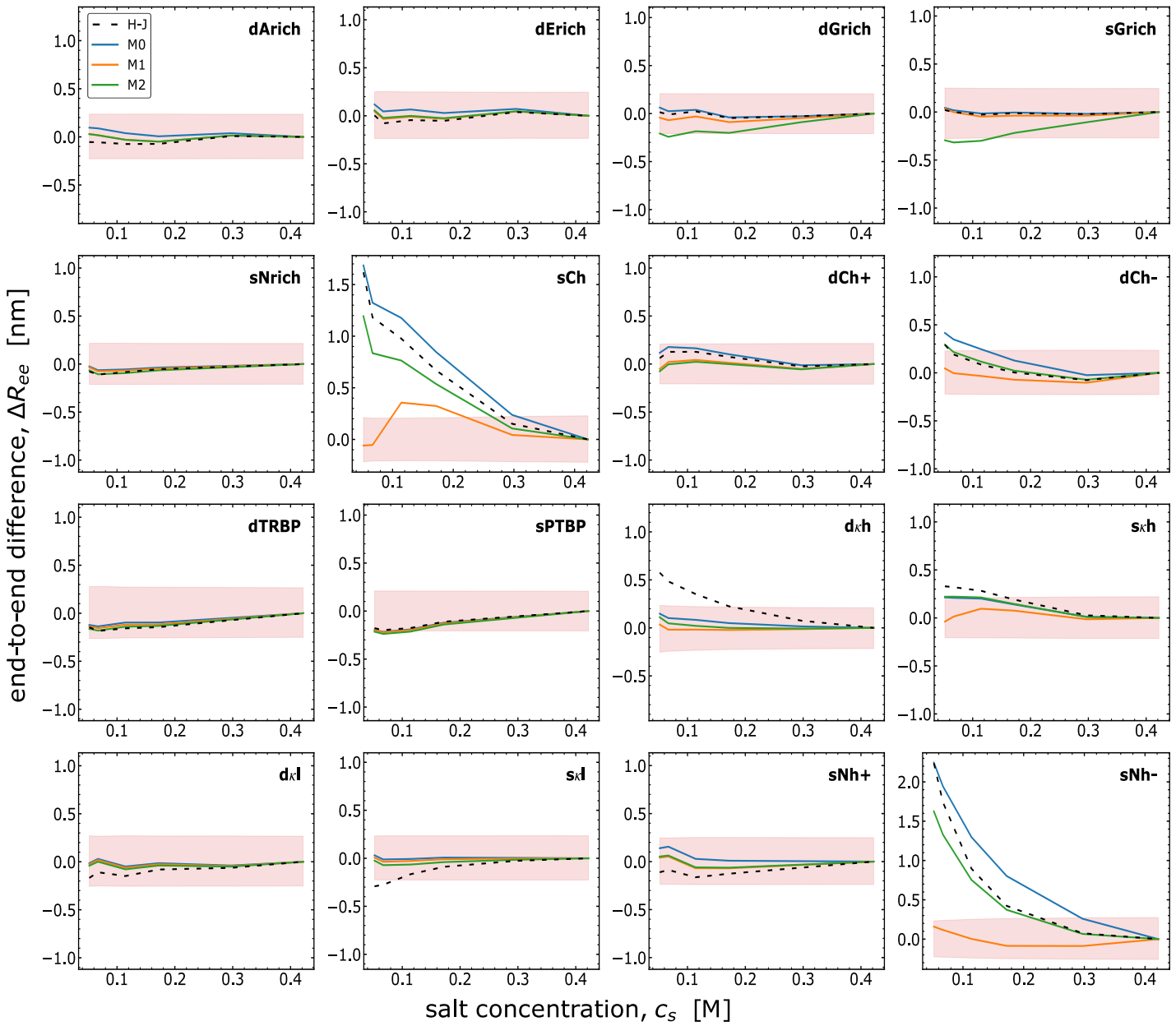

Figure S9: Differences between the end-to-end distances for linker IDRs from all four models (H-J, M0, M1, M2) and end-to-end distances from the measurements show that counterion models (M1, M2) outperform simpler models (H-J, M0). These results are for Cy3B/CF660R fluorophores. The shaded region indicates the span of error bars.

The figure below shows mean errors (chi-square), for each of the 16 IDRs with Cy3B/CF660R dyes, and the overall mean across sequences.

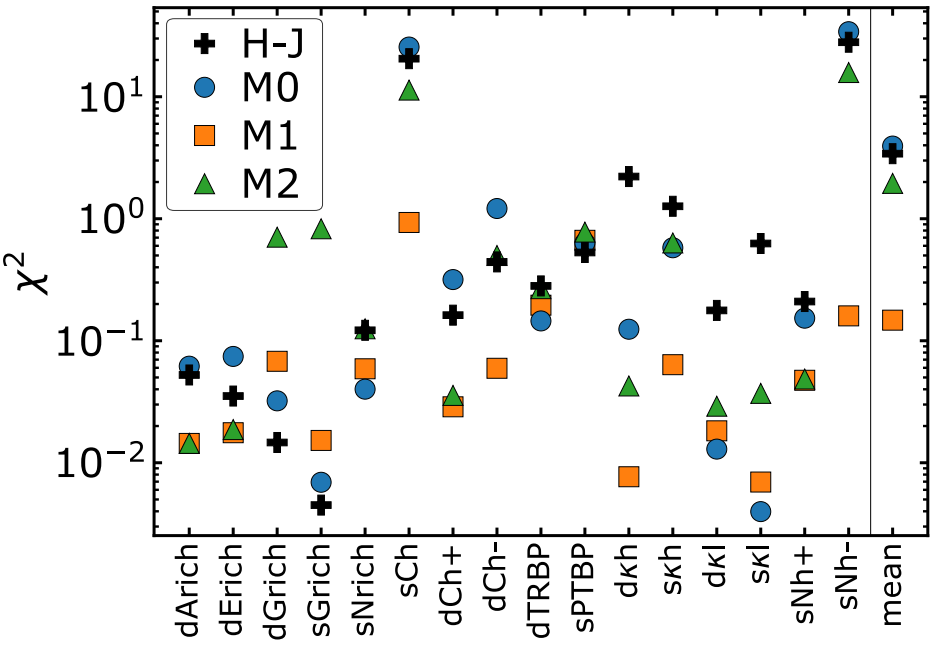

Figure S10: Chi-square errors visually compare the four models H-J, M0, M1, M2 for each IDR with Cy3B/CF660R dyes, with the overall mean across all sequences in the last entry, showing clearly that counterion models (M1, M2) perform better than simpler models (H-J, M0) for most sequences. A log scale is used to highlight the differences between models.

The figure below is a consistency check to ensure that the M3 model (salt-dependent dielectric constant) compares well with data on the linker IDRs collected using the Cy3B/CF660R dye pair.

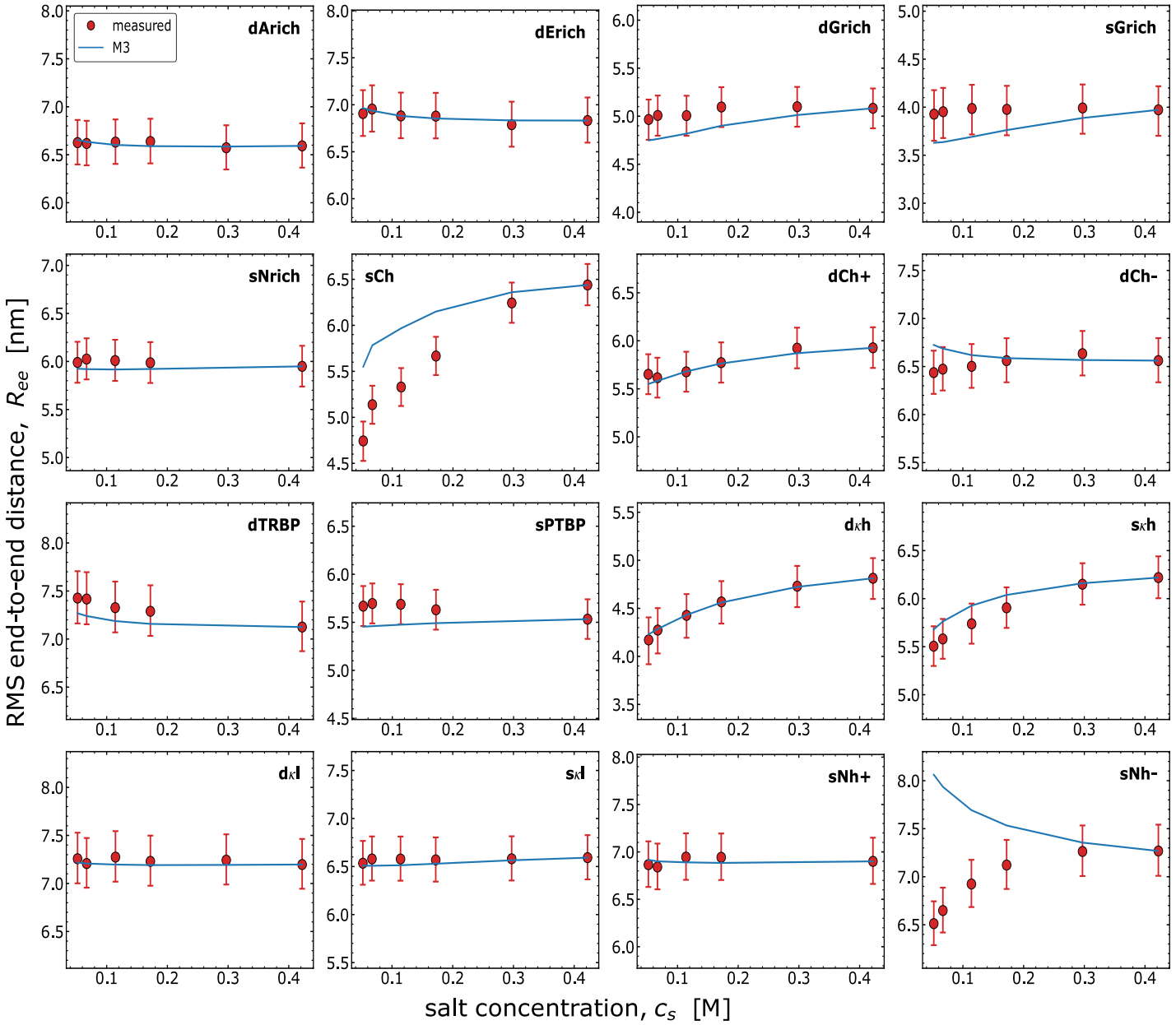

Figure S11: End-to-end distances as a function of salt concentration for all linker IDRs labeled with Cy3B/CF660R, under the salt-dependent δ model (M3), based upon the linear fit (main text equation 24), compare well with FRET measurements. Model M3 performs slightly better than M2, with χ^2 (excluding sNh−) of 0.691 and 1.026, respectively (see main text Figure 4 and Table 2).

Normalized distance maps d^*_{ij} predicted from model M1 and M0 using parameters inferred by fitting data collected for the linker IDRs using the Cy3B/CF660R dye pair.

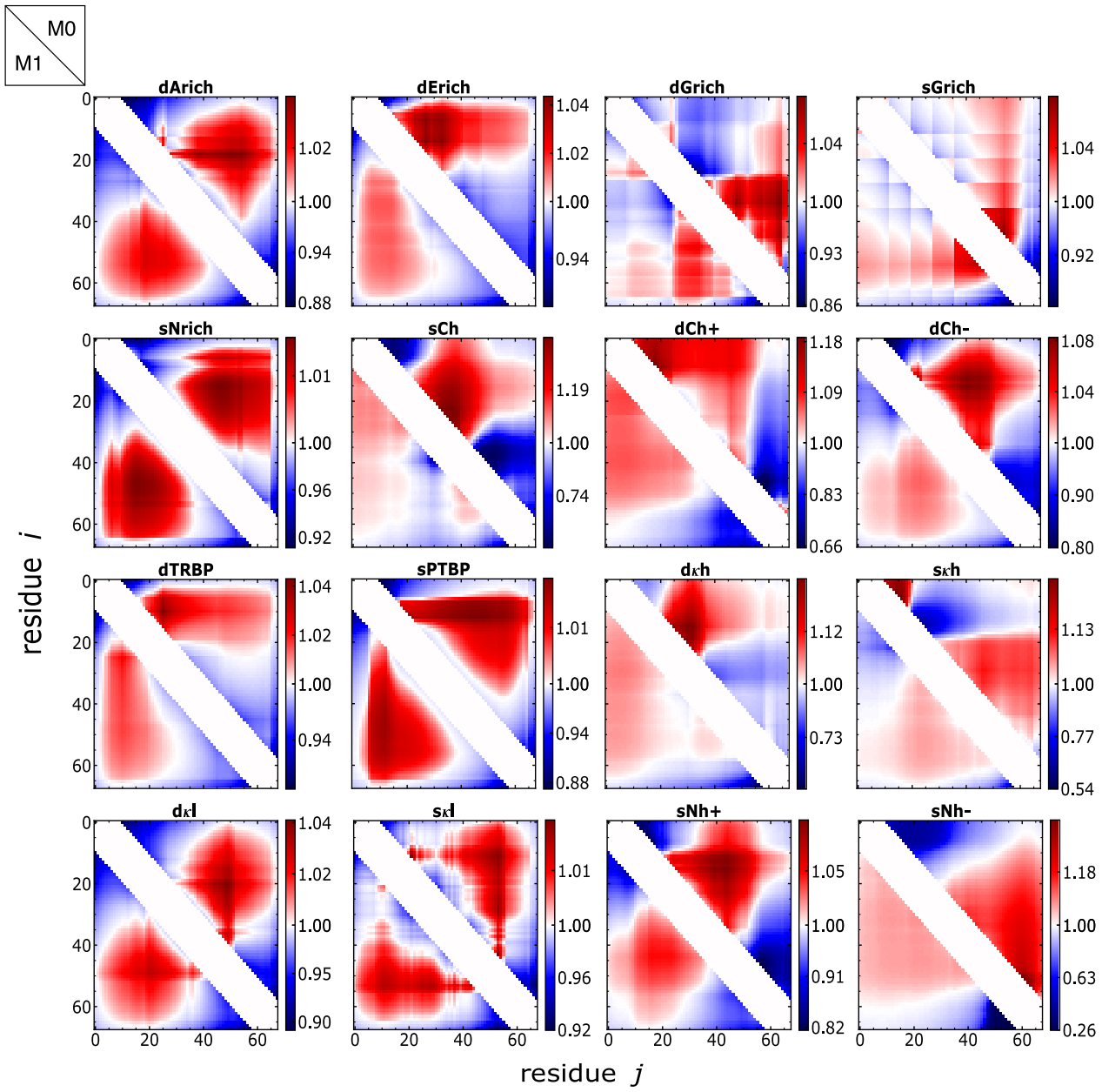

Figure S12: Inter-residue distance maps (normalized by fits to homopolymer scaling) for the linker IDRs under near-physiological conditions (150 mM salt and $T = 20\,^{\circ}\mathrm{C}$), predicted using parameters (determined by matching end-to-end distance data collected using the Cy3B/CF660R dye pair) of model M1 (bottom triangle), and model M0 (upper triangle).

Normalized distance maps d^*_{ij} for the linker IDRs predicted from model M1 compared between the dye pairs Alexa488/594 and Cy3B/CF660R.

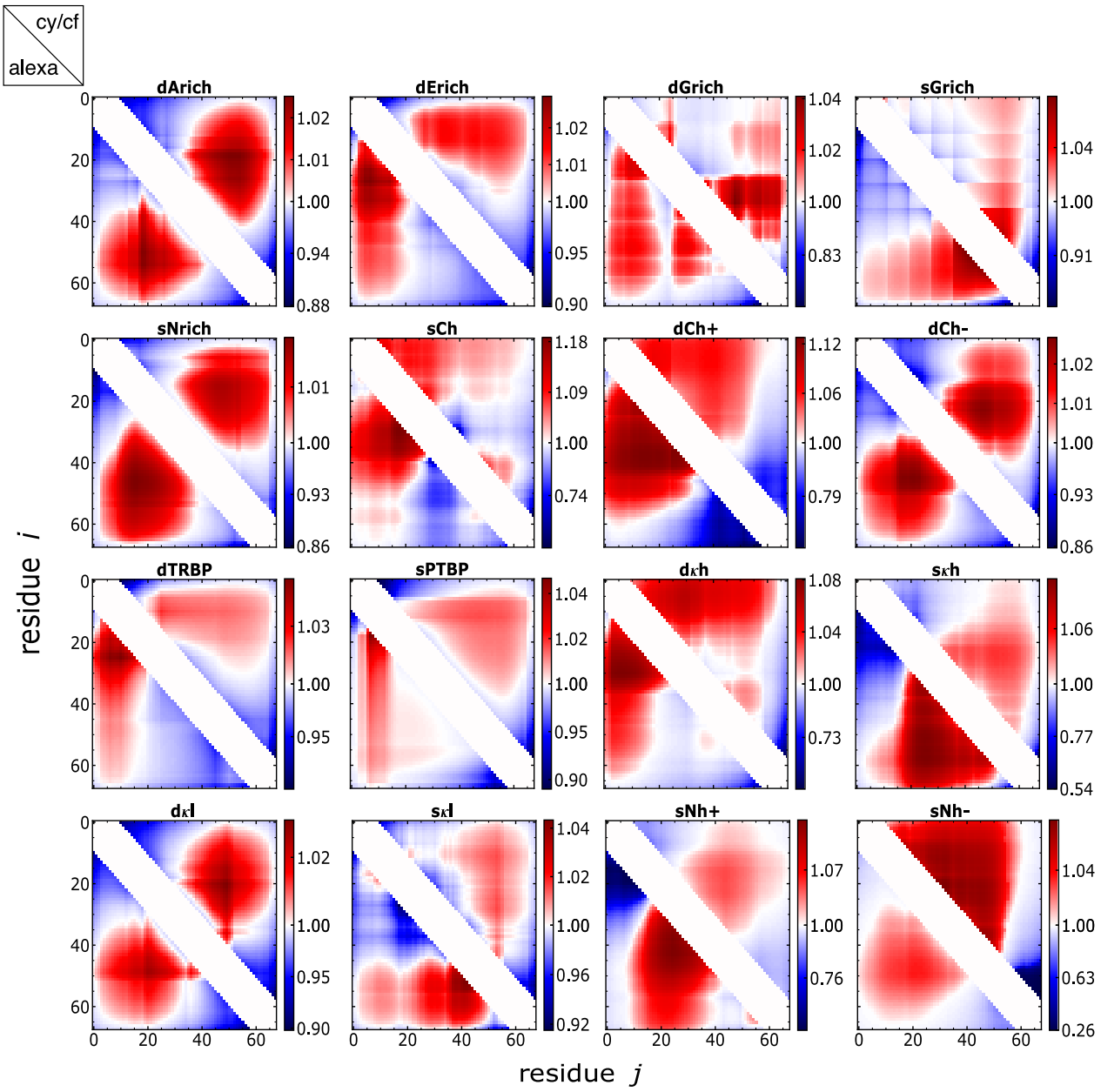

Figure S13: Inter-residue distance maps (normalized by fits to homopolymer scaling) for the linker IDRs under near-physiological conditions (150 mM salt and $T = 20\,^{\circ}\mathrm{C}$), predicted using parameters of model M1 (see Tables S2 and S3) for Alexa488/594 (bottom triangle), and Cy3B/CF660R (upper triangle).

The figure below displays predicted effective charges (degrees of ionization) associated with the prediction of distance between a specific residue pair (segments) of Stm and ProTα.

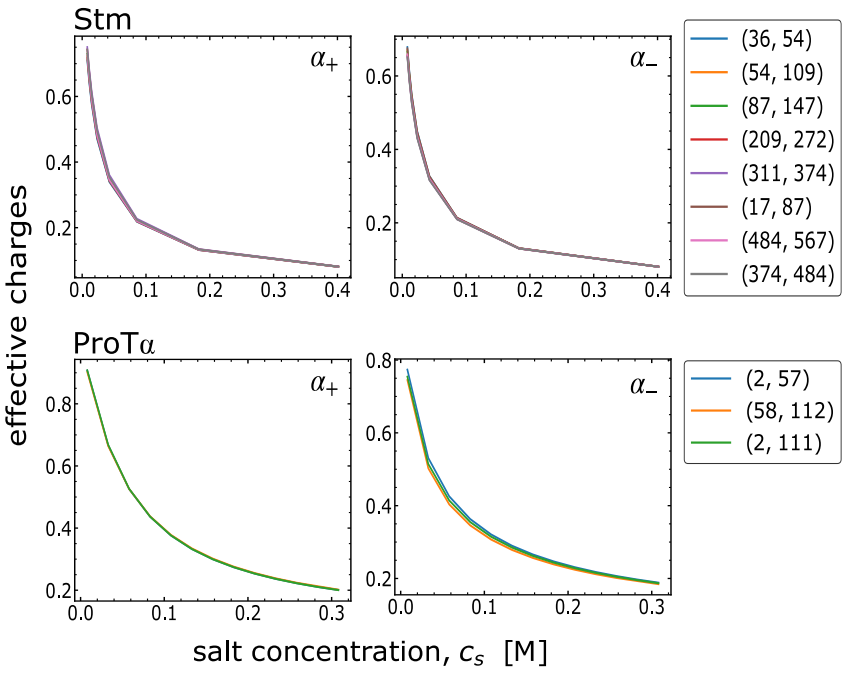

Figure S14: Predicted effective charges (degrees of ionization), for positive and negative charges (α_+, α_-), when free energy is minimized to determine the distance between a specific residue pair (segments) for Stm and ProTα as a function of salt concentration (c_s). These values vary little between different residue pairs. Legends are labeled by residue pairs, (i, j), for each sequence, although curves are difficult to distinguish due to their similarity. These results are not enforced by construction, but are emergent from the model, indicating that the counter-ion theory predicts highly similar results for charges across pairs/segments, even if x_{ij} (and R_{ij}) are distinct between pairs/segments. These results correspond to main text Figure 6, and were obtained self-consistently in the same calculation.

The figure below shows the comparison between inter-residue distances R_{ij} for the linker IDRs predicted from M2 model and the $C_\alpha - C_\alpha$ distance, between the pair of residues $(i, j) = (5, 63)$, measured using the Alexa dye pair.

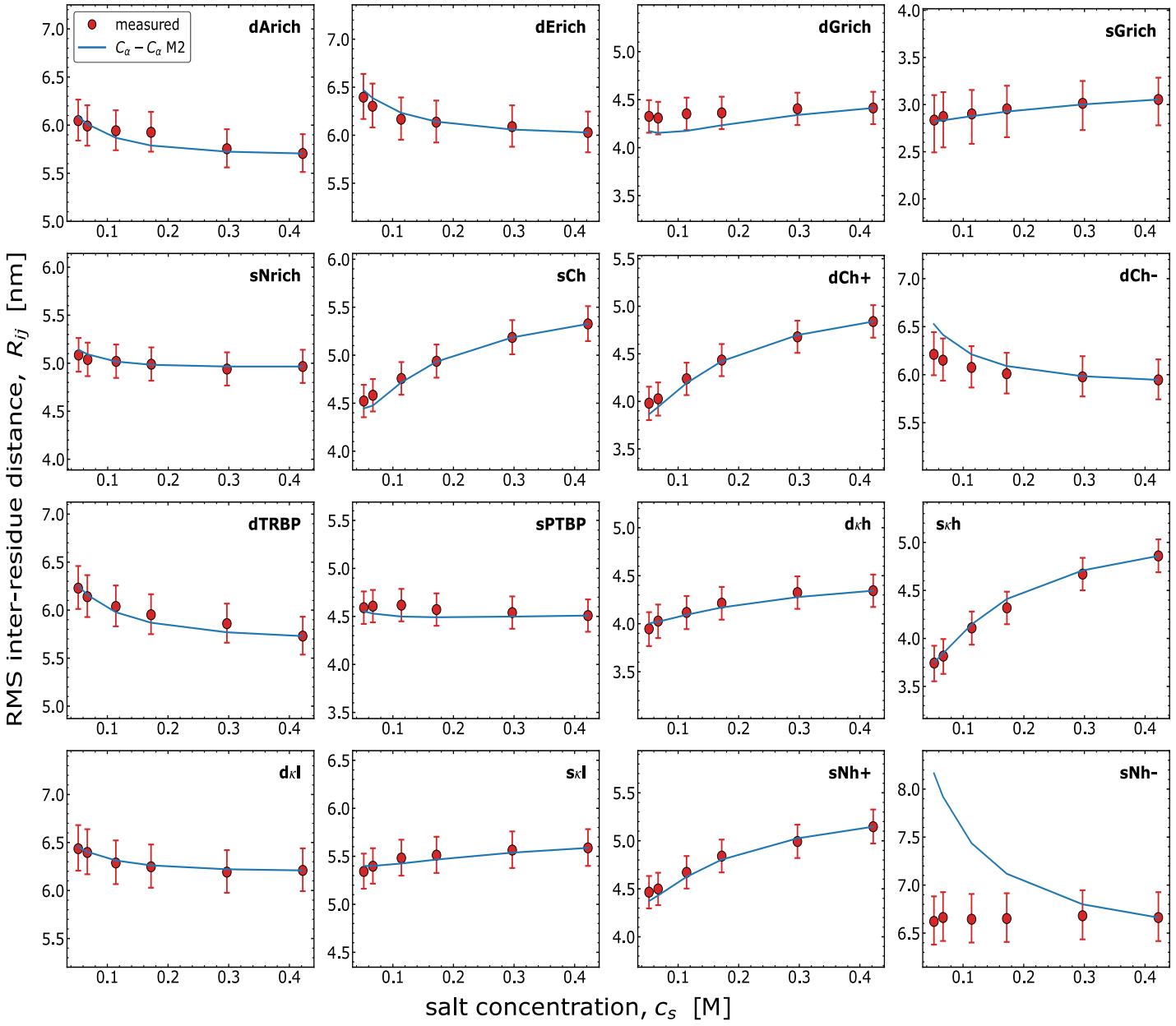

Figure S15: Inter-residue distances $(C_\alpha - C_\alpha)$ as a function of salt concentration for all linker IDRs labeled with Alexa 488/594, modeled using M2, compare well with FRET measurements. This differs from results in main text Figure 2 in three ways: $C_\alpha - C_\alpha$ distances were inferred from FRET efficiencies, dyes were modeled differently (see Sec. S3), and the underlying theory for R_{ij} with residue pair $(i, j) = (5, 63)$ was used (see main text Sec. 2.2 and 3.5). $C_\alpha - C_\alpha$ based on M2 performs similarly to end-to-end M2, with χ^2 (excluding sNh−) of 0.147 and 0.166, respectively (see main text Figure 2 and Table 1).

The figure below shows the comparison between inter-residue distances R_{ij} for the linker IDRs predicted from the M2 model and the $C_\alpha - C_\alpha$ distance, between the pair of residues $(i, j) = (5, 63)$, measured using the Cy3B/CF660R dye pair.

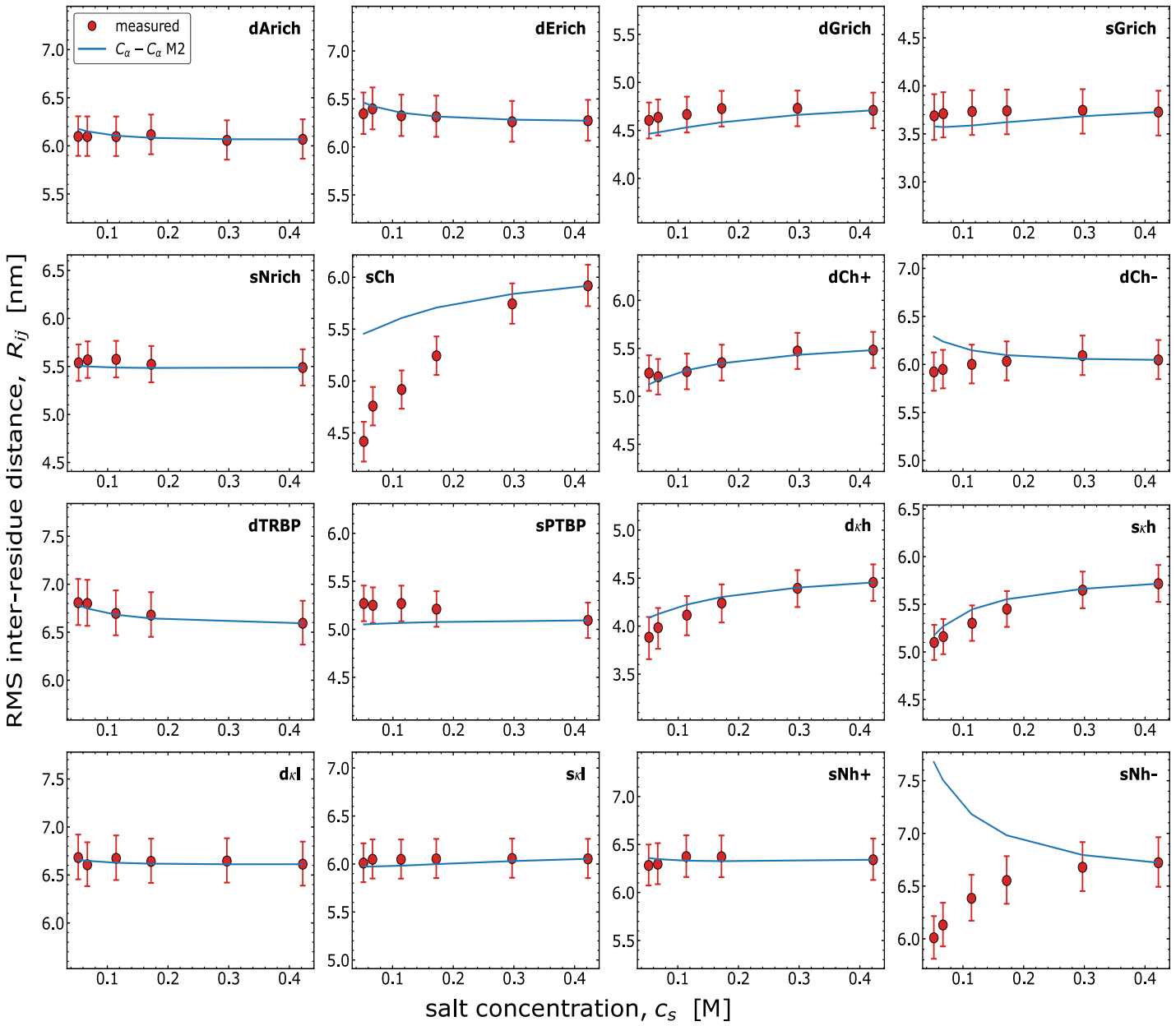

Figure S16: Inter-residue distances ($C_\alpha - C_\alpha$) as a function of salt concentration for all linker IDRs labeled with Cy3B/CF660R, modeled using M2, compare well with FRET measurements. This differs from results in the main text Figure 4 in three ways: $C_\alpha - C_\alpha$ distances were inferred from FRET efficiencies, dyes were modeled differently (see Sec. S3), and the underlying theory for R_{ij} with residue pair $(i, j) = (5, 63)$ was used (see main text Sec. 2.2 and 3.5). $C_\alpha - C_\alpha$ based on M2 performs similarly to end-to-end M2, with χ^2 (excluding sNh−) of 0.956 and 1.026, respectively (see main text Figure 4 and Table 2).

References

[1] Andrea Holla, Erik W. Martin, Thomas Dannenhoffer-Lafage, Kiersten M. Ruff, Sebastian L. B. König, Mark F. Nüesch, Aritra Chowdhury, John M. Louis, Andrea Soranno, Daniel Nettels, Rohit V. Pappu, Robert B. Best, Tanja Mittag, and Benjamin Schuler. Identifying sequence effects on chain dimensions

of disordered proteins by integrating experiments and simulations. *JACS Au*, 4(12):4729–4743, 2024. doi: 10.1021/jacsau.4c00673. URL https://doi.org/10.1021/jacsau.4c00673.

[2] J. K. Ko and J. Ma. A rapid and efficient pcr-based mutagenesis method applicable to cell physiology study. *Am. J. Physiol. Cell Physiol.*, 288(6):C1273–8, 2005. ISSN 0363-6143 (Print) 0363-6143 (Linking). doi: 10.1152/ajpcell.00517.2004. URL https://www.ncbi.nlm.nih.gov/pubmed/15659713.

[3] Tomasz M. Kapłon, Grzegorz Rymarczyk, Małgorzata Nocula-Ługowska, Michał Jakób, Marian Kochman, Marek Lisowski, Zbigniew Szewczuk, and Andrzej Ożyhar. Starmaker exhibits properties of an intrinsically disordered protein. *Biomacromolecules*, 9(8):2118–2125, 2008. doi: 10.1021/bm800135m. URL https://doi.org/10.1021/bm800135m.

[4] A Borgia, M B Borgia, K Bugge, V M Kissling, P O Heidarsson, C B Fernandes, A Sottini, A Soranno, K J Beholder, D Nettles, B B Kragelund, R B Best, and B Schuler. Extreme disordered in an ultrahigh-affinity protein complex. *Nature*, 555:61–66, 2018.

[5] A Chowdhury, A Borgia, S Ghosh, A Sottini, S Mitra, R.S Eapan, MB Borgia, T Yang, N Galvanetto, M Ivanovic, P Lukijanczuk, R Zhu, D Nettels, A Kundagrami, and B Schuler. Driving forces of the complex formation between highly charged disordered proteins. *Proc. Natl. Acad. Sci.*, 120(41): e2304036120, 2023.

[6] Volodymyr Kudryavtsev, Martin Sikor, Stanislav Kalinin, Dejana Mokranjac, Claus A. M. Seidel, and Don C. Lamb. Combining MFD and PIE for accurate single-pair förster resonance energy transfer measurements. *ChemPhysChem*, 13(4):1060–1078, 2012. doi: 10.1002/cphc.201100822. URL https://doi.org/10.1002/cphc.201100822.

[7] B. Schuler. Application of single molecule förster resonance energy transfer to protein folding. *Methods Mol. Biol.*, 350:115–38, 2007. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16957321.

[8] Erik D. Holmstrom, Andrea Holla, Wenwei Zheng, Daniel Nettels, Robert B. Best, and Benjamin Schuler. *Accurate Transfer Efficiencies, Distance Distributions, and Ensembles of Unfolded and Intrinsically Disordered Proteins From Single-Molecule FRET*, pages 287–325. Elsevier, 2018. doi: 10.1016/bs.mie.2018.09.030. URL https://doi.org/10.1016/bs.mie.2018.09.030.

[9] Nam Ki Lee, Achillefs N. Kapanidis, You Wang, Xavier Michalet, Jayanta Mukhopadhyay, Richard H. Ebright, and Shimon Weiss. Accurate FRET measurements within single diffusing biomolecules using alternating-laser excitation. *Biophysical Journal*, 88(4):2939–2953, 2005. doi: 10.1529/biophysj.104.054114. URL https://doi.org/10.1529/biophysj.104.054114.

[10] Matthew Antonik, Suren Felekyan, Alexander Gaiduk, and Claus A. M. Seidel. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *The Journal of Physical Chemistry B*, 110(13):6970–6978, 2006. doi: 10.1021/jp057257. URL https://doi.org/10.1021/jp057257.

[11] Patrick Ernst, Franziska Zosel, Christian Reichen, Daniel Nettels, Benjamin Schuler, and Andreas Plückthun. Structure-guided design of a peptide lock for modular peptide binders. *ACS Chemical Biology*, 15(2):457–468, 2020. doi: 10.1021/acschembio.9b00928. URL https://doi.org/10.1021/acschembio.9b00928.

[12] W Zheng, G H Zerze, A Borgia, J Mittal, B Schuler, and R B Best. Inferring properties of disordered chains from fret transfer efficiencies. *Journal of Chemical Physics*, 148:123329, 2018.

[13] H Hofmann, A Soranno, A Borgia, K Gast, D Nettels, and B Schuler. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci.*, 109(40):16155–16160, 2012.

[14] PG Higgs and JF Joanny. Theory of polyampholyte solutions. *J. Chem. Phys.*, 94:1543–1554, 1991.

[15] E W Martin, A S Holehouse, C R Grace, A Hughes, R V Pappu, and T Mittag. Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation. *J. Am. Chem. Soc.*, 138:15323–15335, 2016.

[16] J Huihui and K Ghosh. An analytical theory to describe sequence-specific inter-residue distance profiles for polyampholytes and intrinsically disordered proteins. *J. Chem. Phys.*, 52:161102, 2020.

[17] NIST computational chemistry comparison and benchmark database. `http://cccbdb.nist.gov/`, 2022. URL `http://cccbdb.nist.gov/`.

[18] M Phillips, M Muthukumar, and K Ghosh. Beyond monopole electrostatics in regulating conformations of intrinsically disordered proteins. *PNAS Nexus*, 3(9):367, 2024.

[19] T J Nott, E Petsalaki, P Farber, D Jervis, E Fussner, A Plochowietz, T D Craggs, D P Bazett-Jones, T Pawson, J D Forman-Kay, and AJ Baldwin. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular Cell*, 57(5):936–947, 2015.

[20] T Firman and K Ghosh. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *J. Chem. Phys.*, 148(12):123305, 2018.

[21] L Houston, M Phillips, A Torres, K Gaalswyk, and K Ghosh. Physics-based machine learning trains hamiltonians and decodes sequence-conformation relation in the disordered proteome. *Journal of Chemical Theory and Computation*, 20(22):10266–10274, 2024.

[22] J Huihui, T Firman, and K Ghosh. Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. *J. Chem. Phys.*, 149:085101, 2018.