Submitted to: PNAS

HYPK: a marginally disordered protein sensitive to charge decoration

Arash Firouzbakht,[1] Austin Haider,[2] Kari Gaalswyk,[3] Sepehr Alaeen,[4] Kingshuk Ghosh[3,*] and Martin Gruebele[1,4,5*]

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, [2]Department of Molecular and Cellular Biophysics, University of Denver, [3]Department of Physics and Astronomy, University of Denver, [4]Center for Biophysics and Quantitative Biology, [5]Department of Physics, Carle-Illinois College of Medicine, Center for Advanced Study, University of Illinois at Urbana-Champaign, USA

[*]Corresponding authors: kingshuk.ghosh@du.edu, mgruebel@illinois.edu

Abstract

Intrinsically Disordered Proteins (IDPs) that lie close to the empirical boundary separating IDPs and folded proteins in Uversky's charge-hydropathy plot may behave as `marginal IDPs' and sensitively switch conformation upon changes in environment (temperature, crowding, charge screening), sequence, or both. In our search for such a marginal IDP we selected HYPK near that boundary as a candidate; PKIα, also near that boundary, has lower secondary structure propensity; and Crk1, just across the boundary on the folded side, has higher secondary structure propensity. We used a qualitative Förster Resonance Energy Transfer (FRET)-based assay together with circular dichroism to simultaneously probe global and local conformation. HYPK shows several unique features indicating marginality: a cooperative transition in end-to-end distance with temperature, like Crk1 and folded proteins, but unlike PKIα; enhanced secondary structure upon crowding, in contrast to Crk1 and PKIα; and a crossover from salt-induced expansion to compaction at high temperature, likely due to a structure-to-disorder transition not seen in Crk1 and PKIα. We then tested HYPK's sensitivity to charge patterning by designing charge-flipped variants including two specific sequences with identical amino acid composition that markedly differ in their predicted size and response to salt. The experimentally observed trends, also including mutants of PKIα, verify the predictions from sequence charge decoration metrics. Marginal proteins like HYPK show features of both folded and disordered proteins that make them sensitive to physico-chemical perturbations and structural control by charge patterning.

Significance statement

Many proteins in eukaryotic cells are intrinsically disordered (IDPs). An interesting subset of these IDPs can be 'marginal' with charge and hydrophobic amino acid content very similar to the folded globular proteins. We find that HYPK, a translation regulator important in Huntington's disease, is marginal. By changing the salt concentration, crowding, or temperature, it can be coaxed to show signs of cooperativity, hallmark feature of many folded proteins. Charge-flipped variants of HYPK reveal just how sensitive size and cooperativity is to the sequence of charges. These findings suggest that pinning down the boundary of marginality requires accounting for charge patterns, not just overall charge, yielding new insights into proteins in the gray zone between unfolded and folded structure.

Keywords: sequence charge decoration; Huntington's disease; compactness;

INTRODUCTION

Intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs) of proteins comprise over a third of the human proteome and serve in a multitude of biological functions including signal transduction, transcription regulation, liquid-liquid phase separation, and chaperoning (1, 2). In addition to their vital cellular functions, aberrant IDP function has been linked to disease, most notably neurodegenerative conditions that are associated with large protein aggregates formed because of interactions mediated by unfolded proteins (3). IDPs and IDRs are also fascinating from a structural standpoint: they vary widely in the range of disordered content under different environmental conditions, from mostly folded but including some flexible linkers, to disordered that fold upon binding (4) enhancing molecular recognition (5), to fully disordered yet functional (6).

There has, therefore, been a concerted effort to better understand the structural ensemble of disordered proteins and describe their structural features in terms of simple parameters such as charge composition, mean hydropathy and charge patterning (7). Uversky first noticed that disordered and folded proteins cluster separately on a plot of fraction of net charge versus mean hydropathy (diagram of states in Figure 1) (8), with disordered proteins having lower mean hydropathy and a higher fraction of charged residues. Following this classification, researchers have become interested in understanding the idiosyncrasies of IDP structure beyond the diagram of states, using theoretical polymer models (9–11), computation (12), experiment (13), or a combination of all three (14). These reports have allowed a sequence-to-ensemble-to-function description of disordered proteins for case studies, and have introduced role of charge (15–17), charge correlation (18) and metrics such as charge patterning (10, 12, 19–21) and aromatic patterning (22) as more accurate predictors of IDP structure and size as well as function.

In this study, we investigate the structural cooperativity and environment- and sequence-sensitivity of “marginally disordered” proteins, ones located near the folded-disordered boundary of the Uversky plot (Figure 1a). We expect these proteins to simultaneously exhibit some aspects of disordered and of folded proteins. We also hypothesized that such proteins could encode a cooperative response (Figure 1b) to environmental perturbations such as temperature, ionic strength, and crowder. An earlier study from the Pielak group has indicated such a possibility for FIgM that gains secondary structure upon crowding, both *in vitro* and *in vivo* (23). Furthermore, we expect biological regulators such as mutations, post-translational modifications, and splicing that modulate sequence parameters such as charge patterning, will cause changes in conformation. In

order of increasing “foldability,” we selected three proteins that could span the gamut from mostly unfolded (PKIα), to structurally ambiguous (HYPK), to a mix of well-folded and short IDR sequences (Crk1) based on the available prior characterization (24–26). We utilized a Förster Resonance Energy Transfer (FRET) assay to qualitatively measure trends in end-to-end compaction or expansion as a function of environmental perturbation (Figure 1b), and we used a sequence-dependent theory of chain dimension (10, 27, 28) to reconcile the trends observed by FRET. We find that all sequences contract by crowding and expand by heating, while HYPK shows a unique cross-over behavior as a function of ionic strength that is correctly explained by theory. Although the three proteins contain varying amount of residual secondary structure, the secondary structure content is invariable in PKIα (always low) and Crk1 (always high) but varies as a function of environmental parameters in HYPK, further hinting at its “marginally disordered” nature. HYPK, despite being disordered, also shows significant cooperativity in its thermal scans like folded proteins.

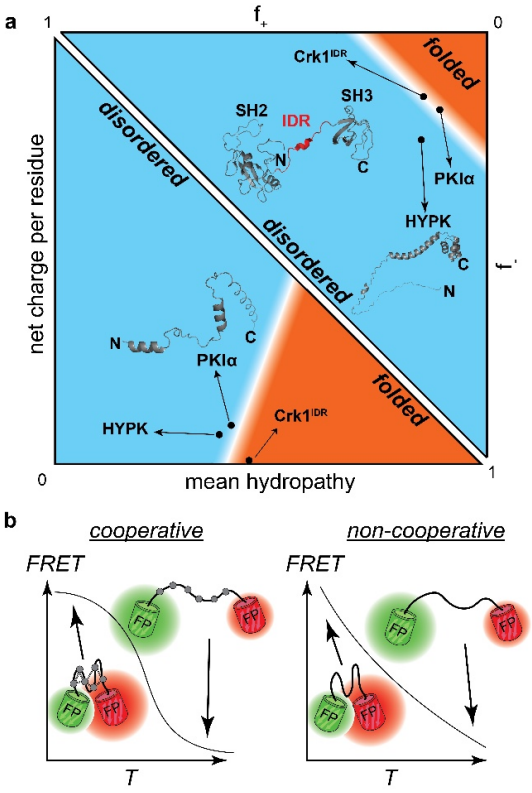

Figure 1. Sequences shown on the diagram of states and a schematic of cooperative and non-cooperative structural change in a FRET assay. (a) Diagram of states of proteins used in this study. The lower left half is the Uversky plot, representing the proteins according to their mean-residue hydropathies and net charge per residues (NCPRs). The upper right half plots the proteins according to their fraction of charged residues. The structures of PKIα and HYPK are predicted by AlphaFold2. For Crk1 we used PDB ID: 2EYY (b) Temperature *T* is an example of environmental parameters that can distinguish a “marginal” IDP’s cooperative behavior, as probed by the FRET assay, from other IDPs.

To further explore the possibility that marginal IDP conformations can be modulated by mutation and test the validity of the theory to account for the effect of charge-patterning on chain dimension, we make several charge-flipped variants of two of the target proteins and show that sequence-dependent theory correctly predicts the relative compaction of the variants as well as the effect of charge-screening mediated by salt on the chain dimension. Designed variants also expand with temperature and collapse upon crowding although with different degree of cooperativity compared to the wild type. These measurements of the wild type and designed mutants provide insights to how physico-chemical and biological regulators can couple and modulate conformations of proteins at the edge of disorder-order boundary, complementing majority of earlier studies of IDPs away from the boundary.

Table 1. Information related to the three IDP targets of the study.

	Protein Kinase Inhibitor α (PKIα)	**Huntingtin-interacting protein K (HYPK)**	**Crk1/Crk1^{IDR}**
Disprot ID	DP00934	DP00546	DP00973
UniProt ID	P61925	Q9NX55	P46108
Length	76	129	204/13
SCD	Rounded to 1	Rounded to 1	n.a.[1] / -0.6
$SCD_{lowsalt}$	Rounded to 7	Rounded to 9	n.a.[1] / -2.1
Q (pH 7)[2]	-7	-9	-4/0
Mean hydropathy[3]	0.42	0.39	0.43/0.43
mean-residue IUPRED score[4]	0.64	0.67	0.42/0.44

[1] n.a. = not applicable to the full Crk1 sequence. [2] In addition, eGFP has a charge of -7 and mCherry of -6 at pH 7.[3] Mean-residue hydropathies were calculated according to Kyte and Doolittle's method, available in Expasy ProtScale (29).[4] The residue-wise IUPRED scores are plotted on each sequence in Figure S1. Q is the net charge of the protein and SCD and $SCD_{lowsalt}$ are two charge patterning metric that is calculated from sequence (see Equation 2 and 3 in the Supplemental Information for more)

RESULTS

Fluorescent protein vs. dye labeling. FRET-labeled proteins have been utilized widely to monitor the end-to-end distance, R_{ee}, of proteins *in vitro* and in-cell (30, 31). Both fluorescent proteins (30) (FPs) and small-molecule dyes are plausible choices for the FRET probes. While small-molecule dyes provide the ability for facile site-specific conjugation to side chains at internal

positions, FPs can be genetically encoded for in-cell experiments and introduce no surface-exposed xenobiotic moiety, potentially making them a suitable probe for comparative measurements *in vitro* and in-cell; FPs do introduce an additional protein interaction (similar to crowding and sticking in cells) and additional charge that needs to be taken into account.

We tested the applicability of each method by purifying our candidate protein Huntingtin-interacting protein K (HYPK), labeled with eGFP and mCherry ("HYPK-FRET") from heterologous expression in *E. coli* (see Methods), as well as HYPK labeled with Alexa Fluor 488 and Alexa Fluor 594 (HYPK-dye). Both constructs reproduced heat-induced expansion in phosphate buffer (Figure S1), but in solutions with added crowder (Figure S1; 100-300 mg/mL PEG-6000), the two probes diverge. The acceptor emission intensities of HYPK-FRET measurements increase at all temperatures by increasing the concentration of crowder. The acceptor emission intensities of HYPK-dye, however, decrease by increasing the concentration of crowder at temperatures less than 30 °C. The opposing trend of HYPK-dye can be attributed to either intermolecular dye-dye or dye-HYPK sticking under crowded conditions, when volume exclusion overcomes the electrostatic repulsion between small negatively charged groups (-2 for Alexa Fluor dyes and -8 for HYPK), or to intramolecular interactions, for example local clusters of positively charged amino acid side chains with Alexa Fluor. Although we can only rationalize the result in hindsight here, there are such positive clusters (e.g. RRR right next to one of the dye attachments, and RRSR mid-sequence in Figure S2) that are plausible candidates.

While FP labels can also interact with the target protein, they have been shown to introduce only mild perturbations of stability (32) and, when properly chosen, avoid intermolecular interaction in-cell (33). Of course, intermolecular interactions would not be an issue at the picomolar concentrations used for single molecule studies of IDPs (31, 34). In conclusion, for our present purpose, FPs are preferable to Alexa Fluor dyes because they produce the expected monotonic FRET curves upon crowding. Moreover, tethered FPs behave more like neighboring proteins would inside the cell, and indeed, our FP-labeled constructs express well for future in-cell measurements (SI Figure S3) of HYPK chain contraction or expansion, where they avoid a labeling and injection step. Hence, we picked FPs as the label of choice for the remainder of the study to minimize interactions and allow extension to in-cell work, rather than ideal FRET distance. One important factor to keep in mind when using any label: the labels may also carry charge (Table 1), which needs to be considered as an additional source of end-to-end electrostatic interaction in the FP-labeled constructs.

All FRET-labeled proteins expand upon heating, contract upon crowding, while HYPK shows cooperativity and unusual sensitivity to charge screening. We sought three protein candidates (Table 1) near the Uversky boundary that span a range of disorder as inferred from IUPred scores and AlphaFold/X-Ray structures (Figure 1a and Figure S2). HYPK comprises a long disordered N-terminal fragment, able to form a C-terminal helical bundle that mediates the interaction to its binding partner NatA (25). It is very close to the "Uversky boundary" and we judged it to be most likely to be environmentally sensitive. We selected cAMP-dependent protein kinase inhibitor alpha (PKIα) also near the boundary, but likely to be a more typical IDP with two regions that have low helix-forming propensity in isolation and rigidify upon binding the target protein (24, 35). Finally, adapter molecule CRK isoform 1 (Crk1) consists of a short disordered linker between folded SH2 and SH3 domains (26), maintaining substantial secondary structure at all temperatures used in this study. Each protein was prepared as a FRET-labeled ("IDP-FRET") and label-free ("IDP") variant (see Methods and Figure S2 for details). Uncorrected FRET efficiencies were measured by exciting eGFP and recording the emission of mCherry and eGFP (Figure 1b):

$$E_{FRET} = \frac{A}{A + D}, \quad (1)$$

where A is the emission intensity of mCherry, and D is the emission intensity of eGFP. Although we do not quantitatively relate E_{FRET} to R_{ee} due to the unknown probability distribution of chain conformations, we expect an inverse relationship between R_{ee} and E_{FRET} for any given protein, and this is the trend we compare with the models.

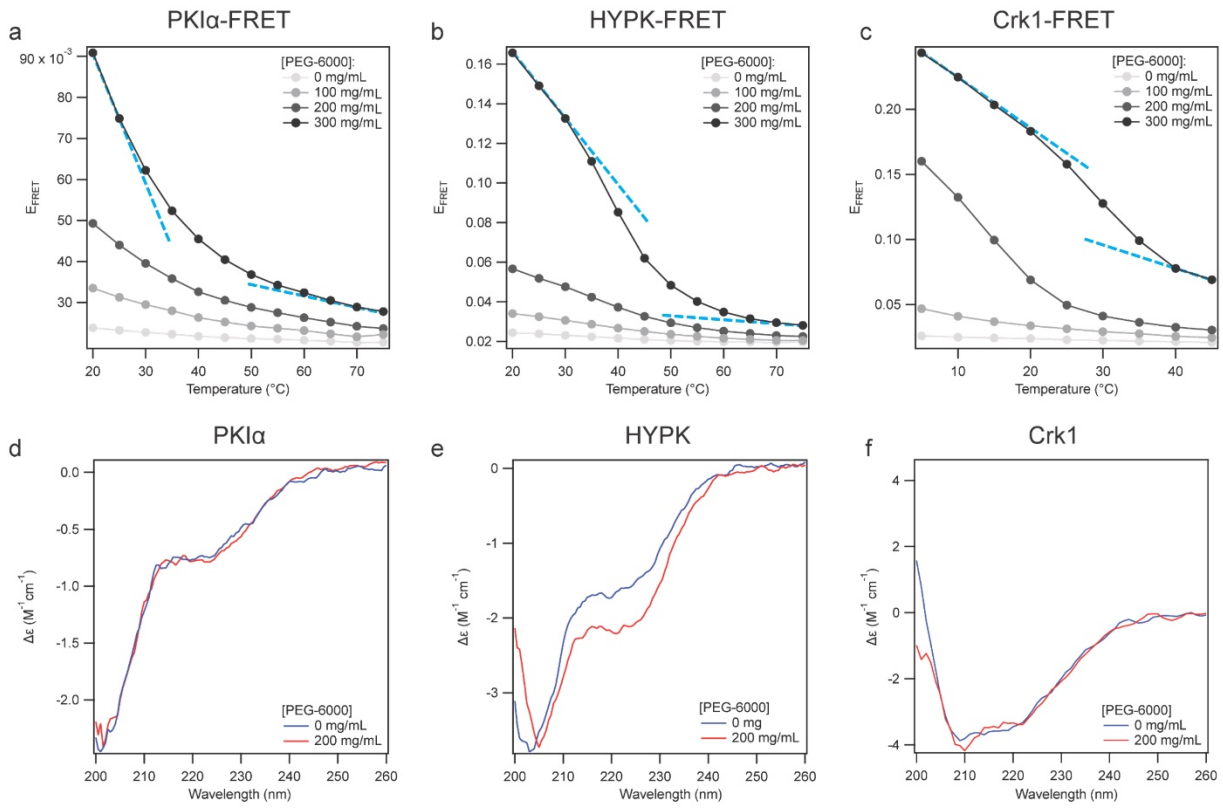

Figure 2. Trends in FRET efficiency and residual secondary structure when crowder concentration is increased (smaller R_{ee} yields higher FRET). (a-c) FRET curves of the IDP probes are shown at different concentrations of crowder as a function of temperature. The blue dotted baselines indicate that the FRET curves of HYPK and Crk-1 have inflection points at high concentration of PEG, whereas the one for PKIα does not. (d-f) CD spectra of the tag-free IDPs recorded at two concentrations of PEG-6000 at 20 °C.

We monitored FRET as a function of temperature and PEG-6000 concentration (crowder) for each FRET probe (Figure 2 a-c). We find that all proteins contract with increasing crowding, in agreement with previous studies (36, 37). We also find that all proteins expand with temperature. Such expansion with temperature is not typical of all IDPs. IDPs (38–40) that are highly charged are known to contract with temperature. IDPs with high net charge such as ProT-α contract due to rapid increase in solvation penalty of the charged groups with temperature (38). Hydrophobicity of IDPs also leads to contraction at higher temperature because the strength of the hydrophobic effect increases with temperature (41). For our IDPs near the folding boundary, these effects are evidently counterbalanced by the increase of chain entropy with temperature, typical of the behavior of folded proteins.

The secondary structure of PKIα (lowest propensity) and of Crk1 (mostly folded) is insensitive to crowder (Figure 2 d-f), hence the increase of R_{ee} with increasing temperature for these two proteins can be attributed mainly to expansion of the disordered chain. The temperature-invariant

secondary structure of PKIα and Crk1 indicates that the two proteins are either mostly disordered (PKIα), or folded except for short IDR stretches (Crk1). In contrast, the helical content of HYPK increases at 20 °C when crowder is added (inferred from the increase in ellipticity at 222 nm in Figure 2e), while at $T > 45$ °C the helix content of HYPK becomes insensitive to crowding (Figure S4), similar to PKIα and Crk1. Therefore, chain contraction and expansion of HYPK is influenced by local secondary structural changes at low temperature, but not at high temperature. HYPK, although disordered, shows its nature as a marginal folder by forming additional structure at higher concentrations of crowder and low temperature.

The marginal nature of HYPK is reinforced by the cooperative character of FRET emission vs. temperature (Figure 2b), reminiscent of folded proteins, while PKIα expands non-cooperatively when heated (Figure 2a), typical of most disordered proteins (42). Crk1-FRET, which lies on the folded side of the Uversky boundary in Fig. 1a, shares the cooperative character of HYPK (Figure 2c). Cooperativity requires an inflection point in the FRET signal, or an extremum in its derivative. HYPK and Crk-1 have an inflection point, but PKIa does not (Figure 2). We plotted the derivatives, showing a clear peak for HYPK (Figure S5) at 300 mg/ml PEG-6000. A normalized FRET curve of HYPK shows a Hill coefficient of 9±1 (see Supplementary Methods and Figure S6), whereas no such fit can be obtained for PKIa (Figure S7).

Next, we studied the effect of ionic strength on the end-to-end distance R_{ee} of these proteins since ionic strength dependent conformational changes are critical for function (43) and they can elucidate complex sequence features. Increasing the ionic strength of the solution reduces electrostatic interaction. The sensitivity of chain scaling to ionic strength can be rationalized by net charge of the sequence. Due to the high net charge of PKIα (-7) and HYPK (-9), increased further by the FP labels, we expected electrostatics to be repulsive which upon screening leads to chain compaction, whereas for Crk1 we expected a smaller R_{ee} perturbation because its intrinsically disordered region $Crk1^{IDR}$ has a net charge of zero (Table 1). In line with predictions, PKIα-FRET contracts at all temperatures when ionic strength is increased from 0 to 200 mM, above which contraction saturates (Figure 3a). Crk1-FRET contracts somewhat less (Figure 3c), explained by the net negative charge due to the labels (Table 1). In contrast, HYPK-FRET contracts by charge screening only above 45 °C (Figure 3b). Below 45 °C, HYPK-FRET expands when a monovalent salt (NaCl) is added. It appears that the formation of helical structure at low temperature (Figure 2) alleviates the net charge repulsion and HYPK behaves more like an ampholyte in a poor solvent. Some folded proteins expand/unfold when salt is added (44, 45), and HYPK shares this feature at temperatures below 45 °C despite its net charge, another sign that it is a marginal IDP.

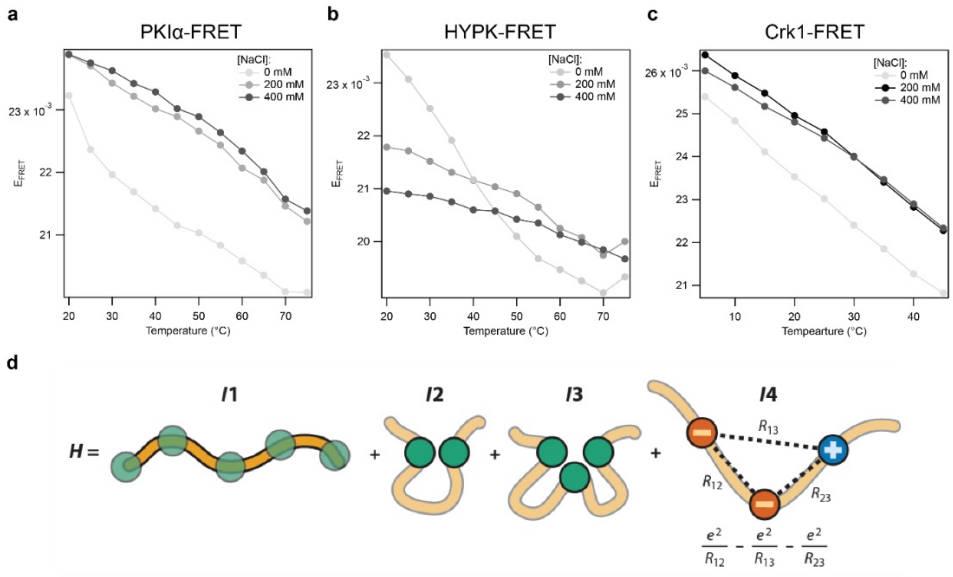

Figure 3. Effect of charge screening on protein dimension. (a-c) FRET curves of the IDP probes are shown as a function of temperature under different ionic strengths. All proteins at all [NaCl] expand with increasing the temperature. PKIα and Crk1 contract at all temperatures by charge screening, yet charge screening causes HYPK to contract at temperatures higher than 50°C and expand at lower temperatures. (d) a cartoon representing the intramolecular interactions that are accounted by the theoretical model. I1 captures chain connectivity, I2 accounts for two body interaction other than electrostatics, where I4 describes two body electrostatics interaction, and I3 accounts for a three-body repulsive interaction to avoid collapse. Details of the theory can be found in (10, 27) and in SI.

Charge patterning explains the conformational changes of charge-flipped IDP variants. Unlike PKIα's, HYPK's behavior is not explained by net charge alone. This motivated us to seek an explanation in more complex sequence features, such as charge patterning. By using mutations that change net charge or charge patterning in a controlled way, we can test the validity of charge patterning theory for marginal IDPs near the Uversky boundary. We thus created different variants of the two proteins HYPK and PKIα based on computational prediction and measured their response to salt.

The initial variants were designed by considering both total charge and their patterning captured by sequence charge decoration (SCD) metric that can be used to compute end-to-end dimension (10, 27). To achieve maximal change, we adopted the strategy of altering net charge by flipping charged residues. For PKIα (with 7 positive and 14 negative charges) we replaced positive charges by negative charges to increase the net charge and create polyelectrolyte-like sequences that are expected to be even more expanded than the wild type (Figure S2). On the other hand, for HYPK (22 positive and 31 negative charges) we replaced negative charges by positive charges to reduce the net charge and create polyampholyte-like sequences (Figure S2).

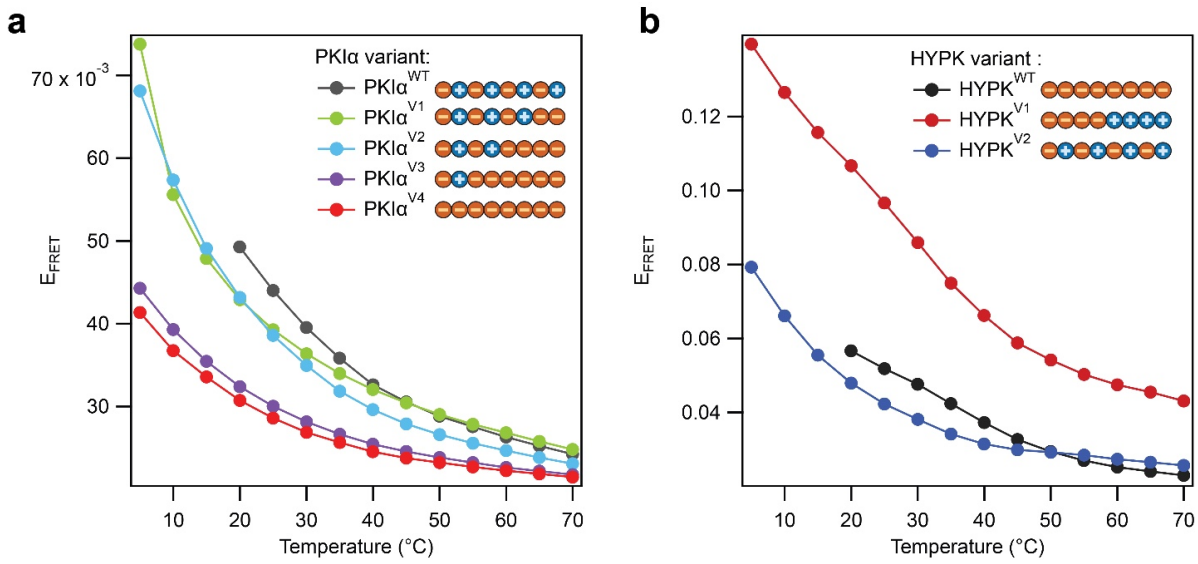

Figure 4. Relative end-to-end distances are accurately predicted by SCD. FRET curves of the (a) PKI α variants and (b) HYPK variants are shown as a function of temperature. The legend includes a schematic of charge decoration of the variants. Actual sequences can be found in Figure S2. The spectra were recorded in the presence of 200 mg/mL PEG-6000.

Of course, such a design strategy faces a combinatorial challenge: we must determine the specific charged residues to be flipped because multiple sequence variants can be created when flipping a fixed number of charges. This is a question of sequence patterning and not just of the net composition. We therefore used the predictive power of Sequence Charge Decoration (SCD; see SI for the definition of SCD) metric to break the degeneracy and facilitate the process of selecting specific charged residues to be flipped. Previous simulations have shown that SCD gives a quantitative measure of electrostatic effect on the chain dimension based on specific placement of charges and not just the net charge (10, 19, 46, 47). SCD also correlates with another well-known charge patterning parameter, κ, defined by Pappu and coworkers (12) when compared within a set of sequences having the same charge composition. We have chosen sequences that have the highest difference in SCD between the mutated and the wild-type sequence. For PKIα, we created variants with four charge flips (from basic to acidic residues). Among 35 possible sequences with four charge flips, we have chosen the sequence (PKIα^{V4}) with the largest SCD, which is expected to produce the most expanded conformation compared to the wild type (Figure S2). We have also selected a specific path in the design space that starts with one of the four charge flips and progressively flips one additional charge at a time to arrive at the variant V4. These intermediate sequences are named PKIα^{V1}, PKIα^{V2}, PKIα^{V3} for one, two and three charge flips (Figure S2). After purifying the PKIα-FRET variants we find that chain expansion trend is perfectly predicted by SCD (Figure 4a and SCD values reported in Figure S2): PKIα^{WT} < PKIα^{V1} < PKIα^{V2} < PKIα^{V3} < PKIα^{V4}. The

CD spectra of the variants indicate that the variants remain disordered, and the trends observed by FRET are due to global rather than local secondary structural changes (Figure S8). The gradual increase of the dimensions of PKIα variants is also reflected in the electrophoretic drift on SDS-PAGE (Figure S9), with more expanded variants drifting slower on the gel.

For HYPK, after converting four acidic residues to basic residues we get a net charge of -1. For this composition (26 basic residues and 27 acidic residues), we have 31465 different possible sequences. From this set, we designed the sequence ($HYPK^{V1}$) with the lowest SCD (-2.72) which is predicted to be more compact compared to the wild-type sequence (SCD = 0.98). We have also designed a second variant ($HYPK^{V2}$) that has the highest SCD (-0.47) among all sequences with 26 positive and 27 negative charges (Figure S2). We predicted that $HYPK^{V2}$ to be more expanded than $HYPK^{V1}$ despite both having identical net charge based on their differences in SCD. This expectation is also consistent with high κ (0.17) of $HYPK^{V1}$ compared to low kappa (0.07) for $HYPK^{V2}$. High value of κ indicates stronger segregation among opposite charges in the sequence, yielding more compact chain dimension.

We next carried out all-atom implicit solvent simulation (using CAMPARI; see Methods and SI) and find $HYPK^{V1}$ (R_{ee} = 4.3 nm) to be indeed more compact compared to $HYPK^{V2}$ (9.4 nm) and wild type HYPK (11.1 nm). Same trend of compaction is also seen in the normalized distance maps beyond just R_{ee} (Figure S10). We further verified our prediction against all-atom explicit solvent MD simulation (see Supplementary Methods) reflecting the same trend of compaction/expansion (see Table S2 in SI). Our theoretical formalism also allows quantitative prediction of R_{ee}(besides qualitative trend, see Supplementary Methods for details) of the mutants that compare reasonably (given the coarse grain nature of the theory compared) with the all-atom explicit water MD simulation and all-atom implicit water MC simulation (Table S2). MD and MC values, although reflect the same trends of swelling/compaction between wild type and mutants as expected from theory, differ among themselves for a given sequence, reflecting differences in the force-field. While SCD-based theoretical predictions and the simulations are done for the tag-free sequence, charges on FPs can influence conformation of IDRs (48). To investigate the possible influence by FP charges we have also simulated the two FP-labeled HYPK variants and found R_{ee} of $HYPK^{V1}$ to be 20.1 nm and 21.4 nm for $HYPK^{V2}$ (Table S3). This additional study ensured the expected trends are not altered by net charges on the fluorescent protein probes although the distances are, as expected, higher than for simulations done in isolation, likely due to electrostatic repulsion between the FPs swelling overall chain dimension.

After establishing the trends by theory and simulation, with and without FP labels, we again turned to experiment, made HYPK^{V1}-FRET and HYPK^{V2}-FRET, and found that at all temperatures, HYPK ^{V1}-FRET is more compact than HYPK^{V2}-FRET, consistent with the prediction and simulation (Figure 4b). Previous work has studied the effect of charge patterning on the hydrodynamic radius or radius of gyration, but not to end-to-end distance (12, 49, 50). Our experiment with extreme values of SCD demonstrate that predictive control over the variation of R_{ee} between two sequences that have the same charge composition but different charge patterning, is possible in a marginally disordered IDP like HYPK. Finally, HYPK wild type is more expanded than HYPK^{V2} at T > 50 °C, consistent with its high SCD (Figure S2). The relative trend between WT and HYPK^{V2} is however reversed at T < 50 °C, again pointing at the role of local structure, such as the extra secondary structure in Fig. 2, not modeled by SCD theory, dictating its conformation.

To determine the degree of cooperativity across variants we computed the derivatives of the FRET curve for HYPK^{v1}, and HYPK^{v2} from Figure 4b as described for Figure 2. HYPK^{v1} show a peak due to a cooperative transition (SI Figure S5c). Thus, HYPK^{v1} maintains a cooperativity similar to the WT in 200 mg/ml PEG-6000. This is in line with both molecules having a simulated (Monte Carlo and all-atom MD) bimodal distribution in distance between different pairs of amino acids, consistent with two structural ensembles (Figure S11, S12). Bimodal distributions are often associated with cooperative transition (51–53). In contrast, HYPK^{v2} does not show a peak in the derivative (Figure S5d). Thus, HYPK^{V2} have lost the cooperative thermal transition of HYPK wild type, even though it has same charge composition as HYPK^{v1}. We conclude that marginality is not fully determined by mean field parameters such as net charge used to determine the Uversky line, but also depends on the sequence patterning.

Our designed sequences are also predicted to have different salt response due to differences in charge patterning reflected in $SCD_{lowsalt}$ metric (See Table 1 and Supplementary Methods for definition). An earlier theory has shown the sign of $SCD_{lowsalt}$ can predict whether a chain will swell or contract upon addition of salt near zero salt, specifically, positive $SCD_{lowsalt}$ predicts contraction with added salt (near low salt regime), while the expansion happens for a negative value (27). The effect of charge screening on the R_{ee} of HYPK and PKIα is correctly predicted by $SCD_{lowsalt}$ (Figure 5). Variants with positive $SCD_{lowsalt}$ (all three PKIα-FRET variants) shrink by addition of salt, whereas HYPK^{V1}-FRET has a negative $SCD_{lowsalt}$ and expands by charge screening. Likewise, HYPK^{V2}-FRET has a small $SCD_{lowsalt}$ and in experiment does not show much sensitivity to charge screening. We did not modify the short, disordered linkers between the SH2 and SH3 domain of

Crk1. Over these small distances, SCD leads to small, predicted effects. Overall, salt dependent trends are well explained by $SCD_{lowsalt}$ metric of the variants and wild type, capturing the coupling between charge patterning and salt screening in the vicinity of the low salt regime.

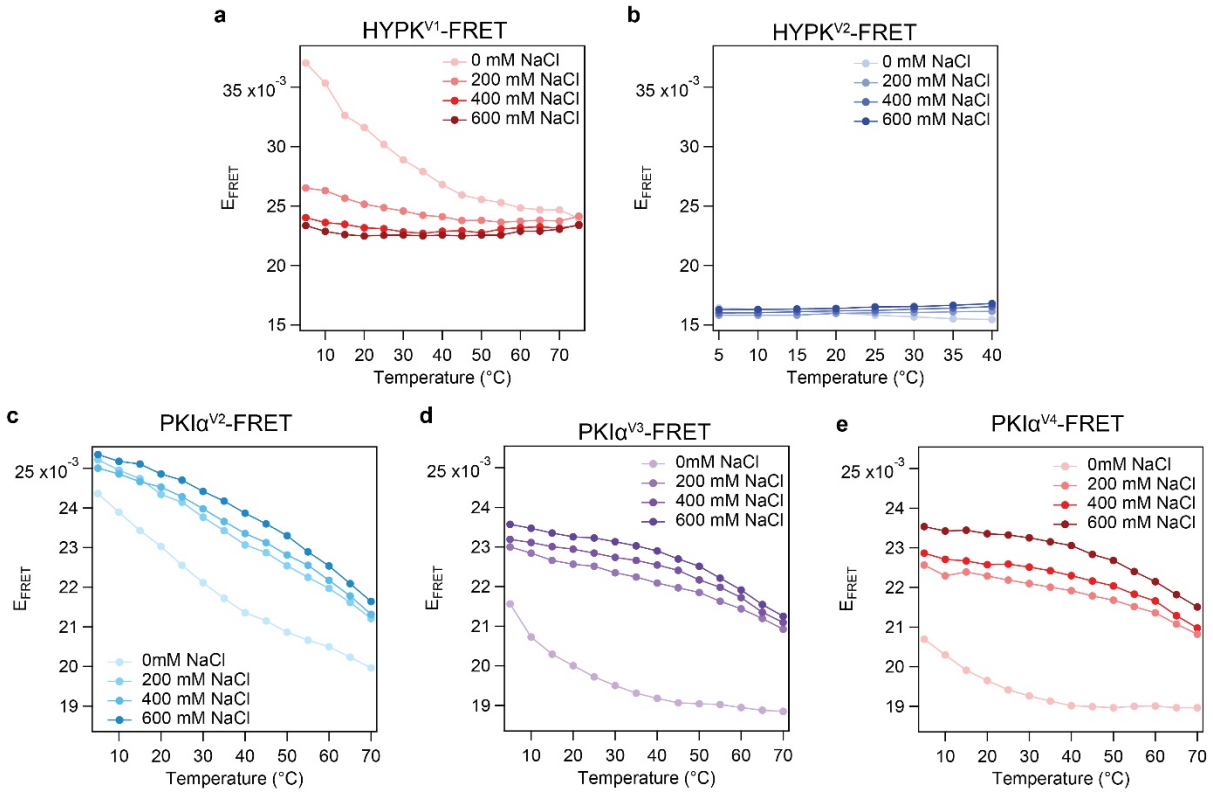

Figure 5. Effect of charge screening is correctly predicted by $SCD_{lowsalt}$. FRET curves of the (a) HYPK^{V1}-FRET, (b) HYPK^{V2}-FRET, (c) PKIα^{V2}-FRET, (d) PKIα^{V3}-FRET, (e) PKIα^{V4}-FRET variants are shown as a function of temperature and salt. As predicted by theory (Fig. S1) and in line with experiment, charge screening contracts the chain at all temperatures for variants with positive $SCD_{lowsalt}$ (PKIα^{V2}-FRET, PKIα^{V3}-FRET, and PKIα^{V4}-FRET), and expands the chain with negative $SCD_{lowsalt}$ (HYPK^{V1}-FRET). The salt-sensitivity of HYPK^{V2}-FRET is negligible, supported by the small $SCD_{lowsalt}$ value of -1.3.

DISCUSSION

We established a FRET assay inversely related to the end-to-end distance of the polypeptide chain to qualitatively compare the compaction of PKIα, marginal IDP HYPK, and IDR-containing but mostly folded Crk1, as a function of temperature, crowding, and ionic strength. We tested the contribution of charge-patterning in governing the end-to-end dimension (R_{ee}) of PKIα and HYPK by designing charge-flipped variants of each after making predictions. We used the Sequence Charge Decoration (SCD) parameter as a predictive guide to develop mutagenesis targets *in silico* and *in vitro.* We find that all three proteins near the Uversky boundary expand by heating

(Figure 2a-c), contrary to many highly charged disordered proteins or the unfolded state of hydrophobic proteins, which collapse with increasing the temperature (38–40). Polar homopolymers, such as polyQ in good solvent, show a similar temperature effect as the IDPs of our study (54), again indicating that our proteins lie in a middle ground between extremely charged (high solvation) and extremely hydrophobic unfolded proteins (54). The heat-induced collapse of the IDP chain has been attributed to the strengthening of hydrophobic interactions or solvation penalty of charged residues with temperature (25,32), and we attribute expansion here to gain in chain configurational entropy, which usually dominates for folded proteins at high temperature.

The effect of crowding (Figure 2d-f) and salt (Figure S4,S13,S14) on the secondary structure and the cooperativity of chain expansion (Figure 2a-c) shows that HYPK is the most “marginally” disordered of the three IDPs we studied near the Uversky boundary, while PKIα is a more typical IDP analogous to α-synuclein (56, 57), and Crk1 behaves like a protein with well-defined folded structure connected by short IDRs. NMR experiments have indeed shown that Crk1 can adopt a more rigid overall structure depending on the number of folded subunits included (26). In contrast, HYPK is sensitive to environment, and has some features of folded proteins, such as cooperativity, despite being disordered. Previous studies have shown that FlgM, an IDP, attains increased secondary structure upon crowding *in vitro* (23) suggesting FlgM, similar to HYPK, may also be a marginal IDP although its thermal scan and cooperativity data is unknown. FlgM has a lower net charge than HYPK (+2 vs -9), yet the mean-residue hydropathies are close (0.43 for FlgM and 0.39 for HYPK), overall placing FlgM on the folded side of the Uversky plot. Interestingly, equilibrium between the disordered and partially structured conformation in FlgM is shifted under *in vivo* condition as well. This raises the possibility that marginal IDPs may have evolved to sensitively respond to a changing cellular environment, adding to the emerging view that some IDPs act as sensors of cellular chemistry (58, 59) and biology.

The effect of charge-screening on the chain expansion/contraction is reconciled by arguments regarding the net charge and charge patterning of each IDP (Figure 3 a-c). All three FRET-labeled proteins (HYPK only at sufficiently high temperature for disorder to dominate) contract with charge-screening, in line with net negative charge of HYPK and PKIα. It is possible that the negatively charged FPs (eGFP is -7 and mCherry is -6) further expand the IDPs relative to the tag-free variants due to additional electrostatic repulsion, and part of the chain contraction with charge-screening is attributed to the effect of FP-FP repulsion, as seen in CAMPARI simulations (Table S3). Nonetheless, we argue that the intrinsic properties of highly charged and relatively high dimension of HYPK and PKIα outweigh the effect of distant FP labels on FRET

measurements at least to observe relative trends. Although $Crk1^{IDR}$ is net neutral, all the folded domains flanking the IDR are negatively charged (eGFP, mCherry, SH2, SH3) and the collapse can be attributed to charge-screened repulsion between folded domains. HYPK expansion at T < 50 °C with adding salt could be either caused by counterion condensation that shields the charged residues and changing the effective charge patterning of the sequence, or by the increased secondary structure content of HYPK at lower temperatures that could change the effective charge patterning of the sequence. Counterion condensation can also alter the nature of effective electrostatic interaction by charge reversal, and/or formation of dipoles giving rise to attractive dipolar interactions and would require an advanced theory, not captured by SCD theory at present. The effect of charge patterning, composition, and their coupling with screening on the conformation of the IDP (Figure 3d) is captured by $SCD_{lowsalt}$ that correctly predicts the trend of PKI α and HYPK (at T > 50 °C).

To demonstrate the predictivity of SCD, we made charge-flipped variants of PKIα and HYPK. We found the expansion or contraction of the R_{ee} of IDP variants (PKIα variants at all T, HYPK variants at high T), correlates with SCD. Most notably, $HYPK^{V1}$ and $HYPK^{V2}$ have equal net charge (-1), while having significantly different sizes that can only be explained by different charge-patterning and not by overall charge: segregated charges in $HYPK^{V1}$, and well-mixed opposite charges in $HYPK^{V2}$. Although there have been reports of the effect of charge-patterning on the R_g and R_H (14, 49, 50), our study is the first experimental validation of the effect of charge-patterning on a measure related to R_{ee}. Furthermore, $HYPK^{V2}$ despite having same charge composition as $HYPK^{V1}$, do not exhibit cooperative transition like $HYPK^{V1}$, indicating cooperativity is also sensitive to sequence patterning and not just to mean-field parameters such as net charge.

As a final validation for the sequence-dependent theoretical model, we measured the FRET curves of the variants at different salt concentrations and found perfect correlation to $SCD_{lowsalt}$ values: all PKIα-FRET variants collapse with charge screening (positive $SCD_{lowsalt}$), $HYPK^{V1}$-FRET expands with charge screening (negative $SCD_{lowsalt}$) while the scaling of $HYPK^{V2}$-FRET is relatively insensitive to charge screening in line with a small value of $SCD_{lowsalt}$. Overall, we show that the relative size of IDP variants is governed by electrostatics interactions and is accurately reconciled by SCD and $SCD_{lowsalt}$ charge patterning metrics, also tested across different segments of E-Cadherin in previous studies (60). Based on these results, the actual boundary of “marginality” may require a third metric, such as charge decoration, in Figure 1a.

MATERIALS AND METHODS

Here is a summary of the methods used in our study. Please see the SI for details.

Molecular biology techniques. The codon-optimized genes of His-tagged, FRET-labeled proteins were synthesized and cloned into pET28b vector using the following template: *His_6-SGSG*-eGFP-*SGSG*-IDP-*SGSG*-mCherry ("*IDP-FRET*"). Additionally, the *idp* genes were subcloned into pET28b-mbp vector, yielding TEV-cleavable MBP-tagged proteins that were used to purify tag-free IDPs. HYPK mutants were produced by gene synthesis, while PKIα mutants were made from site-directed mutagenesis of the $PKI\alpha^{WT}$ vectors.

Protein expression and purification. *IDP-FRET proteins*. His-tagged FP-labeled IDPs were produced by heterologous expression in *E. coli* and subsequent purification via immobilized metal affinity chromatography on an ÄKTA Pure FPLC system equipped with a HisTrap column.

Tag-free IDP proteins. To minimize the degradation of tag-free IDPs during bacterial expression, we opted for producing the proteins with a maltose-binding protein (MBP) tag, with subsequent cleavage and separation of MBP. Briefly, we expressed the MBP-tagged constructs in *E. coli* and purified the resulting proteins on an amylose gravity column. MBP was cleaved by overnight incubation with TEV protease, and the resulting tag-free IDP was separated from the mixture of TEV, MBP, and unreacted MBP-IDP on the ÄKTA Pure FPLC system equipped with a HisTrap column.

Protein-dye conjugation and purification. HYPK was labeled with Alexa Fluor 488 maleimide and Alexa Fluor 594 maleimide on the terminal Cys residues to provide the *HYPK-dye* sample. Since HPLC purification did not yield homogeneous samples, the FRET measurements were run on partially purified samples containing a mixture of HYPK labeled with permutations of both dyes. Briefly, HYPK was reacted with sub-equivalent amount of Alexa Fluor 488, and the mixture was partially purified by HPLC to remove the unreacted dye and some of the unreacted HYPK. The mono-labeled HYPK was reacted with Alexa Fluor 594 in the second step to yield HYPK-dye. The unreacted dye in the second step was separated from HYPK-dye using gel filtration.

Fluorescence and CD spectroscopy. FRET measurements were performed on a FP-8300 spectrofluorometer equipped with a Peltier temperature controller (JASCO). The donor fluorescent probe was excited at 488 nm and fluorescence spectra were collected from 450 to 700 nm. Uncorrected FRET efficiencies to monitor qualitative trends in end-to-end distance were calculated from the following equation:

$$E_{FRET} = \frac{A}{A + D}, \quad (1)$$

where *A* and *D* are the acceptor and donor emission intensities. Experiments were performed with 2.5 μM protein in 10 mM sodium phosphate buffer (pH 7.2) with varying amounts of NaCl and PEG-6000.

Circular dichroism was measured using a J-715 spectropolarimeter with Peltier temperature control (JASCO). All experiments were performed 10 or 20 μM protein concentration in 10 mM sodium phosphate buffer (pH 7.2), with varying amounts of NaCl and PEG-6000.

CAMPARI simulation protocol. CAMPARI version 4 was utilized for all-atom Monte Carlo simulations, employing the ABSINTH implicit solvent model (61, 62). Unless otherwise stated, all simulations used move sets and Hamiltonian parameters identical to those described in previous work (12). Non-bonded and electrostatic cutoffs were set to 12Å. Simulations of IDPs sampled all degrees of freedom available in CAMPARI. Simulations of IDRs constrained the folded domains to exclude backbone degrees of freedom while allowing side chain moves, a method which was reported in previous literature (48, 63). [see Supplementary Methods for more]

ACKNOWLEDGMENTS

A.F., S.A., and M.G. were supported by National Science Foundation grant NSF MCB 2205665. A. H., K.G and K.G. (Kingshuk Ghosh) were supported by National Institutes of Health grant R01 GM138901. K.G. acknowledges help from Alex Holehouse and Andreas Vitalis about CAMPARI. A.F. thanks Dr. Charles Wilson for assistance with protein purification.

REFERENCES:

1. R. Van Der Lee, *et al.*, Classification of intrinsically disordered regions and proteins. *Chemical reviews* **114**, 6589–6631 (2014).

2. V. N. Uversky, What does it mean to be natively unfolded? *European Journal of Biochemistry* **269**, 2–12 (2002).

3. V. N. Uversky, C. J. Oldfield, A. K. Dunker, Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annual Review of Biophysics* **37**, 215–246 (2008).

4. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197–208 (2005).

5. B. A. Shoemaker, J. J. Portman, P. G. Wolynes, Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proceedings of the National Academy of Sciences* **97**, 8868–8873 (2000).

6. A. Borgia, *et al.*, Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61–66 (2018).

7. R. K. Das, K. M. Ruff, R. V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins. *Current opinion in structural biology* **32**, 102–112 (2015).

8. V. N. Uversky, J. R. Gillespie, A. L. Fink, Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Bioinformatics* **41**, 415–427 (2000).

9. A. V. Dobrynin, M. Rubinstein, Theory of polyelectrolytes in solutions and at surfaces. *Progress in Polymer Science* **30**, 1049–1118 (2005).

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

11. P. G. Higgs, J. Joanny, Theory of polyampholyte solutions. *The Journal of Chemical Physics* **94**, 1543–1554 (1991).

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

13. H. Hofmann, *et al.*, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **109**, 16155–16160 (2012).

14. R. K. Das, Y. Huang, A. H. Phillips, R. W. Kriwacki, R. V. Pappu, Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. *Proceedings of the National Academy of Sciences* **113**, 5616–5621 (2016).

15. A. H. Mao, S. L. Crick, A. Vitalis, C. L. Chicoine, R. V. Pappu, Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **107**, 8183–8188 (2010).

16. S. M. Sizemore, S. M. Cope, A. Roy, G. Ghirlanda, S. M. Vaiana, Slow Internal Dynamics and Charge Expansion in the Disordered Protein CGRP: A Comparison with Amylin. *Biophysical Journal* **109**, 1038–1048 (2015).

17. S. Müller-Späth, *et al.*, Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **107**, 14609–14614 (2010).

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

19. T. Firman, K. Ghosh, Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *The Journal of Chemical Physics* **148**, 123305 (2017).

20. Y.-H. Lin, H. S. Chan, Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated. *Biophysical Journal* **112**, 2043–2046 (2017).

21. J. Huihui, K. Ghosh, An analytical theory to describe sequence-specific inter-residue distance profiles for polyampholytes and intrinsically disordered proteins. *The Journal of Chemical Physics* **152**, 161102 (2020).

22. E. W. Martin, *et al.*, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).

23. M. M. Dedmon, C. N. Patel, G. B. Young, G. J. Pielak, FlgM gains structure in living cells. *Proceedings of the National Academy of Sciences* **99**, 12681–12684 (2002).

24. C. Olivieri, *et al.*, Multi-state recognition pathway of the intrinsically disordered protein kinase inhibitor by protein kinase A. *elife* **9**, e55607 (2020).

25. F. A. Weyer, *et al.*, Structural basis of HypK regulating N-terminal acetylation by the NatA complex. *Nat Commun* **8**, 15726 (2017).

26. Y. Kobashigawa, *et al.*, Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. *Nature structural & molecular biology* **14**, 503–510 (2007).

27. J. Huihui, T. Firman, K. Ghosh, Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. *The Journal of Chemical Physics* **149** (2018).

28. K. Ghosh, J. Huihui, M. Phillips, A. Haider, Rules of physical mathematics govern intrinsically disordered proteins. *Annual review of biophysics* **51**, 355–376 (2022).

29. J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132 (1982).

30. C. S. Sørensen, M. Kjaergaard, Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proceedings of the National Academy of Sciences* **116**, 23124–23131 (2019).

31. B. Schuler, A. Soranno, H. Hofmann, D. Nettels, Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annual Review of Biophysics* **45**, 207–231 (2016).

32. K. Dave, H. Gelman, C. T. H. Thu, D. Guin, M. Gruebele, The Effect of Fluorescent Protein Tags on Phosphoglycerate Kinase Stability Is Nonadditive. *J Phys Chem B* **120**, 2878–2885 (2016).

33. S. Sukenik, P. Ren, M. Gruebele, Weak protein–protein interactions in live cells are quantified by cell-volume modulation. *Proc Natl Acad Sci USA*, 201700818 (2017).

34. J. Cubuk, M. D. Stuchell-Brereton, A. Soranno, The biophysics of disordered proteins from the point of view of single-molecule fluorescence spectroscopy. *Essays in Biochemistry* **66**, 875–890 (2022).

35. T. Güttler, *et al.*, NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. *Nat Struct Mol Biol* **17**, 1367–1376 (2010).

36. A. Soranno, *et al.*, Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Biophysical Journal* **106**, 268a (2014).

37. I. König, A. Soranno, D. Nettels, B. Schuler, Impact of in□cell and in□vitro crowding on the conformations and dynamics of an intrinsically disordered protein. *Angewandte Chemie* **133**, 10819–10824 (2021).

38. R. Wuttke, *et al.*, Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proceedings of the National Academy of Sciences* **111**, 5213–5218 (2014).

39. D. Nettels, *et al.*, Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proceedings of the National Academy of Sciences* **106**, 20740–20745 (2009).

40. M. Aznauryan, D. Nettels, A. Holla, H. Hofmann, B. Schuler, Single-Molecule Spectroscopy of Cold Denaturation and the Temperature-Induced Collapse of Unfolded Proteins. *J. Am. Chem. Soc.* **135**, 14040–14043 (2013).

41. N. T. Southall, K. A. Dill, A. D. J. Haymet, A View of the Hydrophobic Effect. *J. Phys. Chem. B* **106**, 521–533 (2002).

42. V. N. Uversky, Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. *Protein J* **28**, 305–325 (2009).

43. R. Vancraenenbroeck, Y. S. Harel, W. Zheng, H. Hofmann, Polymer effects modulate binding affinities in disordered proteins. *Proceedings of the National Academy of Sciences* **116**, 19506–19512 (2019).

44. W. H. Sawyer, J. Puckridge, The dissociation of proteins by chaotropic salts. *J Biol Chem* **248**, 8429–8433 (1973).

45. H. Maity, A. N. Muttathukattil, G. Reddy, Salt Effects on Protein Folding Thermodynamics. *J. Phys. Chem. Lett.* **9**, 5063–5070 (2018).

46. D. S. Devarajan, *et al.*, Effect of charge distribution on the dynamics of polyampholytic disordered proteins. *Macromolecules* **55**, 8987–8997 (2022).

47. J. McCarty, K. T. Delaney, S. P. O. Danielsen, G. H. Fredrickson, J.-E. Shea, Complete Phase Diagram for Liquid–Liquid Phase Separation of Intrinsically Disordered Proteins. *J. Phys. Chem. Lett.* **10**, 1644–1652 (2019).

48. I. Taneja, A. S. Holehouse, Folded domain charge properties influence the conformational behavior of disordered tails. *Current Research in Structural Biology* **3**, 216–228 (2021).

49. K. P. Sherry, R. K. Das, R. V. Pappu, D. Barrick, Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proceedings of the National Academy of Sciences* **114**, E9243–E9252 (2017).

50. W. H. Shi, R. S. Adhikari, D. N. Asthagiri, A. B. Marciel, Influence of Charge Block Length on Conformation and Solution Behavior of Polyampholytes. *ACS Macro Lett.* **12**, 195–200 (2023).

51. H. S. Chan, Z. Zhang, S. Wallin, Z. Liu, Cooperativity, Local-Nonlocal Coupling, and Nonnative Interactions: Principles of Protein Folding from Coarse-Grained Models. *Annual Review of Physical Chemistry* **62**, 301–326 (2011).

52. M. Knott, H. S. Chan, Criteria for downhill protein folding: Calorimetry, chevron plot, kinetic relaxation, and single-molecule radius of gyration in chain models with subdued degrees of cooperativity. *Proteins: Structure, Function, and Bioinformatics* **65**, 373–391 (2006).

53. A. Soranno, Physical basis of the disorder-order transition. *Archives of Biochemistry and Biophysics* **685**, 108305 (2020).

54. G. L. Dignon, W. Zheng, Y. C. Kim, J. Mittal, Temperature-Controlled Liquid–Liquid Phase Separation of Disordered Proteins. *ACS Cent. Sci.* **5**, 821–830 (2019).

55. R. B. Best, Emerging consensus on the collapse of unfolded and intrinsically disordered proteins in water. *Current opinion in structural biology* **60**, 27–38 (2020).

56. B. C. McNulty, G. B. Young, G. J. Pielak, Macromolecular Crowding in the Escherichia coli Periplasm Maintains α-Synuclein Disorder. *Journal of Molecular Biology* **355**, 893–897 (2006).

57. F.-X. Theillet, *et al.*, Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature* **530**, 45–50 (2016).

58. D. Moses, G. M. Ginell, A. S. Holehouse, S. Sukenik, Intrinsically disordered regions are poised to act as sensors of cellular chemistry. *Trends in Biochemical Sciences* (2023) https:/doi.org/10.1016/j.tibs.2023.08.001 (September 15, 2023).

59. D. Moses, *et al.*, Structural biases in disordered proteins are prevalent in the cell. *Nat Struct Mol Biol* **31**, 283–292 (2024).

60. F. Wiggers, *et al.*, Diffusion of a disordered protein on its folded ligand. *Proceedings of the National Academy of Sciences* **118** (2021).

61. A. Vitalis, R. V. Pappu, ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of Computational Chemistry* **30**, 673–699 (2009).

62. A. Vitalis, R. V. Pappu, "Chapter 3 Methods for Monte Carlo Simulations of Biomacromolecules" in *Annual Reports in Computational Chemistry*, R. A. Wheeler, Ed. (Elsevier, 2009), pp. 49–76.

63. J. Cubuk, *et al.*, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* **12**, 1936 (2021).