bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Hidden structure in disordered proteins is adaptive to intracellular changes

David Moses[1,2], Karina Guadalupe[1,2], Feng Yu[2,3], Eduardo Flores[1,2], Anthony Perez[1,2], Ralph McAnelly[1], Nora M. Shamoon[2,4], Estefania Cuevas-Zepeda[1], Andrea Merg[1,2], Erik W. Martin[5,†], Alex S. Holehouse[6,7], Shahar Sukenik[1,2,3,8,*]

1 Department of Chemistry and Biochemistry, University of California, Merced, CA
2 Center for Cellular and Biomolecular Machines, University of California, Merced, CA
3 Quantitative Systems Biology Program, University of California, Merced, CA
4 California State University, Stanislaus, Turlock, CA
5 Department of Structural Biology, St Jude Children's Research Hospital, Memphis, TN
6 Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St Louis, MO
7 Center for Science and Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO
8 Health Sciences Research Institute, University of California, Merced, CA
† Current address: Dewpoint Therapeutics, Boston, MA
* Corresponding author, e-mail: ssukenik@ucmerced.edu

Abstract

Intrinsically disordered protein regions (IDRs) are ubiquitous in all proteomes and essential to cellular function. Unlike folded domains, IDRs exist in an ensemble of rapidly changing conformations. The sequence-encoded structural biases in IDR ensembles are important for function, but are difficult to resolve. Here, we reveal hidden structural preferences in IDR ensembles *in vitro* with two orthogonal structural methods (SAXS and FRET), and demonstrate that these structural preferences persist in cells using live cell microscopy. Importantly, we demonstrate that some IDRs have structural preferences that can adaptively respond to even mild intracellular environment changes, while other IDRs may display a remarkable structural resilience. We propose that the ability to sense and respond to changes in cellular physicochemical composition, or to resist such changes, is a sequence-dependent property of IDRs in organisms across all kingdoms of life.

Keywords

Disordered proteins; live-cell imaging; FRET; SAXS; solution conditions; cellular environment

Introduction

Intrinsically disordered protein regions (IDRs) play key roles in many cellular pathways[1,2] and are vital to cellular function across all kingdoms of life. Compared to well-folded proteins, IDRs lack a stable tertiary structure, have fewer intramolecular interactions, and expose a greater area of their sequence to the surrounding solution[3]. As a result, an IDR exists in an ensemble of conformations that can change rapidly in response to the physical-chemical characteristics of its surroundings[4,5].

An extensive body of work has established the importance of the structural preferences of IDR conformational ensembles to their function[2,6–11]. Transient secondary structure, i.e., mostly *α*-helical segments that form transiently within the ensemble, modulates binding affinity in

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

PUMA[8] and p53[9] as well as liquid-liquid phase separation behavior in TDP-43[11]. Conformational changes produced by longer-range interactions influence IDR function in p53[12], BMAL1[13] and Myc[14].

The mapping between IDR sequence and molecular function is complex[15,16]. The relative importance of amino acid composition compared with specific linear sequence varies on a case-by-case basis. Some IDRs contain well-conserved short linear motifs (SLiMs) that mediate function[17]. Other IDRs that are poorly conserved in terms of sequence, such as transcription activation domains[18] and phosphorylation sites[19], may possess molecular features stemming from amino acid composition that mediate function[18–22]. Yet questions remain as to the relationship between amino acid composition, sequence, and ensemble-level structural preferences of IDRs.

Even less well resolved is how IDR ensembles respond to the complex physical-chemical changes that occur routinely in the intracellular environment[4,5,23,24]. Such changes may arise from natural processes such as the breakdown of the nuclear envelope during cell division[25] or from disease-driven perturbations such as changes in pH or intracellular metabolism in cancer cells[26,27]. A large portion of the experimental work that links IDR sequence, ensemble, and function has been performed *in vitro*, providing quantitative, high-resolution information about the structure of IDRs in a well-controlled environment. Yet it remains unclear if such findings can be applied to the same ensembles in the intracellular environment[28–31]. The complex mix of coexisting ions, small molecules, metabolites, and larger biomolecules make the cellular environment vastly different from a dilute aqueous solution for a highly sensitive IDR[32,33].

Here we aim to elucidate the structural preferences encoded in IDR ensembles and the adaptive response of these structural preferences to physicochemical changes in live cells. We do this by establishing a novel experimental framework that couples Förster resonance energy transfer (FRET) and small-angle X-ray scattering (SAXS) experiments *in vitro*. We interpret these results using homopolymeric dipeptide repeats as a benchmark against which to compare complex IDRs. We then leverage this framework to carry out FRET experiments in live cells and interpret how IDR ensembles respond to perturbations in intracellular physical-chemical composition *in situ*.

Results

IDR ensemble structure is revealed by coupled FRET and SAXS

To facilitate FRET assays, we placed our IDRs of interest between two fluorescent proteins (FPs), mTurquoise2 and mNeonGreen[34,35], that comprise a FRET pair[5] (**Fig. 1A**). As a benchmark against which to compare properties of naturally occurring heteropolymeric IDRs, we first measured homopolymeric peptides in our FRET construct (**Fig. 1**). Specifically, we chose glycine-serine (GS) repeat peptides of various lengths for benchmarking. GS repeats were chosen because of their prevalent use as linkers in synthetic constructs[36] and due to their lack of hydrophobicity, charge, and aromaticity which makes them easy to express and less prone to

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

aggregation[37]. Previous studies have shown that GS-repeat polymers do not form residual structure and act similarly to an ideal Gaussian chain homopolymer in aqueous buffered solutions[38–40].

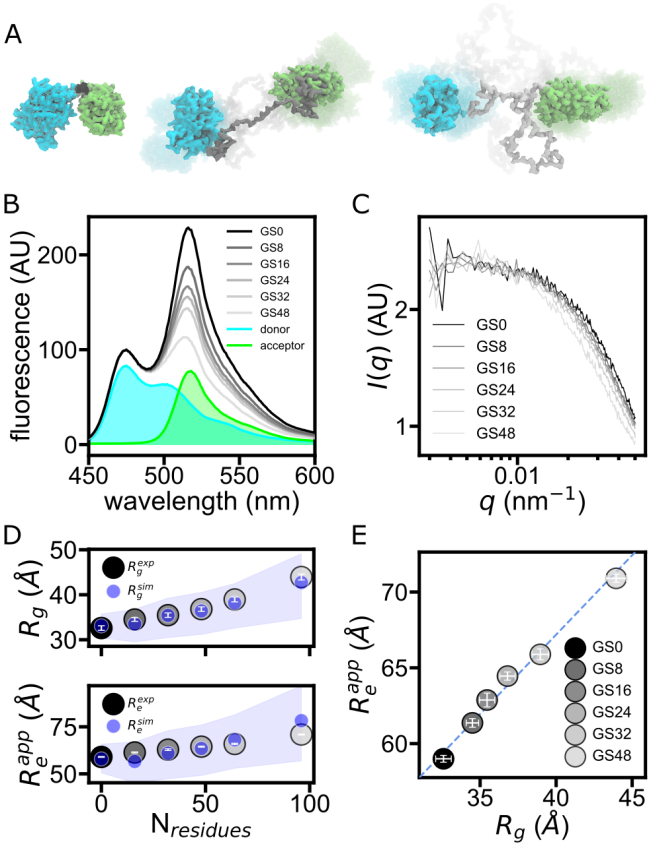


Figure 1. Experimental and simulated results for GS-repeat polymers in phosphate buffer. (A) Snapshots from all-atom simulations of FRET constructs flanking zero (left), 24 (center), or 48 (right) GS repeats (gray chains) in buffer. **(B)** Fluorescence spectra from measurements of FRET constructs incorporating glycine-serine (GS) polymers, notated GSX, where X indicates the number of GS repeats. **(C)** Scattering profiles from SAXS assays of GS constructs. **(D)** End-to-end distances and radii of gyration of FRET constructs incorporating GS-repeat polymers of different lengths derived from FRET and SAXS experiments (R_e^{exp}, R_g^{exp}) and simulations of FRET and SAXS experiments (R_e^{sim}, R_g^{sim}) are plotted against the number of residues in the GS-repeat polymer. Shaded regions represent median 50% of simulation results. **(E)** R_g (top) and R_e^{app} (bottom) derived from fluorescence spectra in **Fig. 1B** and SAXS data shown in **Fig. 1C** (see also Guinier fits in **Fig. S3)**, respectively. For all figures in this work, error bars for R_e^{app} are calculated from results of two repeats; error bars for R_g show fitting error from Guinier analysis (**Fig. S3**).

As previously reported, the FRET efficiency of our constructs decreased linearly with GS polymer length[5] (**Fig. 1B**), indicating increasing apparent end-to-end distance (R_e^{app}) (**Fig. 1D, 1E**). The use of FRET provides a metric for a single, average intramolecular distance which may not be enough to resolve the actual shape of the ensemble[41–44]. To obtain an orthogonal metric

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

that can inform about the ensemble's shape, we carried out size-exclusion chromatography coupled with small-angle X-ray scattering (SEC-SAXS) experiments on the same constructs as we used for FRET (**Fig. 1C**). We note that the enhanced solubility afforded by the use of fluorescent proteins allowed us to easily produce samples at the high protein concentrations that are optimal for SAXS experiments[45]. The chromatograms obtained from SEC showed a consistent, size-dependent increase in retention time (**Fig. S1**), indicating that the sequences increase in dimension with GS polymer length. SAXS intensity curves showed slopes that increased with GS polymer length (**Fig. 1C**), indicating linearly increasing radii of gyration (R_g), in agreement with FRET (**Fig. 1D, 1E**). Kratky plots derived from the SAXS data showed similar molecular form factors for all of the FRET constructs incorporating GS polymers of different lengths between the two FPs (**Fig. S2**), indicating random orientation on the part of the FPs, which constitute the majority of the scattering mass. Taken together, both SAXS and FRET report a linear dependence of ensemble size on GS length.

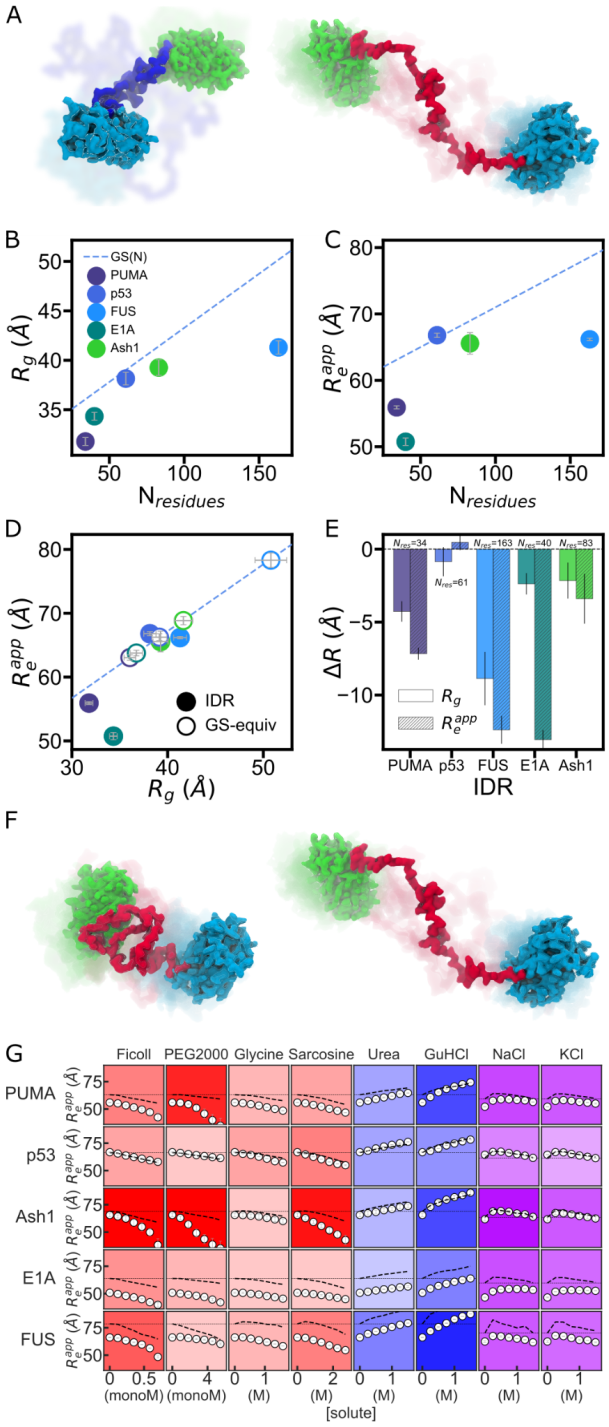
The difference between R_e^{app} and R_g shown in **Fig. 1D** is not surprising, but the ratio between R_e^{app} and R_g is significantly smaller than that expected from a Gaussian chain homopolymer model[46] ($R_e = \sqrt{6}R_g$). We suspected that this deviation is due to the presence of the FPs. To verify that this is not a result of measurement-related artefacts, we conducted simulations of the full-length construct, including both FP labels, at all-atom structural resolution with a simplified molecular forcefield (see *Methods*). We were able to quantitatively reproduce the GS length-dependent R_e^{app} values using ensembles subselected using only the SAXS scattering data (**Fig. 1D**). The internal consistency of our FRET, SAXS, and simulation results, and their agreement with previous findings, establishes a reliable homopolymer standard to which we can now meaningfully compare results for more complex IDRs.

Coupled FRET and SAXS experiments show that naturally occurring IDRs have hidden structure

Having gained an understanding of ensemble dimensions of GS-repeat homopolymers within our FRET construct, we inserted naturally occurring IDRs into our construct and again performed coupled FRET and SAXS assays. Specifically, we tested the BH3 domain of PUMA (abbreviated as PUMA), the N-terminal activation domain of p53 (p53), the low-complexity domain of FUS (FUS), the N-terminal region of the adenovirus hub protein E1A (E1A), and the C-terminal region of the yeast transcription factor Ash1 (Ash1) (**Table S1**). We reasoned that departure by any of these IDRs from the dimensions of a GS-repeat homopolymer of the same length in the same solution condition must indicate differing structural preferences (**Fig. 2A**). We found clear divergence in both R_g and R_e^{app} in all but one of these naturally occurring IDRs from the behavior of the GS-repeat homopolymers (**Fig. 2B-E, S3, S4**). In particular, interpolating along the line of the GS-repeat homopolymer data (blue dashed line in **Fig. 2B-D**), R_g and R_e^{app} of the naturally occurring IDRs do not correspond to those expected of GS-repeat

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

homopolymers of their respective lengths, which we refer to as their "GS-equivalents" (**Fig. 2D, 2E**, **S3**, and **S4**). We also note that R_g and R_e^{app} of the naturally occurring IDRs are generally lower than those of their homopolymer counterparts, and diverge from the values of their GS-equivalents by widely varying magnitudes (**Fig. 2E**).



bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2. Comparison of global dimensions and solution sensitivity of GS-repeat homopolymers and naturally occurring IDRs. (A) Schematic of FRET constructs incorporating two different IDRs (blue and red chains). The blue chain (left) is more collapsed than the red chain (right) due to the prevalence of attractive interactions in the sequence. **(B,C)** R_g and R_e^{app} vs. the number of residues for FRET constructs. Error bars indicate the error associated with Guinier fits shown in **Fig. S3** (B) or FRET measurements shown in **Fig. S4** (C)**.** Blue dashed lines are values for GS homopolymers. **(D)** $R_e^{app}$and R_g of IDR constructs. Hollow circles show $R_e^{app}$and R_g values of GS-repeat polymers of the same length as the IDRs. The blue dashed line is a linear fit of R_g vs. R_e^{app} for the GS-repeat homopolymers (same as **Fig. 1E**). **(E)** Deviations in R_e^{app} and R_g of IDR constructs from their corresponding "GS-equivalent" (GS polymer of the same length) values. **(F)** Schematic of a FRET construct incorporating the same IDR (red chains) in two different solution conditions. The chain on the right is more expanded than the chain on the left due to differing attractive and repulsive interactions between the chain and the two solutions. **(G)** Solution-space scans of IDR constructs. Each cell shows changes in R_e^{app} as a function of increasing solute concentration. Blue background indicates expansion, red indicates compaction, and purple indicates a non-monotonic response, with deeper color indicating a stronger effect. Black dashed curves in each cell represent the R_e^{app} values for a GS-repeat homopolymer of the same length in the same solution conditions based on the results shown in **Fig. S5**. The black dashed horizontal line is the R_e^{app} value for a GS-repeat homopolymer of the same length in dilute buffer. monoM: Concentration of a polymer expressed as a concentration of monomeric units.

We interpreted this varying departure of $R_e^{app}$and R_g from homopolymeric behavior to imply the existence of otherwise hidden structural preferences in these IDR ensembles[42,44,47]. We hypothesized that if such hidden structure truly accounted for this varying behavior, then interaction with different solutes would perturb different ensembles to different extents[4,5,48]. Therefore, we probed this hidden structure using our FRET-based solution-space scanning approach[5] (**Fig. 2F, 2G**). Briefly, the FRET signal from a given construct is measured in increasing concentrations of different biologically-compatible solutes. The resulting changes in FRET report on the sensitivity of the construct to changes in physicochemical conditions, which is determined by the balance between intramolecular interactions of the ensemble and interactions with the solutions. Indeed, our results show distinctive solution-response behavior for each IDR (**Fig. 2G**) that is also different from that of GS-repeat homopolymers (**Fig. 2G, S5**). This wide range of responses to changes in solution conditions agrees with our expectations based on our FRET and SAXS results, further supporting the existence of sequence-dependent hidden structural biases in the ensembles of these naturally occurring IDRs.

Combined, the results of our experiments imply that sequence-dependent structural preferences explain the phenomenon of IDRs of similar lengths having very different global dimensions. In particular, our coupled FRET and SAXS experiments were able to confirm the previously demonstrated[41–44] decoupling of internal structure (represented by R_e^{app}) and overall size (represented by R_g) in complex IDR ensembles. Moreover, the different IDR ensembles, mediated by these differing structural biases, show widely differing responses to changes in solution conditions. Taken together, these results lead us to hypothesize that an IDR's global dimensions and solution-response behavior are encoded in its amino acid composition and sequence.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Structural preferences in IDR ensembles, and their responses to changes in solution composition, are determined by amino acid composition and sequence.

Are sequence-encoded structural preferences simply a consequence of IDR composition (e.g., total content of charged residues), or does the order of the amino acids matter? To answer this question, we generated sequence shuffles of the BH3 domain of PUMA (**Fig. 3A, Table S1**) and measured global and local structural preferences. Sequence shuffling retains the amino acid composition but disrupts sequence patterns, secondary structure elements, and short linear motifs[43,44]. We chose three specific shuffles with different values of parameters related to the patterning of charged or hydrophobic residues (**Fig. 3B**).

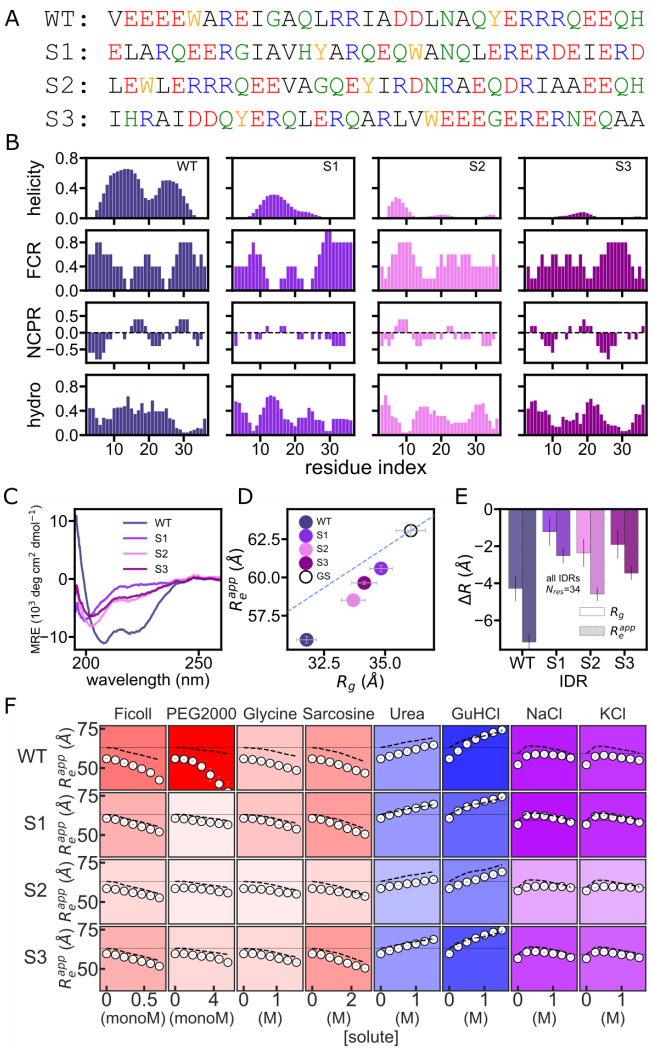


Figure 3. (A) Sequence of wild-type PUMA BH3 domain (WT PUMA) and three sequences (S1, S2, S3) derived by shuffling WT PUMA's sequence. Red: negative charge; blue: positive charge; black: hydrophobic residues; green: polar residues; orange: aromatic residues. **(B)** Molecular features of PUMA WT and shuffles. Predicted helicity calculated using all-atom simulations. Other parameters evaluated by localCIDER[49]. FCR: fraction of charged residues. NCPR: net charge per residue. Values represent the average of that residue number and its four nearest neighbors. **(C)** Circular dichroism spectroscopy signatures of PUMA variants without flanking FPs. **(D)** R_e^{app} and R_g of

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

PUMA constructs. The hollow circle shows the implied R_e^{app} and R_g values of a GS-repeat polymer of the same length as the IDRs. The blue dashed line is a linear fit of R_g vs. R_e^{app} for the GS-repeat homopolymers. **(E)** Deviations in R_e^{app} and R_g of PUMA constructs from their corresponding GS-equivalent values. **(F)** Solution-space scans of PUMA constructs. Each cell shows changes in R_e^{app} as a function of increasing solute concentration. Features are the same as **Fig. 2G**.

We first measured the secondary structure of the label-free constructs using circular dichroism (CD) spectroscopy. While wild-type (WT) PUMA displayed a previously reported tendency toward helicity[8,50] as shown by the minimum at 222 nm, the shuffles showed little or no helical content (**Fig. 3C**). This indicated that at least the local residual structure of the shuffles had been altered due to changing the amino acid sequence. All-atom simulations of the PUMA variants without flanking FPs support the trend seen in R_e^{app} and R_g for all constructs other than S2, which the simulations predict to be more expanded than our experiments show (**Fig. S6**). In agreement with previous work, simulations also confirm the presence of a broken helix in WT PUMA[51], yet find minimal helicity in the three scrambles. FRET and SAXS experiments further showed that all four sequences possess different R_e^{app} and R_g, with the behavior of the shuffles overall much closer to the behavior of GS-repeat homopolymers compared to WT PUMA (**Fig. 3D, 3E**). These results support the hypothesis that amino acid sequence, and not just composition or length, is a key determinant of the behavior of these IDRs.

We again hypothesized that these sequence-dependent differences in structural biases would manifest in differences in solution sensitivity. As expected, solution-space scans showed pronounced differences in the solution sensitivity of R_e^{app} between WT PUMA and the three sequence shuffles (**Fig. 3F**). Most prominently, WT PUMA deviates more from the behavior of a GS homopolymer of the same length (shown as a dashed black curve in **Fig. 3F**) than the sequence shuffles do. Together, these observations support the hypothesis that ensemble-level structural preferences and the solution-response behavior of the four sequences are determined by the arrangement of their amino acids rather than composition alone.

Structural preferences in the cell agree with *in vitro* behavior, but response to intracellular composition changes differs

Our *in vitro* experiments establish a well-defined framework linking IDR structural ensembles to their sensitivity to changes in solution conditions. But the intracellular environment is complex and heterogenous - a sharp departure from single-component aqueous buffers. Our next goal was therefore to leverage our system to measure IDR ensemble structure and its response to solution changes in live cells. The use of genetically encoded FPs in the cell facilitates direct and straightforward comparison with our *in vitro* experiments. We transfected FRET constructs incorporating WT PUMA and the three sequence shuffles into HEK293 cells (**Fig. 4A**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We first analyzed the basal FRET signal of the different PUMA constructs. We found that overall the trend in donor-to-acceptor ratio (D/A) measured *in vitro* persists in live cells, with WT < S2 < S3 < S1 (**Fig. 4B**). D/A ratios relate to the FRET transfer efficiency, and as such relate to the global dimensions of the system; the larger the D/A, the more expanded the ensemble. To compare *in vitro* ensemble behavior with that in cells, we measured D/A for purified proteins in a range of concentrations on the same microscopy setup. We found that D/A levels in live cells were markedly lower, indicating a more compact ensemble *in vivo*, in agreement with previous reports of in-cell IDR ensemble measurements[24,52] (**Fig. 4B, C**). We noted that the spread of the data *in vitro* and in cells correlates well, with WT and S2 having a markedly lower spread than S1 and S3, as measured by the standard deviation of the data (**Fig. 4D**). We also noted that the shape and size of the cell does not affect the D/A value (**Fig. 4E**). Overall, the basal D/A signal reveals that the amino acid sequence plays a role in the ensemble behavior in live cells as well as *in vitro*. Moreover, the ensembles measured in cells retain the trends seen *in vitro* in terms of global dimensions, yet overall are more compact than the ensembles measured *in vitro*.

Our next goal was to measure the structural responses of IDR ensembles to changes in the cell's composition. To alter the intracellular environment, we took a perturbative approach by rapidly changing the volume of the cell through changes to the imaging media. We did this either through the addition of purified water to achieve hypoosmotic conditions and cell volume increase, or through the addition of a concentrated NaCl solution for hyperosmotic conditions and cell volume decrease. We relied on previous work showing that at short times the main effect of an osmotic challenge is the efflux or influx of water from or into the cell[53,54]. We measured the signal of all channels before and after the perturbation (**Fig. 4A**). While the exact changes that occur in the cellular environment as a result of these perturbations are unclear, varying the amplitude of the osmotic challenge controls overall cellular concentrations, since cellular volume acts as a sensitive and robust osmometer[53,55]. We verified this by examining the change in direct acceptor fluorescence before and after exposure to an osmotic challenge (**Fig. 4F**). We found that cells exposed to hypoosmotic media (blue points in **Fig. 4F**) display a reduction in direct acceptor emission due to dilution, while in cells exposed to hyperosmotic media (red points in **Fig. 4F**) the opposite occurs.

The constructs show a markedly different response to changes in cell volume (**Fig. 4G**). The S1 and S3 shuffles display similar trends that are commensurate to those expected from macromolecular crowding: as cell volume decreases, the average D/A of the construct decreases as well[56,57]. WT PUMA and S2, on the other hand, show very small changes in response to the same volume changes. For S2, which seems to be the least responsive of the four sequences, this is commensurate with what was seen *in vitro* (**Fig. 3F**): S2 displayed the smallest changes in D/A of all PUMA variants. However, WT PUMA shows a marked change from its *in vitro* solution change dependence. This construct had the most significant sensitivity *in vitro* to nearly all solutes but especially to polymeric crowders such as PEG and Ficoll that are often used to mimic a crowded cellular environment[29,31]. In the cell, on the other hand, the construct retained nearly constant D/A, and its dimensions remained constant even in response to large cellular volume changes. It is important to note the control at 0.3 Osm includes the

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

addition of media to the well, verifying that the signal is not simply from perturbations caused by pipetting to the well (see Methods).

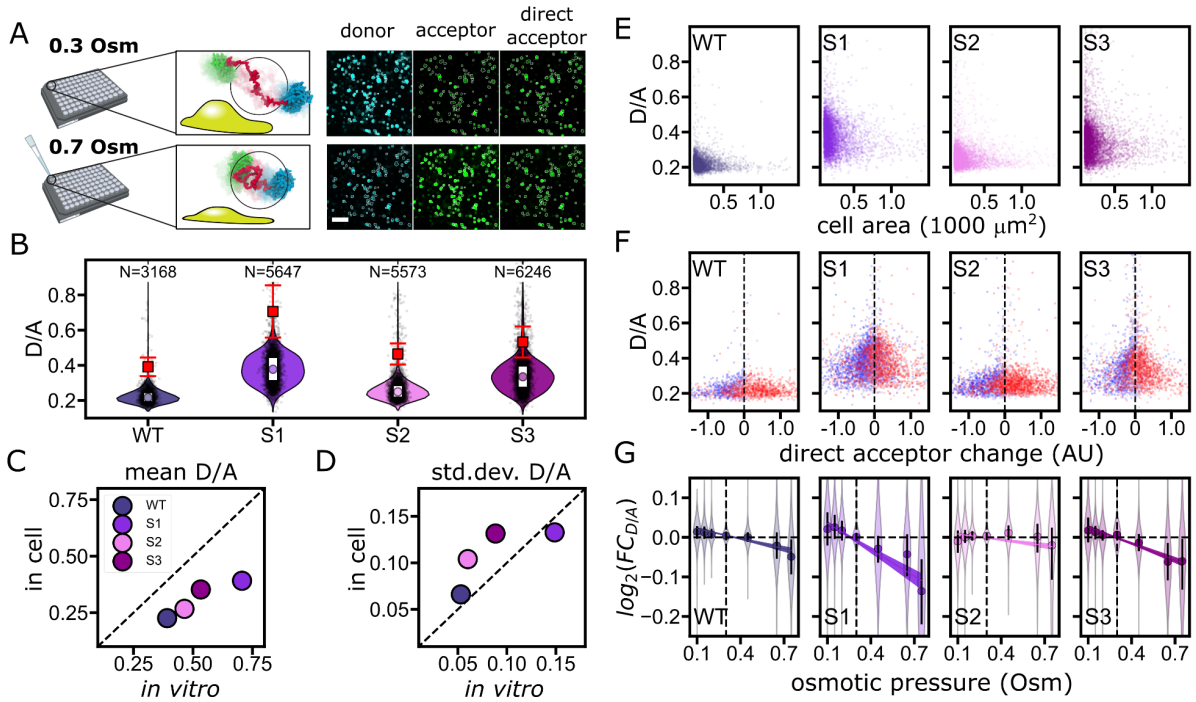


Figure 4. Measurement of WT PUMA and its sequence shuffles in live cells. **(A)** (top) FRET signal measured in cells under isosmotic conditions on three channels: donor and acceptor acquired simultaneously and direct acceptor as a proxy for concentration. (bottom) An osmotic challenge triggered by the addition of water or concentrated NaCl solution to imaging media changes the FRET signal. (**B**) D/A signal measured *in vitro* (red squares, error bars are standard deviation from multiple concentration measurements) and in cells (violin plots, single cells as black points, medians shown as white circles). N shows the number of cells from at least two repeats with at least three replicates each. **(C,D)** Mean (C) and standard deviation (D) of D/A in cells vs. *in vitro*. Dashed line shows equality. (**E**) In-cell D/A vs. cell area obtained from segmentation of live cell images. (**F**) D/A before osmotic challenge vs. change in directA after osmotic challenge. Red/blue points are cells subjected to a hyper/hypo-osmotic challenge, respectively. (**G**) Logarithmic fold-change in D/A signal ($FC_{D/A}$) following osmotic challenge, vs. the magnitude of the challenge. Vertical dashed line denotes isosmotic pressure, and horizontal lines show no change in D/A. Each violin plot represents the spread of the data, with the mean shown by a black circle and the median 50% spanned by the black bar. Lines are linear regressions of the entire dataset with 99.99% confidence intervals. All live cell data is available in **Table S2**.

Overall, the PUMA sequences measured in live cells corroborated the *in vitro* results. In addition, our experiments demonstrated that these IDR ensembles can change in response to even mild changes in their surrounding environment, and that this ability is encoded in the arrangement of their amino acids, and not just the average properties of the sequence.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Discussion

IDRs possess tunable, sequence-encoded structural biases that can determine their cellular function

To establish the *in vitro* dimensions of IDR ensembles, we used a two-pronged approach. First, we combined ensemble FRET and SAXS to obtain related but decoupled metrics of IDR ensemble dimensions: the end-to-end distance (R_e) and the radius of gyration (R_g). Second, we established a baseline for homopolymeric behavior using Gly-Ser sequences of various lengths as reference ensembles against which we could quantitatively compare naturally occurring IDRs. Our *in vitro* experiments showed unequivocally, and in agreement with other works[44,47], that the global dimensions of real IDR ensembles depart in a composition- and sequence-dependent manner from the dimensions of homopolymeric protein sequences of the same length. Next, we showed by shuffling IDR sequences that ensembles contain sequence-dependent hidden structural biases that define these dimensions. Finally, we showed that these structural preferences persist and are even enhanced inside living cells. Taken together, our work cements an emerging consensus that IDRs possess tunable, sequence-encoded structural biases, which we propose ultimately determine their cellular functions.

The use of FPs provides high-throughput, cross-scale examination of IDR ensembles

We use the FPs with the understanding that they may interact with themselves or with the disordered chain between them. While such intrachain interactions can drive changes in ensemble structure[58,59], our data indicates that this is not the major factor affecting the results in our experiments. Prior work has shown that the untethered FPs do not produce a significant FRET signal[5]. In our construct, the GS repeat lengths we assayed show a linear relationship between the number of GS repeats and both R_e^{app} and R_g. The quantitative fit shown by all-atom simulations that include only excluded volume of FPs (**Fig. 1D**) indicates that for GS repeats this approximation may hold, as previously reported for similar sequences[35]. Additional evidence for the relative inertness of FPs in these constructs is that the D/A signal shows a concentration-dependent decrease that can be fit by a self-crowding model[57] (**Fig. S7**) but not by an aggregation/dimerization model. This indicates that chains do not interact with each other, and that instead their behavior is dominated by macromolecular crowding[31,57,60]. Finally, the agreement of CD measurements (**Fig. 3C**) and simulations (**Fig. S6**) of the PUMA variants without FPs with FRET and SAXS measurements of the same IDRs with FPs (both *in vitro* and in live cells) (**Fig. 3D, 4B**) gives us confidence that the trends we are seeing are, at least primarily, a property of the IDR chain itself rather than a product of IDR:FP interactions.

Solution-space scanning reveals a link between hidden structural preferences and sensitivity to solution composition

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The cellular environment is often simplified as a chemically monolithic buffer, yet spatial and temporal regulation of volume, water content, pH, ions, and metabolites accompany key processes in cell biology[55,61,62]. Our work here establishes two related but orthogonal ideas. Firstly, IDR structural preferences observed *in vitro* can be retained in cells. Secondly, both in cells and *in vitro*, those structural preferences can be rewired in response to changes in the solution environment. We argue that adaptive and pathophysiological changes in the cellular environment may alter IDR function by rewiring conformational preferences.

In line with this idea, we show that different sequences have dramatically different responses to changes in solution composition, especially when compared to GS-repeat homopolymers (**Fig. 2G**, **Fig. S5**). This is in line with other observations that show both sensitivity[5,8,63,64] and insensitivity[65,66] of IDP function to different solution environments. In our constructs, the most prominent changes were seen in E1A, which displayed very strong deviations from GS-repeat homopolymers both in R_e^{app} and R_g and in its response to solute changes, and p53 NTAD, which shows the closest behavior to a GS homopolymer in each of these metrics (**Fig. 2**). We propose that the deviations shown between IDR ensembles and their homopolymer equivalents in response to changes in their solution environment can act as an additional metric for the existence of hidden structure in IDR ensembles.

Hidden structure is encoded in amino acid sequence

By comparing the behavior of PUMA's BH3 domain and three shuffles of its sequence, we were able to focus on differences in ensemble-level structural preferences that were solely sequence-dependent. *In vitro* FRET, SAXS, and CD experiments showed that shuffling the sequence produced differing basal global dimensions and transient secondary structure despite identical length and amino acid content (**Fig. 3C-E**). Furthermore, the adaptive response of these sequences to different solution conditions shows that the sensing ability of IDRs is also encoded in amino acid sequence (**Fig. 3F**).

The encoding of ensemble sensitivity in sequence implies that, like any other sequence-encoded protein property, IDR sequence sensitivity has been subjected to evolutionary pressure. We propose that certain sequences have been evolutionarily designed to act as sensors and actuators of the cellular environment, while other sequences have been designed to buffer out these changes, as shown by the in-cell behavior of the PUMA BH3 domain (**Fig. 4G**). Indeed, for the PUMA BH3 domain, resisting these changes may be critical, as it has previously been shown that its residual structure correlates with its activity[8]. As our understanding of sensing modalities for IDRs expands, we expect to be able to design IDR-based sensors for specific physicochemical intracellular conditions. This has already been demonstrated for the case of osmotic sensing[67].

Hidden structure persists in the cell, but adaptive response differs from *in vitro*

Perhaps surprisingly, the sequences tested in live cells using our FRET assays produced results that closely paralleled the trends we had seen *in vitro*. This indicates that *in vitro* experiments, at

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

least in this instance, provide a robust model of IDR behavior in the cellular environment. This is shown by the fact that overall the trends in D/A persist (**Fig. 4B**). However, the response of the ensemble to osmotically-induced volume changes in live cells showed a different trend from what we had seen *in vitro*. While WT PUMA showed the highest sensitivity to the presence of solutes *in vitro* (**Fig. 3F**), this sequence displayed a robust resistance to cellular volume changes (**Fig. 4G**). This is especially interesting considering it has previously been shown that the structure of the PUMA BH3 ensemble, as well as its activity, is sensitive to salt concentrations[8]. One possibility is that this resilience in the cell may have evolved in WT PUMA specifically to prevent changes in activity due to routine changes in the cellular environment that may alter its function[8].

Conclusions and outlook

Where can the dynamic hidden structure in IDRs affect their function? The cellular environment is subject to constant changes in physicochemistry. With their sensing abilities, IDRs can act as sensors and actuators of events. For example, transcription factor proteins, responsible for the regulation and initiation of all transcription in the cell, often contain a disordered activation domain[22,68]. Adapting to the cellular environment through changes in ensemble structure provides a rapid, specific, and streamlined way for disordered transcription factor regions to regulate transcription[69].

An additional implication of the evolved ability to sense and respond to changes in the environment is that a misregulated intracellular environment may disparately affect IDR function. Metabolic rewiring, a hallmark of cancer, viral infection, and other pathologies, can dramatically alter the physicochemical composition of the cell[70,71]. Even if this change would alter the activity of only a small subset of IDRs, their role as central signalling hubs could cause cellular malfunction. In this way, IDR sequences can be drivers of pathology in a deleterious cellular environment, even in the absence of mutations. We propose that this phenomenon is a previously overlooked driver of proteopathy.

Taken together, our results shed light on the enigmatic relationship between amino acid sequence, structural preferences, ensemble dynamics, and function of IDR conformational ensembles. Our experiments provide insights as to how IDRs can work as sensors and actuators in cells, pointing in particular to the role of ensemble-level structural dynamics.

Supplementary data

Supplementary Figures S1-S10 are available in Supplementary Information. Supplementary Tables S1-2 are available as csv files. Source data and code to produce all figures in this manuscript are available online at https://github.com/sukeniklab/Moses_2021.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Acknowledgements

We thank H.B. Schmidt, M. Thompson, J.A. Caro, and A. LiWang for helpful comments and discussion. We are indebted to J. Hopkins, S. Chakravarthy, and all BioCAT beamline staff at the Advanced Photon Source at Argonne National Laboratory, I. Rajkovic and all BioSAXS beamline staff at SLAC National Accelerator Laboratory for assistance with SAXS measurements. Research reported in this publication was supported by the NIH under award R35GM137926 to SS. KG is supported by a fellowship from NSF-CREST Center for Cellular and Biomolecular Machines (CCBM) at UC Merced, Grant No. NSF-HRD-1547848. This research used the Advanced Photon Source at Argonne National Laboratory under Contract No. DE-AC02-06CH11357, Proposal No. 75514. We acknowledge computing time on the MERCED cluster at UC Merced, NSF Grant ACI-1429783, and on the XSEDE computational infrastructure framework, Grant No. TG-MCB190103 to ASH and SS, supported by NSF Grant ACI-154856.

Materials and Methods

All-atom simulations of constructs with fluorescent proteins

To verify that our *in vitro* SAXS and FRET results report on the same conformational ensemble, we performed all-atom simulations of full-length constructs that include both fluorescent proteins using an identical amino acid sequence to the experimental constructs. Fluorescent protein models were constructed from PDB files 4AR7 (mTurquoise2)[72] and 5LTR (mNeonGreen)[73]. Simulations were performed using the ABSINTH implicit solvent model and CAMPARI Monte Carlo simulation engine[74].

Given the size of these proteins, simulating them at full-length and all-atom resolution raises a number of challenges. Given that our objective here was to determine whether SAXS and FRET were in agreement in the context of a simple homopolymeric linker, we took advantage of the ABSINTH implicit forcefield's ability to tune specific components of the Hamiltonian. Specifically, we performed simulations in which all excluded-volume interactions were present (i.e., the repulsive component of the Lennard-Jones potential was turned on). However, the attractive component of the Lennard-Jones potential was only turned on for residues within the glycine-serine (GS) linker, and limited only to intra-linker interactions by varying the inherent Lennard-Jones parameters of all atoms outside of the GS linker. Beyond these two components, all additional non-bonded Hamiltonian terms (i.e., long and short-range electrostatics and solvation effects) were turned off, dramatically lowering the computational cost of simulations. By systematically tuning the overall strength of the attractive GS-linker intramolecular interactions, we in effect performed simulations for GS homopolymers for all relevant homopolymer interaction strengths and GS-repeat lengths from 0 to 48 (i.e., 0 residues to 96 residues).

We initially performed simulations using a GS0 construct, where the only backbone degrees of freedom available were associated with the set of flexible residues that connect the two beta-barrels. Specifically, all backbone dihedral angles for amino acids within the two

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

beta-barrels were deactivated, but all sidechain dihedral moves were allowed. The residues between the two beta-barrels that had their backbone degrees of freedom sampled consist of amino acids 227 to 255 (GITLGMDELYKEGLSKLMVSKGEEDNMAS) in the GS0 construct[5]. After running thousands of short independent simulations in which these twenty-nine amino acids were sampled with variable intramolecular interaction strengths, we subselected an ensemble of 1000 distinct conformations which, on average, reproduced the experimentally measured SAXS scattering data for the GS0 construct (**Fig. S8A**). This GS0 ensemble was then used to define the starting configurations of the FPs and the 'handles' (non-GS component of the linker) for all other GS simulations.

For each of the other GS-repeat lengths (8, 16, 24, 32, 48), we performed simulations in which the attractive Lennard-Jones potential was scaled from 0.30 to 0.62 in steps of 0.02. This range straddles intramolecular interaction strengths that cause the longer GS chains to behave as a self-avoiding random coil (attractive LJ scaling parameter = 0.3) and a compact globule (attractive LJ scaling parameter = 0.62). For each combination of GS length and LJ strength, we performed 1000 independent simulations in which the fluorescent proteins and associated handles defined in the GS0 simulations were also fixed in place. As such, in total we performed 85,000 independent simulations for GS lengths 8, 16, 24, 32 and 48.

Each simulation was run in a spherical droplet with a radius of 500 Å to avoid any possible finite size effects. Given the absence of any electrostatic components, no ions were included in the simulations. Each simulation was run for 100,000 Monte Carlo steps. The first 50,000 steps were discarded as equilibration, and conformations were then sampled every 5000 steps. As such, each independent simulation generated 10 conformations, such that each GS/LJ combination generated a 10,000 conformer ensemble. Other than the repulsive component of the Lennard-Jones potential and (for some atoms) the attractive component of the Lennard-Jones potential, all other modes of nonbonded interactions were switched off. As such, each individual simulation takes on the order of 10 minutes.

After 17,000 separate simulations were performed for each GS-repeat length, we calculated predicted scattering profiles for each of the 17,000 simulations using FoXS software, as done previously[75,76]. For each GS-repeat length and for each of the 17,000 sub-ensembles, we calculated χ^2_{free} to assess how well each sub-ensemble compared to the experimentally measured scattering data. Using a chi-squared free threshold of 3.2 (a large value that reflects the relatively small error in the experimentally measured SAXS data), we generated sub-ensembles with scattering curves that quantitatively reproduced the experimental data at each of the GS-repeat lengths (**Fig. S8**).

Finally, using the SAXS-matched sub-ensembles, we computed the inter-barrel distance based on the distance between two residues in the center of the beta-barrel (**Fig. S9**). Distances were calculated between alpha-carbon atoms, such that we subtracted a 6 Å offset to approximately account for the distance between the alpha-carbon atoms and the anticipated chromophore centers. The resulting inter-beta-barrel distances, shown as R_e^{sim} in **Fig. 1D**, are in excellent

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

agreement with distances measured from ensemble FRET experiments. Taken together, this approach shows that the ensembles that best describe the SAXS data also correctly describe the distances inferred from FRET, confirming that these orthogonal methods are reporting on the same underlying conformational ensemble. The final sub-ensembles for each GS-repeat length and the associated data are provided in https://github.com/sukeniklab/Moses_2021. Simulation analysis was performed with SOURSOP (https://soursop.readthedocs.io/).

All-atom simulation of IDR region only

Simulations of label-free WT PUMA and its shuffles were done using the CAMPARI simulation suite and the ABSINTH forcefield[74,77]. For each sequence, five independent simulations were run at 310 K using $8x10^7$ Monte Carlo steps (following $1x10^7$ steps of equilibration) starting from random conformations to ensure proper sampling. Protein conformations were written out every 12,500 steps. The end-to-end distance and the helicity of the simulated conformation ensembles were determined using the MDTraj python library[78].

FRET construct design and cloning

The FRET backbone for bacterial expression (fIDR_pET-28a(+)-TEV) or for mammalian expression (fIDR_pCDNA3.1(+)) was prepared by ligating mTurquoise2 and mNeonGreen into pET28a-TEV or pCDNA backbone using 5' NdeI and 3' XhoI restriction sites. Genes encoding for IDR regions were obtained from GenScript (Piscataway, NJ) and ligated between the two fluorescent proteins using 5' SacI and 3' HindIII restriction sites. Cloned plasmids were amplified in XL1 Blue (Invitrogen) cell lines using manufacturer-supplied protocol. Sequences of all IDR sequence inserts are available in **Table S1**.

FRET construct expression and purification

BL21 (DE3) cells were transformed with fIDR_pET-28a(+)-TEV plasmids according to manufacturer protocol and grown in LB medium with 50 µg/mL kanamycin. Cultures were incubated at 37 °C while shaking at 225 rpm until OD600 of 0.6 was reached (approx. 3 h), then induced with 1 mM IPTG and incubated for 20 h at 16 °C while shaking at 225 rpm. Cells were harvested by centrifugation for 15 min at 3,000 rcf, the supernatant was discarded, and the cells were lysed in lysis buffer (50 mM NaH_2PO_4, pH 8, 0.5 M NaCl) using a QSonica Q700 Sonicator (QSonica, Newtown, CT). Lysate was centrifuged for 1 h at 20,000 rcf and the supernatant collected and flowed through a column packed with Ni-NTA beads (Qiagen). The FRET construct was eluted with 50 mM NaH_2PO_4, pH 8, 0.5 M NaCl, 250 mM imidazole, and further purified using size-exclusion chromatography on a Superdex 200 PG column (GE Healthcare) in an AKTA go protein purification system (GE Healthcare). The purified FRET constructs were divided into 200 µL aliquots, flash-frozen in liquid nitrogen, and stored at -80 °C in 20 mM sodium phosphate buffer, pH 7.4, with the addition of 100 mM NaCl. Protein concentration was measured after thawing and before use using UV-vis absorbance at 434 and 506 nm (the peak absorbance wavelengths for mTurquoise2 and mNeonGreen, respectively; the molar absorbance coefficients for mTurquoise2 and mNeonGreen are 30,000 $cm^{-1}M^{-1}$ and 116,000 $cm^{-1}M^{-1}$, respectively.[79] Calculations of concentration based on λ = 434 nm produced slightly

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

higher values than calculations based on λ = 506 nm, so the concentrations based on the measurement at λ = 506 nm were used), and purity was assessed by SDS-PAGE after thawing and before use.

Preparation of solutions for solution-space scanning

Solutes were purchased from Alfa Aesar (Sarcosine, PEG2000), GE Healthcare (Ficoll), Thermo Scientific (Guanidine Hydrochloride), and Fisher BioReagents (Glycine, Potassium Chloride, Sodium Chloride, Urea), and used without further purification. Stock solutions were made by mixing the solute with 20 mM sodium phosphate buffer, pH 7.4, with the addition of 100 mM NaCl except for experiments where the concentration of NaCl and KCl were varied, which began free of additional salt. The same buffer was used for all dilutions.

***In vitro* FRET experiments**

In vitro FRET experiments were conducted in black plastic 96-well plates (Nunc) with clear floors using a CLARIOstar plate reader (BMG LABTECH). Buffer, stock solution, and purified protein solution were mixed in each well to reach a volume of 150 μL containing the desired concentrations of the solute and the FRET construct, with a final concentration of 1 μM protein. Fluorescence measurements were taken from above, at a focal height of 5.7 mm, with gain fixed at 1020 for all samples. For each FRET construct, two repeats with 6 or 12 replicates each were performed in neat buffer, and two repeats were done in every other solution condition. Fluorescence spectra were obtained for each FRET construct in each solution condition by exciting the sample in a 16-nm band centered at λ = 420 nm, with a dichroic at λ = 436.5 nm, and measuring fluorescence emission from λ = 450 to 600 nm, averaging over a 10 nm window moved at intervals of 0.5 nm. Base donor and acceptor spectra for each solution condition were obtained using the same excitation and emission parameters on solutions containing 1 μM mTurquoise2 or mNeonGreen alone, and measuring fluorescence emission from 450 to 600 nm[79,80].

Calculation of FRET efficiencies and end-to-end distances

The FRET efficiency (E_f) of each FRET construct in each solution condition was calculated by linear regression of the fluorescence spectrum of the FRET construct with the spectra of the separate donor and acceptor emission spectra in the same solution conditions (in order to correct for solute-dependent effects on fluorophore emission). E_f was calculated using the equation:

$$E_f = 1 - \frac{F_d}{\frac{Q_d f_d}{Q_d f_a} F_s + F_d}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where F_d is the decoupled donor contribution, F_s is the decoupled acceptor contribution, f_d is the area-normalized donor spectrum, f_a is the area-normalized acceptor spectrum, Q_d = 0.93 is the quantum yield of the donor, and Q_a = 0.8 is the quantum yield of the acceptor[34,80].

The data for each series of solution conditions consisting of increasing concentrations of a single solute was processed in the following manner:

(1) Raw spectra for the free donor and free acceptor in the various solution conditions were loaded, and the averages of all repeats in each solution condition were computed. These averages are referred to as the "raw" donor and acceptor spectra below because they will be further corrected.
(2) The donor and acceptor peak intensities were assumed to change in a linear fashion with increasing solute concentration, so peak height of donor- or acceptor-only spectra vs. concentrations were linearly fit.
(3) To correct for artifacts (such as variations in FRET construct concentration between different wells) that may contribute to unexpected differences in fluorescence intensity, a correction factor was applied to each raw donor and acceptor spectrum to bring the peak intensity to the linear fit described in step 2, resulting in "corrected" donor and acceptor spectra. Importantly, we have seen in our previous work that this correction corrects well-to-well variations in raw data but has a negligible effect on overall values and trends[5].
(4) The raw FRET construct fluorescence spectra for the series were loaded.
(5) To compensate for unintended direct excitation of the acceptor by excitation at the donor excitation frequency, the corrected acceptor spectrum for each solution condition was subtracted from the FRET construct spectrum for each solution condition, resulting in "corrected" FRET construct spectra.
(6) The corrected donor, acceptor and FRET construct spectrum for each solution condition was fitted with a linear regression function to determine the decoupled contributions of the donor and acceptor to the FRET construct spectrum.
(7) E_f of each FRET construct in each solution condition was calculated using the equation shown above.
(8) The apparent end-to-end distance (R_e^{app}) of each FRET construct in each solution condition was calculated according to the equation:

$$E_f = \frac{R_0^{\ 6}}{(R_0^{\ 6} + R_e^{\ 6})}$$

where R_0 = 62 Å is the Förster distance of the FRET pair mTurquoise2 and mNeonGreen[34].

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Small-angle X-ray scattering experiments

Small-angle X-ray scattering (SAXS) experiments were performed at BioCAT (beamline 18ID at the Advanced Photon Source, Chicago). The experiments were performed with in-line size exclusion chromatography (SEC-SAXS) (**Fig. S1**) to separate monomeric protein from aggregates and improve the accuracy of buffer subtraction. Experiments were conducted at 20 °C in 20 mM sodium phosphate, pH 7.4, with 100 mM NaCl. Samples of approximately 300 µL were loaded, at concentrations in mg/mL approximately equal to 240 divided by the molecular weights of the constructs in kD (for example, a typical construct of molecular weight 60 kD would have a target concentration for SEC-SAXS of 240/60 = 4 mg/mL), onto a Superdex 200 Increase 10/300 column (GE Life Sciences) run at 0.6 mL/min. The column eluent passed through the UV monitor and proceeded through the SAXS flow cell which consists of a 1.5 mm ID quartz capillary with 10 µm walls. The column to X-ray beam dead volume was approximately 0.1 mL. Scattering intensity was recorded using a Pilatus3 1M (Dectris) detector placed 3.5 m from the sample providing access to a q-range from 0.003-0.35 $Å^{-1}$. 0.5 second exposures were acquired every 2 seconds during the elution. Data was reduced at the beamline using BioXTAS RAW[81,82] version 2.1.1. The contribution of the buffer to the X-ray scattering curve was determined by averaging frames from the SEC eluent which contained baseline levels of integrated X-ray scattering, UV absorbance and conductance. Frames were selected as close to the protein elution as possible and, ideally, frames pre- and post-elution were averaged. Final scattering profiles (**Fig. 1C**) were generated by subtracting the average buffer trace from all elution frames and averaging curves from elution volumes close to the maximum integrated scattering intensity; these frames were statistically similar in both small and large angles. Buffer subtraction and subsequent Guinier fits (**Fig. S3**) as well as Kratky transformations (**Fig. S2**) were done in BioXTAS RAW software. Radii of gyration (R_g) were calculated from the slope of the fitted line of the Guinier plot using the equation:

$$ln[I(q)] = ln[I(0)] - (\frac{R_g^2}{3})\, q^2$$

Mammalian cell culture

HEK293T cells were cultured in Corning treated flasks with Dulbecco's modified Eagle medium (Advanced DMEM:F12 1X, Gibco Cat. No. 12634-010) supplemented with 10% FBS (Gibco Cat. No. 16000-044) and 1% penicillin/streptomycin (Gibco Cat. No. 15140-122). For live-cell microscopy experiments, 5,000 cells were plated in a µ-Plate 96-well black treated imaging plate (Ibidi Cat. No. 89626) and allowed to adhere overnight (~16 hours) before transfection. Cells were incubated at 37 °C and 5% CO_2. Before transfection, the media was switched out with new warmed media. XtremeGene HP (Sigma Cat. No. 6366236001) was used to transfect FRET construct plasmids into HEK293T cells per manufacturer's protocol. Cells were incubated at 37 °C and 5% CO_2 for 48 hours. NaCl stock solution was prepared by dissolving NaCl (Fisher Bioreagents CAS 7647-14-5) in 1X PBS (Gibco Cat. No. 70011-044) and filtering using a 0.2 µm filter. The solutions used for perturbations were obtained by diluting 1X PBS with autoclaved DI

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

water to achieve hypoosmotic conditions or by adding NaCl stock solution for hyperosmotic conditions. To prepare for imaging, cells were rinsed twice with 1X PBS and left in 200 µL PBS (300 mOsm) for imaging.

Live-cell microscopy

Imaging was done on a Zeiss epifluorescent microscope using a 10X 0.3 NA dry objective. Excitation was done with a Colibri LED excitation module and data was collected on a duocam setup with two linked Hamamatsu flash v3 sCMOS cameras. The cells were imaged at room temperature before and after perturbation with 75 ms exposure times. Imaging was done by exciting mTurquoise2 at 430 nm (donor and acceptor channels, **Fig. 4A**) or mNeonGreen at 511 nm (direct acceptor channel, **Fig. 4A**). Emitted light was passed on to the camera using a triple bandpass dichroic (467/24, 555/25, 687/145). When measuring FRET, emitted light was split into two channels using a downstream beamsplitter with a 520 nm cutoff. For each perturbation, the cells were focused using the acceptor channel and imaged before manually adding water (hypoosmotic conditions), PBS (isosmotic condition) or NaCl solution (hyperosmotic conditions) with a pipette and pipetting up and down 10 times to ensure mixing. The final osmolarities that were used for the perturbations were: 100 mOsm, 150 mOsm, 200 mOsm, 300 mOsm (isosmotic), 450 mOsm, 650 mOsm, and 750 mOsm with NaCl as the osmotic agent. Imaging was typically completed in ~ 1 minute.

Image analysis

Images were analyzed using ImageJ[83]. Images collected before and after osmotic challenge, containing three channels each, were stacked and aligned using the StackReg plugin with rigid transformation. The aligned image was segmented based on the donor channel before perturbation. Segmentation was done using several methods to ensure that the results were robust. The methods included the ImageJ built-in implementations of the Triangle and MaxEntropy algorithm, as well as a fixed threshold that selected only pixels with an intensity of 2,000 or higher. All methods gave nearly identical results, so the fixed threshold method was finally selected for the data shown in **Fig. 4**. The resulting mask was corrected using the Open and Watershed binary algorithms. Cells were selected using the Analyze Particles option of ImageJ, selecting only those that were 100-2,000 μm^2 in size, and with a circularity of 0.3 or higher. The resulting ROIs were averaged in each channel at each timepoint. The resulting cells were filtered to remove cells with an intensity over 40,000 (which tended to include saturated pixels), and cells where the absolute change in direct acceptor emission was over 1,500 (which tended to be cells that moved or lifted off the coverslip during measurement). To correct for donor bleedthrough and cross-excitation, cells were transfected with the mTurquoise2 or mNeonGreen construct only, the cells were imaged and analyzed using the same protocol as previously mentioned, and correlation plots were generated to determine percent bleedthrough and cross-excitation (**Fig. S10**). The final filtering step removed cells with a corrected D/A ratio that was negative or higher than 2. The resulting dataset is available as **Table S2**. Analysis code is available in an ImageJ macro at https://github.com/sukeniklab/Moses_2021.

***In vitro* concentration dependence experiments.** To prepare the slides, pieces of ultra-thin 10

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

μm double-sided tape (Nitto Cat. No. NT5601) were hole-punched and stuck onto Fisherbrand plain microscope slides (Cat. No. 12-550D) so that the holes in the tape could serve as wells. Protein aliquot samples were diluted into a series of varying concentrations using 20 mM sodium phosphate, 100 mM NaCl, pH 7.4 buffer. Fluorescent beads (Phosphorex Cat. No. 2225) were added to the prepared aliquots to ensure focus on the bottom of the coverslip. Samples were loaded into the wells on the microscope slides and covered with a cover glass (VWR 16004-302). Imaging parameters were the same parameters as were used for the live-cell microscopy experiments.

Label-free peptide synthesis and purification

WT PUMA and shuffled sequences were prepared via standard microwave-assisted solid-phase peptide synthesis protocols using a Liberty Blue automated microwave peptide synthesizer (CEM, NC, USA) and ProTide Rink Amide resin (CEM). Fmoc-deprotection was achieved by treatment with 4-methylpiperidine (20% v/v) in dimethylformamide (Sigma-Aldrich), and Fmoc-amino acids were activated using N,N'-Diisopropylcarbodiimide (Sigma-Aldrich) and Oxyma Pure (CEM). Peptides were N-terminally acetylated and C-terminally amidated. After synthesis, the peptidyl resins were filtered and rinsed with acetone and air-dried. The crude peptides were cleaved from the resin for 4 hours at room temperature with a 92.5% trifluoroacetic acid (TFA), 2.5% H_2O, 2.5% 3,6-dioxa1,8-octane-dithiol, 2.5% triisopropylsilane cleavage solution, precipitated with cold diethyl ether, and centrifuged at 4000 rpm for 10 min at 4 °C. After centrifugation, the supernatants were discarded, and the pellets were dried under vacuum overnight. Crude peptides were purified by high-performance liquid chromatography (HPLC) using an Agilent 1260 Infinity II HPLC instrument equipped with a preparative scale Phenomenex Kinetex XB-C18 column (250 x 30 mm, 5 μm, 100 Å). Peptides were eluted with a linear gradient of acetonitrile-water with 0.1% TFA. The target fractions were collected, rotovapped, and lyophilized. Purified peptides were analyzed by mass spectrometry using a Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Scientific).

CD spectroscopy

Lyophilized protein constructs were weighed and dissolved to a final concentration of 20 μM in a 20 mM sodium phosphate, 100 mM sodium chloride buffer at pH 7.4. CD spectra were measured using a JASCO J-1500 CD spectrometer with 0.1 cm quartz cell (Starna Cells, Inc., Atascadero, CA) using a 0.1 nm step size, a bandwidth of 1 nm, and a scan speed of 200 nm/min. Each spectrum was measured 7 times and averaged to increase signal-to-noise ratio. The buffer control spectrum was subtracted from each protein spectrum.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18**, 343–384 (2005).
2. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
3. van der Lee, R. *et al.* Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* **114**, 6589–6631 (2014).
4. Holehouse, A. S. & Sukenik, S. Controlling Structural Bias in Intrinsically Disordered Proteins Using Solution Space Scanning. *J. Chem. Theory Comput.* **16**, 1794–1805 (2020).
5. Moses, D. *et al.* Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment. *J. Phys. Chem. Lett.* **11**, 10131–10136 (2020).
6. Babu, M. M., Kriwacki, R. W. & Pappu, R. V. Structural biology. Versatility from protein disorder. *Science* **337**, 1460–1461 (2012).
7. Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 9614–9619 (2015).
8. Wicky, B. I. M., Shammas, S. L. & Clarke, J. Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9882–9887 (2017).
9. Borcherds, W. *et al.* Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat. Chem. Biol.* **10**, 1000–1002 (2014).
10. Mittag, T., Kay, L. E. & Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **23**, 105–116 (2010).
11. Conicella, A. E. *et al.* TDP-43 α-helical structure tunes liquid-liquid phase separation and function. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5883–5894 (2020).
12. Schrag, L. G. *et al.* Cancer-Associated Mutations Perturb the Disordered Ensemble and Interactions of the Intrinsically Disordered p53 Transactivation Domain. *J. Mol. Biol.* **433**, 167048 (2021).
13. Xu, H. *et al.* Cryptochrome 1 regulates the circadian clock through dynamic interactions with the BMAL1 C terminus. *Nat. Struct. Mol. Biol.* **22**, 476–484 (2015).
14. Panova, S. *et al.* Mapping Hidden Residual Structure within the Myc bHLH-LZ Domain Using Chemical Denaturant Titration. *Structure* **27**, 1537–1546.e4 (2019).
15. Riback, J. A. *et al.* Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **358**, 238–241 (2017).
16. Best, R. B. *et al.* Comment on ‘Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water’. *Science* vol. 361 (2018).
17. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).
18. Sanborn, A. L. *et al.* Simple biochemical features underlie transcriptional activation domain diversity and dynamic, fuzzy binding to Mediator. *Elife* **10**, (2021).
19. Cho, M.-H., Wrabl, J. O., Taylor, J. & Hilser, V. J. Hidden dynamic signatures drive substrate selectivity in the disordered phosphoproteome. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23606–23616 (2020).
20. Zarin, T. *et al.* Proteome-wide signatures of function in highly diverged intrinsically

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

disordered regions. *Elife* **8**, (2019).
21. Zarin, T. *et al.* Identifying molecular features that are associated with biological function of intrinsically disordered protein regions. *Elife* **10**, (2021).
22. Staller, M. V. *et al.* A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Syst* **6**, 444–455.e6 (2018).
23. Vancraenenbroeck, R., Harel, Y. S., Zheng, W. & Hofmann, H. Polymer effects modulate binding affinities in disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19506–19512 (2019).
24. Gnutt, D., Gao, M., Brylski, O., Heyden, M. & Ebbinghaus, S. Excluded-volume effects in living cells. *Angew. Chem. Int. Ed Engl.* **54**, 2548–2551 (2015).
25. Smoyer, C. J. & Jaspersen, S. L. Breaking down the wall: the nuclear envelope during mitosis. *Curr. Opin. Cell Biol.* **26**, 1–9 (2014).
26. Hsu, P. P. & Sabatini, D. M. Cancer cell metabolism: Warburg and beyond. *Cell* **134**, 703–707 (2008).
27. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
28. Monteith, W. B., Cohen, R. D., Smith, A. E., Guzman-Cisneros, E. & Pielak, G. J. Quinary structure modulates protein stability in cells. *Proceedings of the National Academy of Sciences* **112**, 1739–1742 (2015).
29. Gruebele, M., Dave, K. & Sukenik, S. Globular Protein Folding In Vitro and In Vivo. *Annu. Rev. Biophys.* **45**, 233–251 (2016).
30. Davis, C. M., Gruebele, M. & Sukenik, S. How does solvation in the cell affect protein folding and binding? *Curr. Opin. Struct. Biol.* **48**, 23–29 (2018).
31. Rivas, G. & Minton, A. P. Macromolecular Crowding In Vitro, In Vivo, and In Between. *Trends Biochem. Sci.* **41**, 970–981 (2016).
32. Zimmerman, S. B. & Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol.* **222**, 599–620 (1991).
33. Cayley, S., Lewis, B. A., Guttman, H. J. & Record, M. T., Jr. Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo. *J. Mol. Biol.* **222**, 281–300 (1991).
34. Mastop, M. *et al.* Characterization of a spectrally diverse set of fluorescent proteins as FRET acceptors for mTurquoise2. *Sci. Rep.* **7**, 11999 (2017).
35. Sørensen, C. S. & Kjaergaard, M. Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23124–23131 (2019).
36. van Rosmalen, M., Krom, M. & Merkx, M. Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. *Biochemistry* **56**, 6565–6574 (2017).
37. Das, R. K., Ruff, K. M. & Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **32**, 102–112 (2015).
38. Sørensen, C. S. & Kjaergaard, M. Measuring Effective Concentrations Enforced by Intrinsically Disordered Linkers. *Methods Mol. Biol.* **2141**, 505–518 (2020).
39. Möglich, A., Joder, K. & Kiefhaber, T. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

formation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12394–12399 (2006).

40. Evers, T. H., Van Dongen, E. M. W. M., Faesen, A. C., Meijer, E. W. & Merkx, M. Quantitative understanding of the energy transfer between fluorescent proteins connected via flexible peptide linkers. *Biochemistry* **45**, 13183–13192 (2006).
41. Ruff, K. M. & Holehouse, A. S. SAXS versus FRET: A Matter of Heterogeneity? *Biophysical journal* vol. 113 971–973 (2017).
42. Song, J., Li, J. & Chan, H. S. Small-Angle X-ray Scattering Signatures of Conformational Heterogeneity and Homogeneity of Disordered Protein Ensembles. *J. Phys. Chem. B* **125**, 6451–6478 (2021).
43. Song, J., Gomes, G.-N., Shi, T., Gradinaru, C. C. & Chan, H. S. Conformational Heterogeneity and FRET Data Interpretation for Dimensions of Unfolded Proteins. *Biophysical Journal* vol. 113 1012–1024 (2017).
44. Fuertes, G. *et al.* Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proceedings of the National Academy of Sciences* vol. 114 E6342–E6351 (2017).
45. Martin, E. W., Hopkins, J. B. & Mittag, T. Small-angle X-ray scattering experiments of monodisperse intrinsically disordered protein samples close to the solubility limit. *Methods Enzymol.* **646**, 185–222 (2021).
46. Flory, P. J. *Principles of Polymer Chemistry*. (Cornell University Press, 1953).
47. Gomes, G.-N. W. *et al.* Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET. *J. Am. Chem. Soc.* **142**, 15697–15710 (2020).
48. Sukenik, S., Sapir, L., Gilman-Politi, R. & Harries, D. Diversity in the mechanisms of cosolute action on biomolecular processes. *Faraday Discuss.* **160**, 225–37; discussion 311–27 (2013).
49. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. & Pappu, R. V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **112**, 16–21 (2017).
50. Rogers, J. M., Steward, A. & Clarke, J. Folding and binding of an intrinsically disordered protein: Fast, but not 'diffusion-limited'. *J. Am. Chem. Soc.* **135**, 1415–1422 (2013).
51. Harmon, T. S. *et al.* GADIS: Algorithm for designing sequences to achieve target secondary structure profiles of intrinsically disordered proteins. *Protein Eng. Des. Sel.* **29**, 339–346 (2016).
52. König, I. *et al.* Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat. Methods* **12**, 773–779 (2015).
53. Sukenik, S., Ren, P. & Gruebele, M. Weak protein–protein interactions in live cells are quantified by cell-volume modulation. *Proceedings of the National Academy of Sciences* 201700818 (2017) doi:10.1073/pnas.1700818114.
54. Sukenik, S., Salam, M., Wang, Y. & Gruebele, M. In-Cell Titration of Small Solutes Controls Protein Stability and Aggregation. *J. Am. Chem. Soc.* **140**, 10497–10503 (2018).
55. Son, S. *et al.* Direct observation of mammalian cell growth and size regulation. *Nat. Methods* **9**, 910–912 (2012).
56. Schuler, B., König, I., Soranno, A. & Nettels, D. Impact of in-cell and in-vitro crowding on the conformations and dynamics of an intrinsically disordered protein. *Angew. Chem. Int.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ed Engl. (2021) doi:10.1002/anie.202016804.

57. Soranno, A. *et al.* Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4874–4879 (2014).
58. Martin, E. W. *et al.* Interplay of folded domains and the disordered low-complexity domain in mediating hnRNPA1 phase separation. *Nucleic Acids Res.* **49**, 2931–2945 (2021).
59. Mittal, A., Holehouse, A. S., Cohan, M. C. & Pappu, R. V. Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins. *J. Mol. Biol.* (2018) doi:10.1016/j.jmb.2018.05.012.
60. Zosel, F., Soranno, A., Buholzer, K. J., Nettels, D. & Schuler, B. Depletion interactions modulate the binding between disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 13480–13489 (2020).
61. Zamboni, N., Saghatelian, A. & Patti, G. J. Defining the metabolome: size, flux, and regulation. *Mol. Cell* **58**, 699–706 (2015).
62. Kim, D.-H. *et al.* Volume regulation and shape bifurcation in the cell nucleus. *J. Cell Sci.* **128**, 3375–3385 (2015).
63. Banks, A., Qin, S., Weiss, K. L., Stanley, C. B. & Zhou, H. X. Intrinsically Disordered Protein Exhibits Both Compaction and Expansion under Macromolecular Crowding. *Biophys. J.* **114**, 1067–1079 (2018).
64. Salvi, N., Abyzov, A. & Blackledge, M. Solvent-dependent segmental dynamics in intrinsically disordered proteins. *Sci Adv* **5**, eaax2348 (2019).
65. Schnatwinkel, J. & Herrmann, C. The interaction strength of an intrinsically disordered protein domain with its binding partner is little affected by very different cosolutes. *Phys. Chem. Chem. Phys.* (2020) doi:10.1039/D0CP03040F.
66. Goldenberg, D. P. & Argyle, B. Minimal effects of macromolecular crowding on an intrinsically disordered protein: A small-angle neutron scattering study. *Biophys. J.* **106**, 905–914 (2014).
67. Cuevas-Velazquez, C. L. *et al.* Intrinsically disordered protein biosensor tracks the physical-chemical effects of osmotic stress on cells. *Nat. Commun.* **12**, 5438 (2021).
68. Brodsky, S. *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol. Cell* (2020) doi:10.1016/j.molcel.2020.05.032.
69. Ignacio Gutiérrez, J. *et al.* SWI/SNF senses carbon starvation with a pH-sensitive low complexity sequence. *bioRxiv* 2021.03.03.433592 (2021) doi:10.1101/2021.03.03.433592.
70. Ortmayr, K., Dubuis, S. & Zampieri, M. Metabolic profiling of cancer cells reveals genome-wide crosstalk between transcriptional regulators and metabolism. *Nat. Commun.* **10**, 1841 (2019).
71. Sanchez, E. L. & Lagunoff, M. Viral activation of cellular metabolism. *Virology* **479-480**, 609–618 (2015).
72. Clavel, D. *et al.* Structural analysis of the bright monomeric yellow-green fluorescent protein mNeonGreen obtained by directed evolution. *Acta Crystallogr D Struct Biol* **72**, 1298–1307 (2016).
73. Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods* **10**, 407–409 (2013).
74. Vitalis, A. & Pappu, R. V. ABSINTH: a new continuum solvation model for simulations of polypeptides in aqueous solutions. *J. Comput. Chem.* **30**, 673–699 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.24.469609; this version posted November 25, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

75. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).
76. Peran, I. *et al.* Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12301–12310 (2019).
77. Mittal, A., Das, R., Vitalis, A. & Pappu, R. ABSINTH Implicit Solvation Model and Force Field Paradigm for Use in Simulations of Intrinsically Disordered Proteins. *Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods* 181 (2014).
78. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
79. Cranfill, P. J. *et al.* Quantitative assessment of fluorescent proteins. *Nat. Methods* **13**, 557–562 (2016).
80. Lambert, T. J. FPbase: a community-editable fluorescent protein database. *Nat. Methods* **16**, 277–278 (2019).
81. Nielsen, S. S. *et al.* BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *Journal of Applied Crystallography* vol. 42 959–964 (2009).
82. Hopkins, J. B., Gillilan, R. E. & Skou, S. : improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **50**, 1545–1553 (2017).
83. Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image Processing with ImageJ. *Biophotonics International* **11**, 36–42 (2004).