Commonly-used FRET fluorophores promote collapse of an otherwise disordered protein

Joshua A Riback[a], Micayla A Bowman[b], Adam M Zmyslowski[c], Kevin W Plaxco[d], Patricia L Clark[b]*, Tobin R Sosnick[c,e]*

[a]Graduate Program in Biophysical Sciences, University of Chicago, Chicago, IL 60637.

[b]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556.

[c]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637.

[d]Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106

[e]Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637.

Corresponding authors:
Tobin R. Sosnick
Department of Biochemistry and Molecular Biology
Institute for Biophysical Dynamics
University of Chicago
929 E. 57th Street
Chicago, IL 60637
trsosnic@uchicago.edu
773-218-5950

Patricia L. Clark
Department of Chemistry and Biochemistry
University of Notre Dame
251 Nieuwland
Notre Dame, IN 46556
Pclark1@nd.edu
574-631-8353

Keywords: IDP, SAXS, FRET, unfolded state

ABSTRACT: The dimensions that unfolded and intrinsically disordered proteins (IDPs) adopt at low or no denaturant remains controversial. We recently developed an innovative analysis procedure for small-angle X-ray scattering (SAXS) profiles and found that even relatively hydrophobic IDPs remain nearly as expanded as the chemically denatured ensemble, rendering them significantly more expanded than is inferred from many fluorescence resonance energy transfer (FRET) studies. Here we show that fluorphores typical of those added to IDPs for FRET studies contribute to this discrepancy. Specifically, we find that labeling a highly expanded IDP with Alexa488 causes its ensemble to contract significantly. We also tested the recent suggestion that FRET and SAXS results can be reconciled if, for unfolded proteins (and as opposed to the case for ideal random flight homopolymers), the radius of gyration (R_g) can vary independently from the chain's end-to-end distance (R_{ee}). Our analysis indicates, however, that SAXS is able to accurately extract R_g, ν and R_{ee} even for heteropolymeric, protein-like sequences. From these studies we conclude that mild chain contraction and fluorophore-based interactions at lower denaturant concentrations, along with improved analysis procedures for both SAXS and FRET, can explain the preponderance of existing data regarding the nature of polypeptide chains unfolded in the absence of denaturant.

Significance Statement:

Proteins can adopt a disordered ensemble, either prior to folding or as a part of their function. Simulations and fluorescence resonance energy transfer (FRET) studies often describe these disordered conformations as more compact than the fully random-coil state, whereas small-angle X-ray scattering studies (SAXS) indicate an expanded ensemble closely approximating the dimensions expected for the random coil. Resolving this discrepancy will enable more accurate predictions of protein folding and function. Here we reconcile these views by showing that the addition of common FRET fluorophores reduces the apparent dimensions of a disordered protein. Detailed analysis of both techniques, along with accounting for a moderate amount of fluorophore-induced contraction, demonstrates that disordered and unfolded proteins often remain well solvated and largely expanded in the absence of denaturant, properties that presumably minimize misfolding and aggregation.

Protein disorder is an essential component of diverse cellular processes (1-4). Unlike well-folded proteins, which populate a well-defined functional state, intrinsically disordered proteins (IDPs) and regions (IDRs) sample a broad ensemble of rapidly interconverting conformations (3-9) with biases that are poorly understood and difficult to measure. Of particular interest is the extent to which IDPs undergo compaction under physiological conditions (i.e., in the absence of denaturants). Such compaction would have broad implications for our understanding of protein folding, interactions and stability as well as the action of denaturants. Moreover, understanding the extent of collapse in disordered ensembles has profound implications for the development of realistic simulations of protein folding and interpretation of SAXS and FRET measurements (10, 11).

Our current understanding of the physiochemical principles that underlie whether a given polypeptide chain will fold, adopt a collapsed but disordered ensemble, or behave as an expanded, fully-solvated self-avoiding random walk (SARW) is insufficient to explain the existing data. Most of this understanding is derived from studies of proteins unfolded by high concentrations of chemical denaturants such as urea and guanidine hydrochloride (Gdn). Under these conditions the consensus is that proteins behave as SARWs, corresponding to a Flory exponent (ν) of 0.60 in the relationship $R_g \propto N^{\nu}$ (where R_g = radius of gyration and N = polypeptide chain length). Consensus is lacking, in contrast, regarding the behavior of disordered polypeptide chains at lower or no denaturant. Specifically, while numerous FRET (18-21, 25-35) and computational studies (18-24, 36) have argued that the expanded, disordered ensemble seen at high denaturant collapses at low or no denaturant ($\nu < 0.5$) (12-24), almost equally numerous SAXS studies report no or only minor collapse under these same conditions (11, 37-42).

A variety of recent studies have attempted to reconcile this widely-recognized and seemingly important discrepancy (**Fig. 1A**). For example, the application of more realistic simulations and analytical models

has reduced the denaturant dependence of FRET-derived distances (**Fig. 1A, bottom**) (41, 43-45). Likewise, improved analysis of SAXS data provides evidence for minor contraction when the guanidine hydrochloride (Gdn) concentration falls below 2 M (**Fig. 1A, bottom**), further reducing the apparent disagreement between the two methods (46, 47). Significant discrepancies nevertheless persist between the two techniques, even when the same procedures are used to analyze a single protein by both SAXS and FRET (**Figs. 1A-B**, **S1-S2**; **Movie S1**).

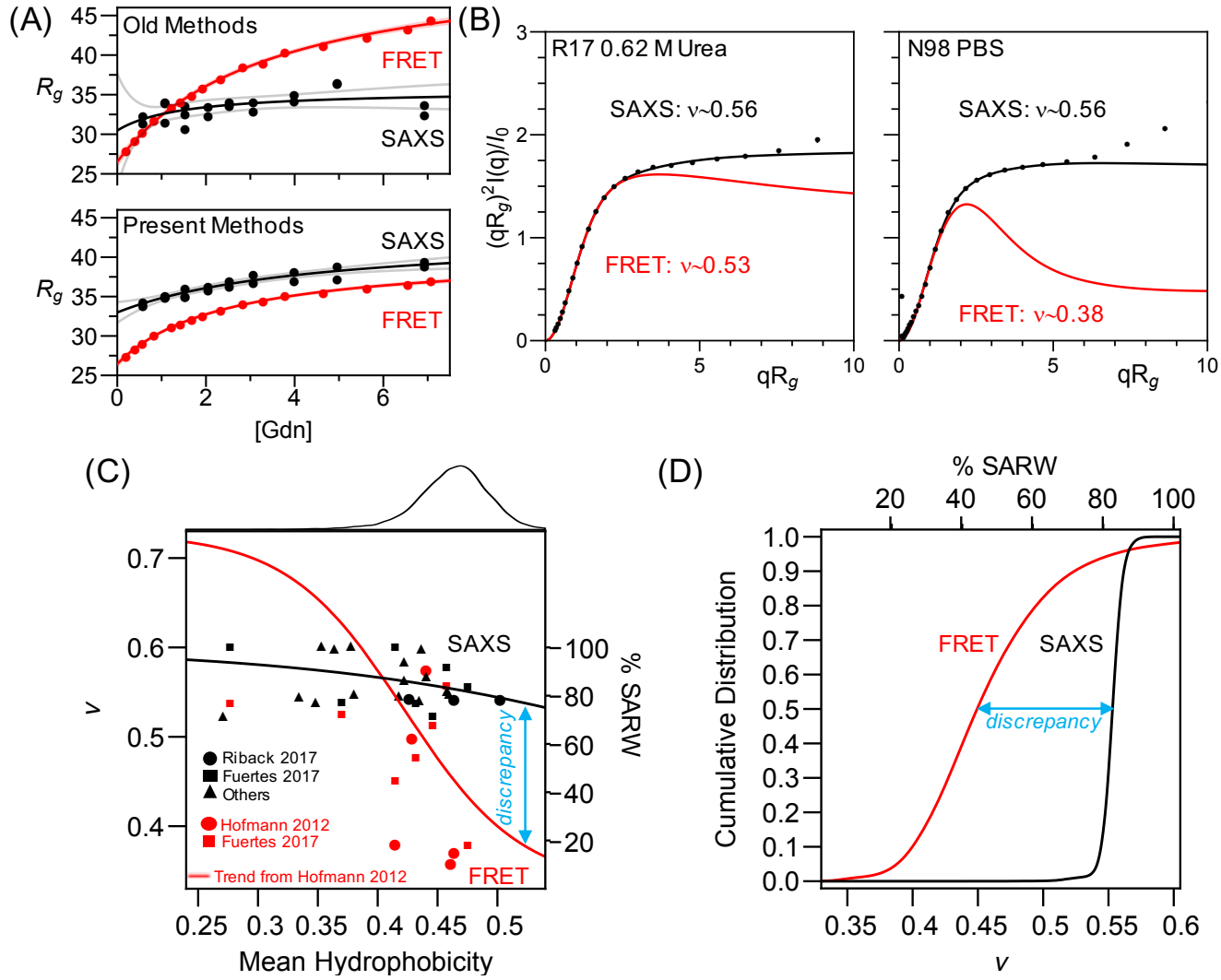


Fig. 1. New analysis reduces, but does not eliminate, the discrepancy between SAXS and FRET-derived measures of IDP dimensions. In all panels, SAXS results are shown in black; FRET results are in red. **(A)** R17 SAXS and FRET data (from (44)). Top, comparison of results obtained when FRET data is fit assuming a Gaussian chain and SAXS data is fit using the Guinier approximation. Bottom, SAXS and FRET data fit using our new MFF analysis method (46). Black line is best fit hyperbolic trend line; grey lines are 95% confidence intervals. **(B)** SAXS profiles for R17 (left, taken from (44)) and N98 (right, taken from (43)) fit with the MFF are significantly different than the expected behavior using values of ν taken from similar analysis of FRET data. **(C)** Trends of hydrophobicity (Kyte-Doolittle) versus ν in water for SAXS data of foldable protein sequences fit by applying the MFF to published data (43, 46, 52-67). Also shown are results from FRET studies calculated as in (31) for published data (31, 43). Red trend line for FRET results is from (31). Black trend line is best fit to SAXS results shown. Top, histogram of hydrophobicity of representative proteins in the PDB (dataset from (46)). **(D)** Cumulative distributions of ν for the representative proteins from the PDB, inferred from the trend lines shown in (C).

To comprehensively compare results from recent SAXS and FRET analyses of IDPs, we collected published datasets for a variety of disordered proteins (**Fig. 1C-D**; **Table S3**). When analyzed using our simulations and molecular form factor (MFF), SAXS studies consistently find $\nu > 0.53$ (mean = 0.55) whereas ν derived from FRET studies typically falls below the random walk value of 0.50 (mean = 0.46). This discrepancy of 0.09 is substantial given that the entire range of ν varies only from 0.6 (for a SARW) through 0.5 (for a non-self-avoiding random walk) to 0.33 (for a perfect sphere; it is somewhat higher for non-spherical compact states). In general, SAXS results suggest that the conformational ensembles of a majority of unfolded proteins and IDPs (of protein-like composition) are more expanded than a non-self-avoiding random walk, whereas FRET suggests otherwise (**Fig 1D**).

The above and other, similar results have led us and others to consider the factors that might contribute to the persistent discrepancy between SAXS- and FRET-based views of IDPs (36, 40, 43, 48). One alternative, herein denoted the "heteropolymer-decoupling hypothesis," posits that the heteropolymeric nature of proteins leads to variation of the normally fixed relationship between R_g and the polypeptide chain end-to-end distance (R_{ee}). Specifically, for a homopolymer adopting a SARW the ratio $G = (R_{ee}/R_g)^2$ is expected to be fixed at a value of 6.3 irrespective of the polymer's length. Recent simulations, however, suggest that, unlike the case for homopolymers, this ratio can vary significantly for heteropolymers (36, 40, 43, 44), with this "decoupling" offering a possible explanation for the discrepancy between SAXS (which measures R_g) and FRET (which measures R_{ee}). In contrast, a second hypothesis, herein denoted the "fluorophore-interaction hypothesis," suggests that, in the absence of denaturant, the FRET fluorophores interact with each other and/or with the polypeptide chain, causing the conformational ensemble of dye-modified constructs to contract artifactually (10, 46, 48, 49).

To address these hypotheses we have used SAXS to characterize the radius of gyration of fluorophore-modified IDPs in the presence and absence of dye modification. Doing so we find that labeling with fluorophores commonly used for FRET studies alters the conformational ensemble populated in the absence of denaturant, decreasing its SAXS-measured dimensions by 10-20%. When coupled with improved analysis procedures employing realistic ensembles for both SAXS and FRET, this contraction is sufficient to bring results from SAXS and FRET studies into agreement. In parallel, we present SAXS measurements on polyethylene glycol (PEG), confirming prior reports that the addition of fluorophores likewise compacts this otherwise SARW polymer (10), a finding that was recently questioned (43). We also show that SAXS can extract R_g, ν and R_{ee} with an accuracy of better than 3% when analyzed using a new MFF developed for heteropolymers. These simulations are accurate enough to reproduce scattering data without the need to select only a sub-ensemble of conformations, as commonly used in other data fitting procedures. Finally, we demonstrate the extent that one can use small deviations from ideality in SAXS data to infer biases within the heteropolymer conformational ensemble.

Results

Fluorophore-labeling induces collapse of a highly expanded ensemble. To directly test the fluorophore-interaction hypothesis we have measured SAXS profiles of an unmodified IDP and the same IDP when site-specifically modified with one or two copies of the commonly employed FRET dye Alexa488. We chose this dye because it is relatively hydrophilic (44), and thus considered less likely to form interactions that would alter the conformational ensemble than any of the other commonly used FRET fluorophores. As our test protein we used PNt, a well-behaved IDP comprising the amino terminal 334 residues of pertactin (50). To produce the mono-dye-modified protein we introduced a cysteine residue at position 117 and modified it with the appropriate thiol-reactive Alexa488 variant (PNtC-Alexa488). To produce a dye-pair-modified protein we introduced cysteines at positions 29 and 117 (PNtCC-Alexa488). Alkylation was used to produce dye-free constructs (PNtC-Alkd and PNtCC-Alkd) that we used as controls, in addition to the unmodified parent protein (PNt).

The addition of Alexa488 reduces the SAXS-measured dimensions of the IDP under both aqueous and intermediate denaturant conditions (**Fig. 2A**, **Table S1**). Specifically, R_g and ν decrease nearly twice as much for PNtCC-Alexa488 than for either PNtCC-Alkd or PNt (**Fig. 2B**; **Table S1**). These data indicate that labeling with Alexa488 leads to increased contraction of the PNt conformational ensemble, implying that stabilization of fluorophore-mediated interactions within the IDP conformational ensemble. Of note, whereas the unlabeled protein in 2 M Gdn is in a good solvent condition (ν > 0.5), with protein-protein interactions weaker than protein-solvent interactions, fluorophore-labeling leads to measurable intramolecular interactions even at this relatively high denaturant concentration. The magnitude of this denaturant-dependent chain expansion is qualitatively similar to that observed by FRET for a variety of other proteins, consistent with a common origin (**Fig. 1B**) (43, 44). We also observed a fluorophore-dependent decrease in average R_g and ν for the single-labeled construct PNtC-Alexa488 (**Fig. 2**). This result indicates that the PNt conformational ensemble is affected by fluorophore-protein interactions, not only fluorophore-fluorophore interactions. Of note, this contraction occurs despite the fact that the steady-state fluorescence anisotropy of PNtCC-Alexa488 is, at 0.11 and 0.08 in 0 and 2 M denaturant, respectively (**Table S2**), below the threshold usually used to indicate free rotation of protein-attached dyes (43, 51).

It thus appears that addition of even relatively hydrophilic fluorophores commonly employed for FRET measurements can significantly influence the dimensions of disordered polypeptide chains (43, 44, 51).

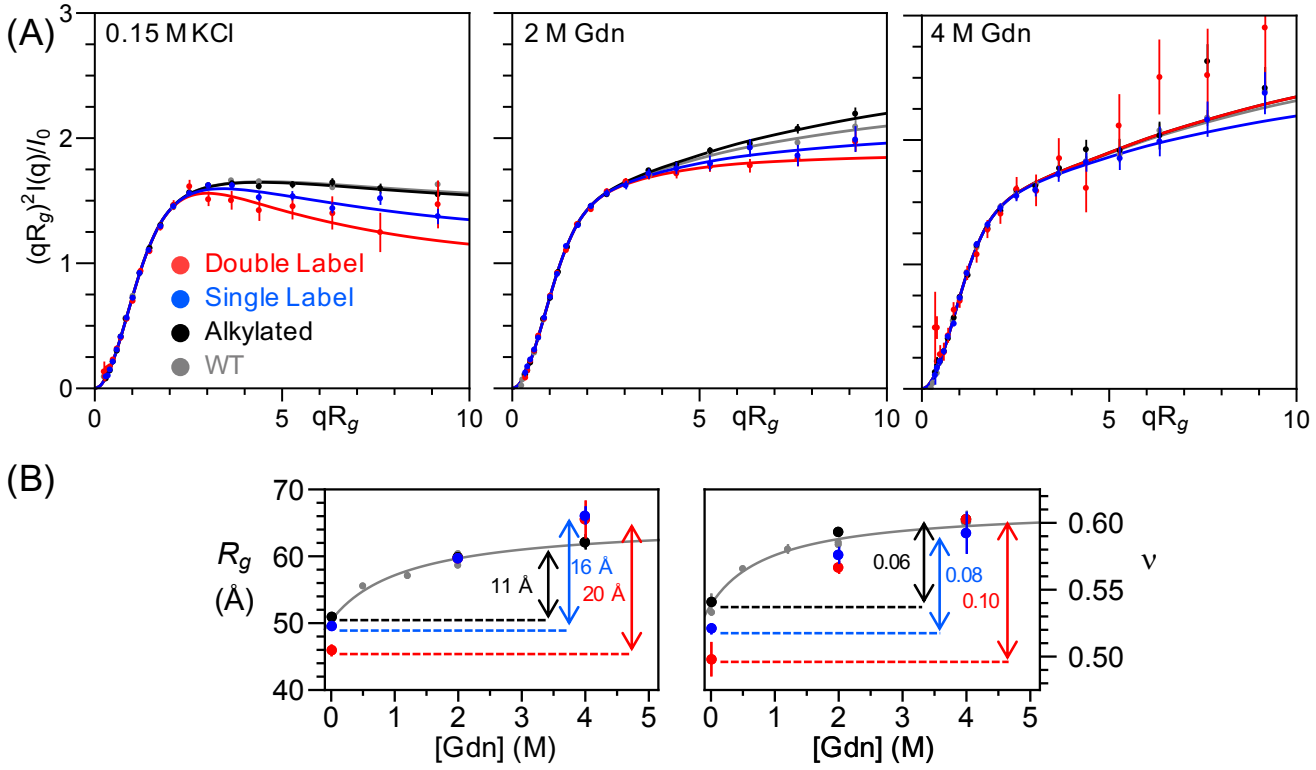


Fig. 2. The addition of Alexa488 alters the scattering of PNt. (A) Dimensionless Kratky plots of wild type PNt (gray), PNtCC-Alkylated (black), PNtC-Alexa488 (single label, red), and PNtCC-Alexa488 (double label blue) in 0.15 M KCl, 2 M Gdn, and 4 M Gdn. Results for alkylated PNt are indistinguishable from wild type PNt but there are significant differences for PNt labeled with one (blue) or two (red) Alexa488 fluorophores. (B) R_g and ν as a function of Gdn concentration. Gray curves are reproduced from previous analysis of wild type PNt (46).

PEG dimensions are independent of polymer concentration. In an earlier study we reported that addition of Alexa488/594 to PEG resulted a denaturant-dependent change in FRET (10), similar to that seen in FRET measurements of unfolded proteins. No contraction was observed, however, when the equivalent unlabeled polymer was studied using small angle neutron scattering. It has been proposed that the high (3 mM) concentrations of PEG used in this scattering study masked what would otherwise be denaturant-dependent changes in R_g (43). To test this, we measured SAXS profiles over a range of PEG and denaturant concentrations and found no evidence for a significant change in polymer dimensions (**Fig. 3**). Likewise, under all conditions we observe a Flory exponent of 0.60, further confirming that PEG behaves as a SARW independent of denaturant concentration. The fluorophore-interaction hypothesis thus remains the simplest interpretation of the denaturant-dependent changes in FRET observed for fluorophore-labeled PEG (10).

Testing the heteropolymer-decoupling hypothesis. Taken together, the above observations indicate that fluorophores added to an IDP lead to significant contraction of its conformational ensemble, contributing to the different conclusions drawn from prior SAXS and FRET studies. These observations, however, do not rule out the possibility that heteropolymer-decoupling also contributes to the SAXS-FRET discrepancy. To investigate how amino acid sequence composition and pattern can alter the SAXS profile and test our ability to extract information from such deviations we used our Cβ-level polypeptide chain simulations (46) to simulate the scattering for unfolded ensembles of 50 protein sequences of 250-650 residues randomly chosen from the PDB. We treated each sequence as a binary hydrophobic/polar (HP) chain, where the only favorable interactions are between Cβ atoms of aliphatic and/or aromatic residues. For each of the 50 sequences, 30 different Cβ interaction strengths were used. These simulations yielded a range of deviations from G(ν) obtained from homopolymer simulations (**Fig. S4A**). From the 1500 resulting ensembles, we determined the R_g, ν, and R_{ee} both directly from the atomic coordinates of the simulated ensemble and by fitting the hydrated SAXS profile (with added realistic random errors) of each ensemble using our MFF. The inferred end-to-end distance (R_{ee}^{inf}) was determined using the relationship $R_{ee}^{inf}(R_g, \nu)=G(\nu)^{1/2}*R_g$, where G(ν) was obtained from homopolymer simulations.

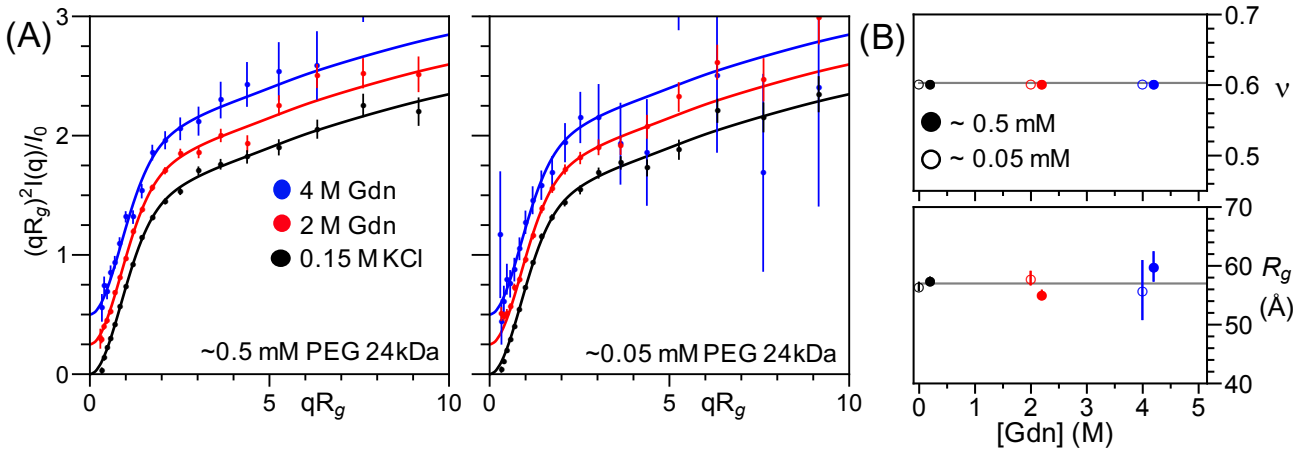


Fig. 3. The SAXS profiles of PEG are denaturant independent. (A) Dimensionless Kratky plots of 24 kDa PEG at 0.5 mM and 0.05 mM in 0.15 M KCl, 2 M Gdn, and 4 M Gdn. The normalized scattering profile of 24 kDa PEG is unchanged from 0-4 M Gdn over a broad range of PEG concentrations. Scattering profiles have been offset vertically for clarity. (B) R_g and ν as a function of Gdn concentration for 0.5 mM and 0.05 mM PEG. Open and closed points are offset horizontally for clarity.

Compared to the true values calculated from the atomic coordinates, fits obtained using our MFF yield values of R_g, ν, and R_{ee} with a mean absolute deviation of only 1.3 Å, 0.011 and 4.2 Å, respectively,

representing a 3%, 2% and 4% mean absolute error (**Fig S3**). The largest deviations are observed for more compact structures; for more extended conformations with ν >0.54, there is ~2% error.

To further reduce the small error associated with the application of our MFF derived from homopolymers to the scattering of heteropolymers, we generated a new molecular form factor, MFF_{het}, using the heteropolymer simulations described above. Application of this slightly modified MFF lowers errors in fitted R_g, ν, and R_{ee} to 0.5 Å, 0.005 and 2.7 Å, respectively, representing 1%, 1% and 2% mean absolute error (**Fig 4A-C**). These results demonstrate that SAXS analysis returns accurate values of R_{ee}, in addition to R_g and ν, even for disordered heteropolymers.

Measuring deviations from ideality in heteropolymers. MFF_{het} accurately captures the overall dimensions of disordered heteropolymers. Nevertheless, small but measureable deviations are observed for proteins in our test set with less well-mixed HP patterns (**Fig. S4**). These differences can be seen in the

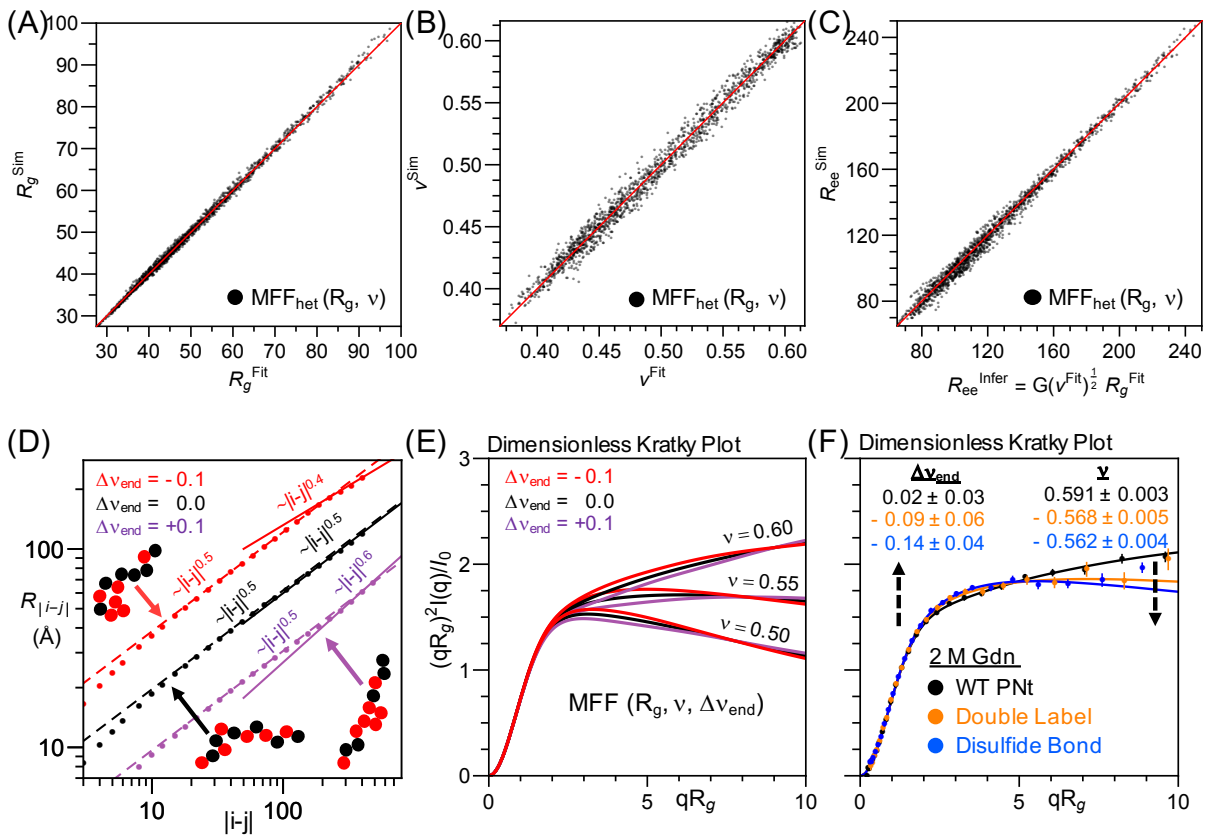


Fig. 4. **Fitting simulated SAXS data on realistic heteropolymer sequences demonstrates SAXS profiles are a robust measure of R_g, ν, and R_{ee} while also informing on the degree of heterogeneity.** **(A-C)** Comparison of R_g, ν, and R_{ee} calculated from coordinates of HP-model simulations versus fit with our $MFF_{het}(R_g, \nu)$ to SAXS profiles with randomly added experimental errors. **D)** Deviations in ν_{ends} are observed for heteropolymers with less well-mixed HP patterns (obtained by a fit to the slope of the dependence of the intra-chain distance, $R_{|i-j|}$, on sequence separation, |i–j| where |i–j| > N/2). **E)** Effects of $\Delta\nu_{end}$ at different values of ν. **F)** Experimental data fit to $MFF(R_g,\nu, \Delta\nu_{end})$ demonstrates fluorophore labeling and loop formation via disulfide bonds in PNt induces significant and measurable deviations.

intra-molecular distance distribution plot, where the slope at separation distances |i-j|>N/2 can be different than the average slope, which defines the global ν value **(Fig. 4D)**. We define change in slope as

$\Delta\nu_{end}$ (**Fig. 4D**). Negative values of $\Delta\nu_{end}$ correlate with more hydrophobic residues at the ends of the polypeptide sequences (**Figs. 4D, S4C**) and with deviations in G(ν) (**Fig. S4A**) (R^2~0.84). The SAXS profile is most sensitive to $\Delta\nu_{end}$ at low qR_g (**Fig. 4E**).

To extract this extra information from the SAXS data we generated a more general three-parameter form factor, MFF(R_g, ν, $\Delta\nu_{end}$) (**Fig. 4E-F**; **Movie S2**). To demonstrate its ability to yield useful information, we fit data from PNt, PNtCC-Alexa488 and a circularized (disulfide-bonded) PNtCC at 2 M Gdn (**Fig. 4F**). $\Delta\nu_{end}$ decreases from ~0 for PNt to ~ -0.1 for PNtCC-Alexa488 and PNtCC, consistent with the increase in interactions at the N-terminus of the chain. These data demonstrate that for disordered polymers, SAXS is sensitive to heteropolymer deviations (**Fig. 4E-F**) while still able to measure R_g and ν accurately (**Fig. 4A-C**).

Discussion

Labeling with Alexa488, a fluorophore commonly used to support FRET measurements, can alter the conformational ensemble of a disordered protein, decreasing R_g and ν even when the fluorescence anisotropy is low relative to accepted limits for free dye rotation (43, 51). In combination with prior studies (10), similar conclusions can be inferred for PEG, a known SARW. These findings, along with our prior result that disordered chains undergo a mild expansion in denaturants (46) and improved methods for extracting R_g values from FRET data, provide a sufficient framework for resolving discrepancies between SAXS and FRET on the dimensions of disordered proteins.

Consistent with our findings of fluorophore-induced effects, others have found that molecular dimensions inferred by FRET can be dependent on the fluorophore pair used, with more hydrophobic fluorophores suggesting a more collapsed state (44). All-atom molecular dynamics simulations with a Alexa488/594 fluorophore pair, for example, resulted in a 10% contraction of an IDP even in 1 M urea (68). Likewise, a recent study found that smFRET signals from both DNA and PEG, which are often referred to as "spectroscopic rulers," are dependent on solvent conditions under which the dimensions of the chains were expected to be invariant (49). In apparent disagreement with our data, however, Fuertes *et al.* (43) conducted SAXS measurements on five IDPs with and without Alexa488/594 and concluded that, *on average,* the alterations seen upon the addition of fluorophores were minimal. When considered for each protein separately, however, the differences appear significant relative to the narrow range of all possible values of ν. Specifically, for the five proteins characterized in that study, $\nu_{unlabel}$-ν_{label} =0.08, 0.03, 0.03, -0.02, -0.04 (or 0.09, 0.06, 0.03, -0.02, -0.08 when analyzed using our procedures). That is, more than half of these IDPs exhibit a fluorophore-induced contraction similar in magnitude to the contraction we observe for fluorophore-labeled PNt in water. Together, these data suggest a consistent picture of fluorophore-induced contraction contributing to differences in the magnitude and denaturant dependence of R_g inferred from SAXS and FRET.

The other factor that has been suggested to contribute to the discrepancy between SAXS and FRET results is deviations from the proportional relationship between R_g and R_{ee} that arise when moving from homopolymers to heteropolymers (43). Underlying this view is the observation that, if one reweighs the ensemble (i.e., calculates R_g using only a subset of conformations), many possible values of R_{ee} are consistent with any given R_g (and vice-versa). Rather than selecting a sub-ensemble to fit the data, however, we have instead taken an alternative approach (46). We generate physically plausible ensembles at the outset and examine whether they fit the data in their entirety. We find that the MFF derived from our ensembles accurately matches the entire scattering profile (rather than just the R_g), which provides strong support for our procedure. Since we can determine the values of R_g and R_{ee} directly from the underlying ensembles, we have a procedure to obtain these two parameters by fitting the SAXS data with our MFF.

This MFF is imperfect in the sense that slightly different ensembles can be fit using the same R_g and ν parameters. But the error is very low for these two parameters relative to their true values (**Fig. 4A-C**). Inclusion of heteropolymer effects does not change this conclusion. From this result, we conclude that

SAXS is well suited to extract both R_g and R_{ee} for disordered heteropolymers, while circumventing potential artifacts due to fluorophore interactions with polypeptide chains. This conclusion does not negate the potential of FRET to measure dynamics, binding and conformational changes; it does, however, reinforce that caution must be exercised when employing FRET to infer quantitative distances in the original, unlabeled biomolecule.

Nearly a dozen IDP SAXS datasets reported here and previously (46) have been shown to fit well to our general MFF (**Tables S1, S3**). This finding suggests that the interactions that drive chain contraction are spread along protein sequences. Water-soluble, well-folded protein sequences tend to be well-mixed heteropolymers, with relatively small stretches of consecutive hydrophobic residues (69). These well-mixed sequences tend to behave as homopolymers when measured by global, low resolution methods such as SAXS. Indeed, we have demonstrated that, with sufficient data quality, poorly mixed sequences can be identified by their deviation from our MFF (**Fig. 4D-F**). Larger deviations can occur for some IDPs, especially those with partial folding, unusual sequence patterning (e.g., block copolymers) and/or under crowded conditions that may serve specific functions (70, 71).

These results and analysis presented here demonstrate that water is a good solvent for many foldable protein sequences, rendering them expanded with $\nu \geq 0.54$. This property is likely beneficial to the cell as it helps to reduce misfolding and unwanted associations, while simultaneously facilitating protein synthesis and transport. Moreover, many proteins fold (in both a thermodynamic and kinetic sense) even in the presence of modest amounts of denaturants – and some even in 6 M Gdn (42, 72), where $\nu \sim 0.6$. Hence, folding is a robust process that does not critically depend on solvent quality as long as the native protein structure is stable relative to the unfolded ensemble.

Materials and Methods

Protein purification. PNtCC and PNtC were expressed in *E. coli* BL21(DE3)pLysS and purified from inclusion bodies as described previously (46, 50, 73), with the following modifications. After inclusion body solubilization, PNt constructs were refolded in 50 mM Tris pH 7.2 with 50 mM ß-mercaptoethanol (ßME). Prior to the final size exclusion chromatography step, 20 mM ßME was added to the protein stock solution.

Alkylation. Purified PNtCC or PNtC (70 uM) in 50 mM Tris pH 8, 5 mM EDTA was reduced with 5 mM TCEP for 30 min at room temperature while stirring. Alkylation was initiated by addition of 10 mM iodoacetamide in water. PNt constructs were incubated for 30 min at room temperature while stirring in the dark. The alkylation reaction was quenched with the addition of 20 mM fresh DTT. Excess reagents were removed by size exclusion chromatography (Superdex 16/60 S200). Alkylation efficiency (100%) was determined by mass spectrometry (MS).

Alexa488 labeling. PNt constructs were reduced as above. Alexa488 C5 maleimide (ThermoFisher Scientific) was resuspended in DMSO to 5 mg/ml and added to the reduced protein dropwise while stirring, to a final ratio of 5:1 fluorophore:cysteine. The reaction proceeded overnight at 4°C while stirring in the dark. Free fluorophore was separated from fluorophore-labeled protein by size-exclusion chromatography (Superdex 16/60 S200), protected from light at all times. Labeling efficiency (100% for PNtCC; >50% for PNtC) was determined by mass spectrometry.

Steady-state anisotropy. Steady-state anisotropy measurements were performed on a QM-6 T-format fluorometer (Horiba) at room temperature in 50 mM Tris pH 7.5 and 2 M Gdn where indicated. Samples containing 1 uM Alexa-488 were excited with 494 nm vertically polarized light. Anisotropy (r) was calculated using the emission at 516 nm by the equation

$$r = \frac{I_{\parallel} - GI_{\perp}}{I_{\parallel} - 2GI_{\perp}}$$

Where G is the instrument correction factor. G was calculated from the emission of free-Alexa488 at 516 nm after excitation with 494 nm horizontally polarized light by the equation

$$G = \frac{I_{HV}}{I_{HH}}$$

SAXS. Data were collected at the BioCAT beamline at the Advanced Photon Source (Argonne National Lab) using a GE Lifesciences Superdex 200 SEC column with the scattering presented as $q=4\pi \sin\theta/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength (1 Å).

Code for simulations and fitting scattering data. Code and associated files necessary to produce simulations and the analysis can be accessed at www. (to be added prior to publication). Additionally, our web-server http://sosnick.uchicago.edu/SAXSonIDPs is available for fitting SAXS data with our MFF(ν,R_g).

Simulations and generation of MFF. Calculations were performed at the University of Chicago Research Computing Center using a version of our *Upside* molecular dynamics program (**arXiv:1610.07277**) modified for Cβ-level interactions, as done previously (46).

Acknowledgements

We thank Srinivas Chakravarthy for assistance in the SAXS measurements and Matthew Champion for assistance with the mass spectrometry experiments. This work was supported by National Institutes of Health (NIH) Research Grant R01 GM055694 (TRS), the W. M. Keck Foundation (PLC) and NSF grants GRF DGE-1144082 (JAR) and MCB 1516959 (CR Matthews). Use of the Advanced Photon Source, an Office of Science User Facility, operated for the Department of Energy (DOE) Office of Science by Argonne National Laboratory, was supported by the DOE under Contract No. DEAC02- 06CH11357. This project was supported by the NIH (2P41RR008630-18 and 9 P41 GM103622-18).

References

1. van der Lee R, *et al*. (2014) Classification of intrinsically disordered regions and proteins. *Chem Rev* 114(13):6589-6631.
2. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, & Obradovic Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41(21):6573-6582.
3. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, & Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* 323(3):573-584.
4. Oates ME, *et al*. (2013) D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res* 41(Database issue):D508-516.
5. Ripaud L, *et al*. (2014) Overexpression of Q-rich prion-like proteins suppresses polyQ cytotoxicity and alters the polyQ interactome. *Proc Natl Acad Sci U S A* 111(51):18219-18224.
6. Marsh JA & Forman-Kay JD (2010) Sequence determinants of compaction in intrinsically disordered proteins. *Biophys J* 98(10):2383-2390.
7. Pappu RV, Wang X, Vitalis A, & Crick SL (2008) A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch Biochem Biophys* 469(1):132-141.
8. Mao AH, Lyle N, & Pappu RV (2013) Describing sequence-ensemble relationships for intrinsically disordered proteins. *The Biochemical journal* 449(2):307-318.
9. Uversky VN, Gillespie JR, & Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41(3):415-427.
10. Watkins HM, *et al*. (2015) Random coil negative control reproduces the discrepancy between scattering and FRET measurements of denatured protein dimensions. *Proc Natl Acad Sci U S A* 112(21):6631-6636.

11. Yoo TY, *et al*. (2012) Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state. *J. Mol. Biol*. 418(3-4):226-236.
12. Sadqi M, Lapidus LJ, & Munoz V (2003) How fast is protein hydrophobic collapse? *Proc. Natl. Acad. Sci. U S A* 100(21):12117-12122.
13. Haran G (2012) How, when and why proteins collapse: the relation to folding. *Curr Opin Struct Biol* 22(1):14-20.
14. Voelz VA, *et al*. (2012) Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. *J. Am. Chem. Soc*. 134(30):12565-12577.
15. Teufel DP, Johnson CM, Lum JK, & Neuweiler H (2011) Backbone-driven collapse in unfolded protein chains. *J Mol Biol* 409(2):250-262.
16. Montgomery Pettitt B (2013) The unsolved "solved-problem" of protein folding. *J Biomol Struct Dyn* 31(9):1024-1027.
17. Goluguri RR & Udgaonkar JB (2016) Microsecond rearrangements of hydrophobic clusters in an initially collapsed globule prime structure formation during the folding of a small protein. *J Mol Biol*.
18. O'Brien EP, Ziv G, Haran G, Brooks BR, & Thirumalai D (2008) Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci. U S A* 105(36):13403-13408.
19. Nettels D, *et al*. (2009) Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl. Acad. Sci. U S A* 106(49):20740-20745.
20. Merchant KA, Best RB, Louis JM, Gopich IV, & Eaton WA (2007) Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc Natl Acad Sci U S A* 104(5):1528-1533.
21. Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, & Pande VS (2010) Unfolded-state dynamics and structure of protein L characterized by simulation and experiment. *J. Am. Chem. Soc*. 132(13):4702-4709.
22. Bowman GR & Pande VS (2010) Protein folded states are kinetic hubs. *Proc Natl Acad Sci U S A* 107(24):10890-10895.
23. Lindorff-Larsen K, Piana S, Dror RO, & Shaw DE (2011) How fast-folding proteins fold. *Science* 334(6055):517-520.
24. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, & Shaw DE (2012) Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc*. 134(8):3787-3791.
25. Chung HS, Louis JM, & Eaton WA (2009) Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc Natl Acad Sci U S A* 106(29):11837-11844.
26. Nettels D, Gopich IV, Hoffmann A, & Schuler B (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. U S A* 104(8):2655-2660.
27. Sherman E & Haran G (2006) Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. U S A* 103(31):11539-11543.
28. Ziv G, Thirumalai D, & Haran G (2009) Collapse transition in proteins. *Phys. Chem. Chem. Phys*. 11(1):83-93.
29. Ziv G & Haran G (2009) Protein folding, protein collapse, and tanford's transfer model: lessons from single-molecule FRET. *J. Am. Chem. Soc*. 131(8):2942-2947.
30. Pugh SD, Gell C, Smith DA, Radford SE, & Brockwell DJ (2010) Single-molecule studies of the Im7 folding landscape. *J Mol Biol* 398(1):132-145.
31. Hofmann H, *et al*. (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U S A* 109(40):16155-16160.
32. Magg C, Kubelka J, Holtermann G, Haas E, & Schmid FX (2006) Specificity of the initial collapse in the folding of the cold shock protein. *J. Mol. Biol*. 360(5):1067-1080.

33. Waldauer SA, Bakajin O, & Lapidus LJ (2010) Extremely slow intramolecular diffusion in unfolded protein L. *Proc Natl Acad Sci U S A* 107(31):13713-13717.
34. Dasgupta A & Udgaonkar JB (2010) Evidence for Initial Non-specific Polypeptide Chain Collapse During the Refolding of the SH3 Domain of PI3 Kinase. *J. Mol. Biol.* 403(3):430-445.
35. Huang F, *et al.* (2009) Time-resolved fluorescence resonance energy transfer study shows a compact denatured state of the B domain of protein A. *Biochemistry* 48(15):3468-3476.
36. Reddy G & Thirumalai D (2017) Collapse Precedes Folding in Denaturant-Dependent Assembly of Ubiquitin. *J Phys Chem B* 121(5):995-1009.
37. Plaxco KW, Millett IS, Segel DJ, Doniach S, & Baker D (1999) Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nature Struct. Biol.* 6(6):554-556.
38. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, & Sosnick TR (2004) Early Collapse is not an Obligate Step in Protein Folding. *J. Mol. Biol.* 338(2):369-382.
39. Kathuria SV, *et al.* (2014) Microsecond barrier-limited chain collapse observed by time-resolved FRET and SAXS. *J Mol Biol* 426(9):1980-1994.
40. Song J, Gomes GN, Shi T, Gradinaru CC, & Chan HS (2017) Conformational Heterogeneity and FRET Data Interpretation for Dimensions of Unfolded Proteins. *Biophys J* 113(5):1012-1024.
41. Song JH, Gomes GN, Gradinaru CC, & Chan HS (2015) An Adequate Account of Excluded Volume Is Necessary To Infer Compactness and Asphericity of Disordered Proteins by Forster Resonance Energy Transfer. *Journal of Physical Chemistry B* 119(49):15191-15202.
42. Skinner JJ, *et al.* (2014) Benchmarking all-atom simulations using hydrogen exchange. *Proc Natl Acad Sci U S A* 111(45):15975-15980.
43. Fuertes G, *et al.* (2017) Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc Natl Acad Sci U S A.*
44. Borgia A, *et al.* (2016) Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *J Am Chem Soc* 138(36):11714-11726.
45. Zheng W, *et al.* (2018) Inferring properties of disordered chains from FRET transfer efficiencies. *J Chem Phys* 148(12):123329.
46. Riback JA, *et al.* (2017) Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* 358(6360):238-241.
47. Zheng W & Best RB (2018) An Extended Guinier Analysis for Intrinsically Disordered Proteins. *J Mol Biol* 430(16):2540-2553.
48. Takahashi S, Yoshida A, & Oikawa H (2018) Hypothesis: structural heterogeneity of the unfolded proteins originating from the coupling of the local clusters and the long-range distance distribution. *Biophys Rev* 10(2):363-373.
49. Qu S, *et al.* (2018) Solvent effect on FRET spectroscopic ruler. *J Chem Phys* 148(12):123331.
50. Renn JP, Junker M, Besingi RN, Braselmann E, & Clark PL (2012) ATP-independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. *Chem Biol* 19(2):287-296.
51. Aznauryan M, *et al.* (2016) Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc Natl Acad Sci U S A* 113(37):E5389-5398.
52. Gates ZP, *et al.* (2017) Perplexing cooperative folding and stability of a low-sequence complexity, polyproline 2 protein lacking a hydrophobic core. *Proc Natl Acad Sci U S A* 114(9):2241-2246.
53. Contreras-Martos S, *et al.* (2017) Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Sci Rep* 7(1):4676.
54. Martin EW, *et al.* (2016) Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J Am Chem Soc* 138(47):15323-15335.
55. Kolonko M, *et al.* (2016) Intrinsic Disorder of the C-Terminal Domain of Drosophila Methoprene-Tolerant Protein. *PLoS One* 11(9):e0162950.

56. Shell SS, Putnam CD, & Kolodner RD (2007) The N terminus of Saccharomyces cerevisiae Msh6 is an unstructured tether to PCNA. *Mol Cell* 26(4):565-578.
57. Renner M, *et al*. (2017) Structural dissection of human metapneumovirus phosphoprotein using small angle x-ray scattering. *Sci Rep* 7(1):14865.
58. Watson MC & Curtis JE (2014) Probing the average local structure of biomolecules using small-angle scattering and scaling laws. *Biophys J* 106(11):2474-2482.
59. Yang C, van der Woerd MJ, Muthurajan UM, Hansen JC, & Luger K (2011) Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic Acids Res* 39(10):4122-4135.
60. Konno T, Tanaka N, Kataoka M, Takano E, & Maki M (1997) A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. *Biochim Biophys Acta* 1342(1):73-82.
61. Bressan GC, *et al*. (2008) Human regulatory protein Ki-1/57 has characteristics of an intrinsically unstructured protein. *J Proteome Res* 7(10):4465-4474.
62. Gazi AD, *et al*. (2008) Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. *J Biol Chem* 283(49):34062-34068.
63. Lens Z, *et al*. (2010) Solution structure of the N-terminal transactivation domain of ERM modified by SUMO-1. *Biochemical and biophysical research communications* 399(1):104-110.
64. Boze H, *et al*. (2010) Proline-rich salivary proteins have extended conformations. *Biophys J* 99(2):656-665.
65. Alborghetti MR, *et al*. (2010) Human FEZ1 protein forms a disulfide bond mediated dimer: implications for cargo transport. *J Proteome Res* 9(9):4595-4603.
66. Wells M, *et al*. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci U S A* 105(15):5762-5767.
67. Moncoq K, *et al*. (2004) SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with a transient structure primer? *Biophys J* 87(6):4056-4064.
68. Zheng W, *et al*. (2016) Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *J Am Chem Soc* 138(36):11702-11713.
69. Schwartz R & King J (2006) Frequencies of hydrophobic and hydrophilic runs and alternations in proteins of known structure. *Protein Sci* 15(1):102-112.
70. Das RK, Ruff KM, & Pappu RV (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* 32:102-112.
71. Banks A, Qin S, Weiss KL, Stanley CB, & Zhou HX (2018) Intrinsically Disordered Protein Exhibits Both Compaction and Expansion under Macromolecular Crowding. *Biophys J* 114(5):1067-1079.
72. Watters AL, *et al*. (2007) The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* 128(3):613-624.
73. Junker M, *et al*. (2006) Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc Natl Acad Sci U S A* 103(13):4918-4923.

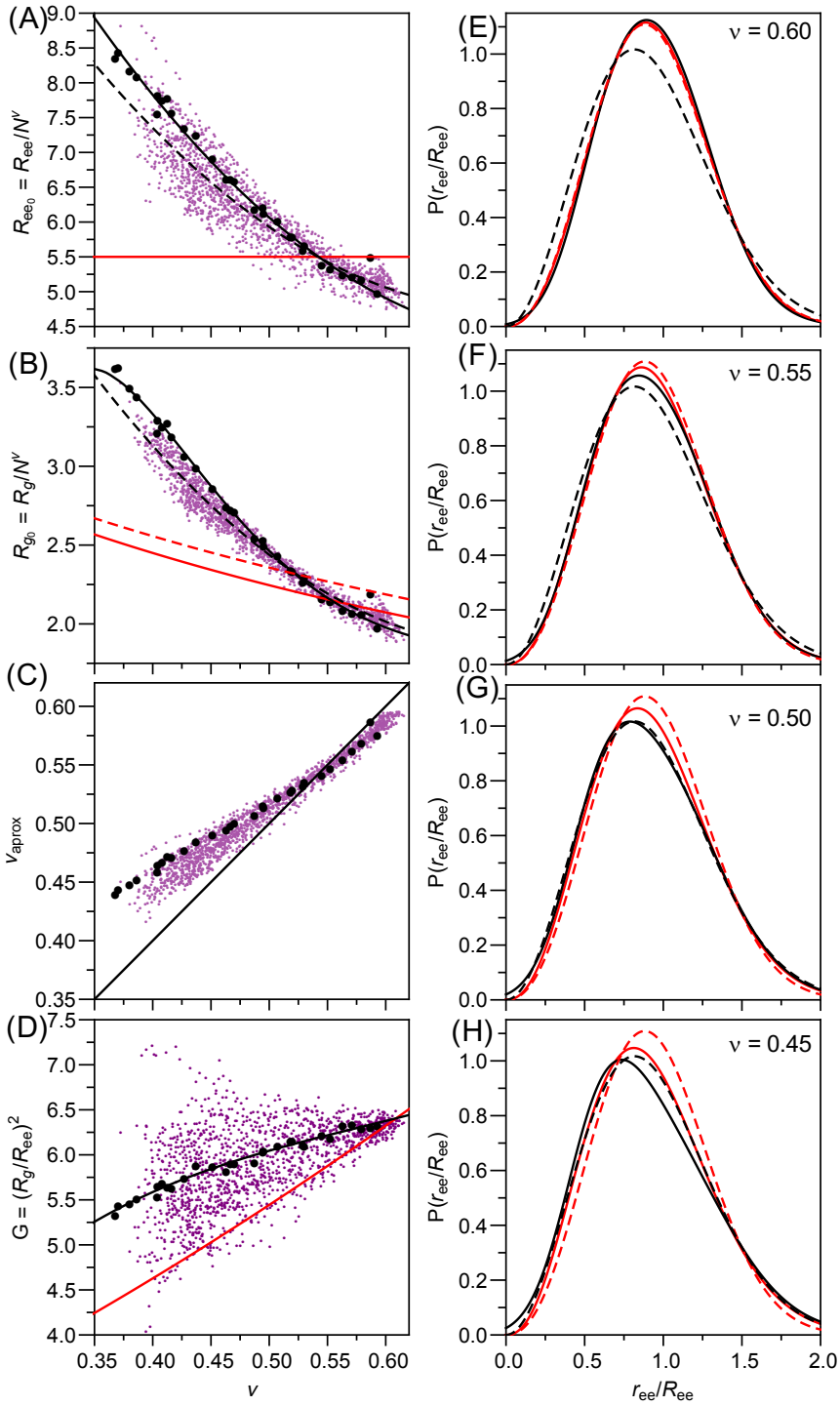


Figure S1: **Comparison of polymer models utilized in SAXS and FRET. (A-D)** Comparison of ν vs. various polymer quantities. Black dots and solid lines, ALL-H PNt simulations and resulting trend from Riback et al., 2017 (46). Purple dots, values from HP-model simulations. Dashed black lines, trends from HP-model simulations with near-even HP distributions. Red line, trends from Borgia et al., 2016 (44). **(B)** Dashed red line, trend from Hoffman et al 2012 (31). **(C)** Comparison of ν determined from the $R_{|i-j|}$ plot and ν approximated assuming R_0=5.5 and the procedure in ref. 31. Black line is an x=y line for reference. **(D)** Red line, trend from Zheng et al., 2018 (45). **(E-H)** Dimensionless (r_{ee}/R_{ee}) end-to-end distribution. Red dashed, SARW model; black dashed Gaussian model; solid red, model from Zheng et

al., 2018 (45). Solid black, model presented herein based on ALL-H PNt simulations at the noted ν values.

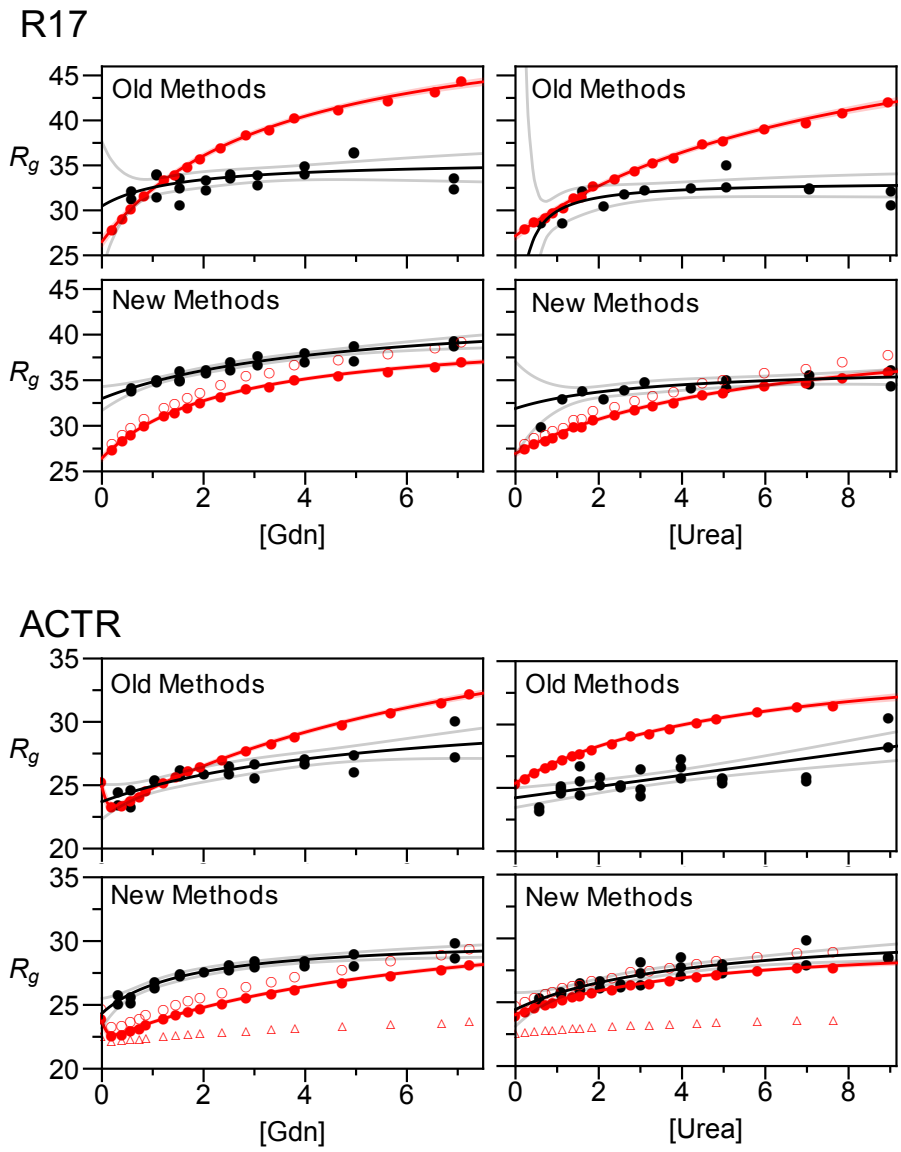


Figure S2: **Comparison of published SAXS and FRET data on the same systems**. Legend equivalent to Figure 1A. R17 and ACTR SAXS and FRET data (from Borgia et al., 2016 (44)). Top, 'Old Methods' Comparison between FRET data fit assuming a Gaussian chain and SAXS data fit with the Guinier approximation. Bottom (filled circles), 'New Methods' SAXS and FRET fit with MFF from ref. 46). Lines shown are best fit hyperbolic trend and mean prediction lines in opaque and slightly transparent, respectively. Open circles are models from Zhang et al 2018 (45). Open triangles correspond to Song et al., 2017 (40) determined at http://dice.utm.utoronto.ca.

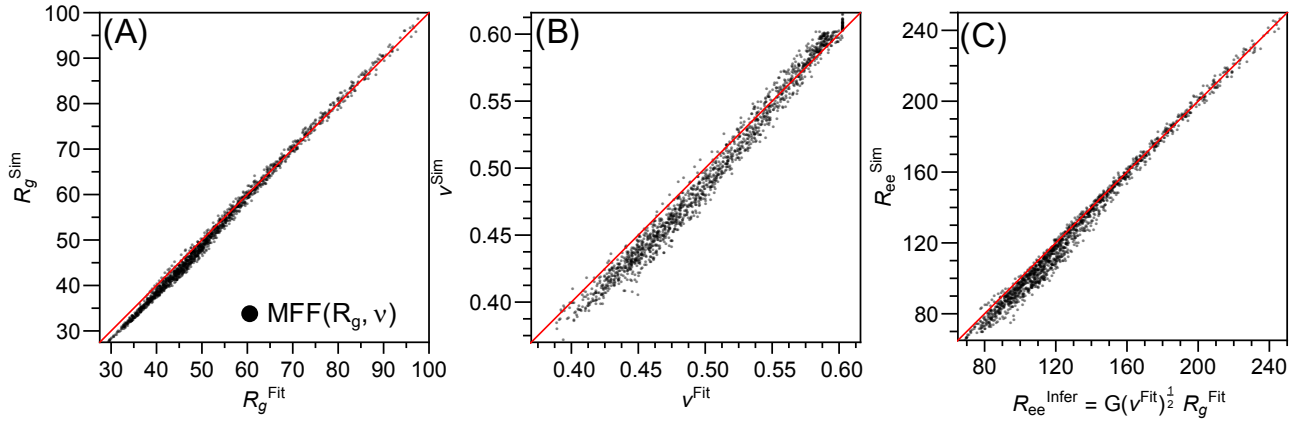


Figure S3: Comparison between extracted parameters from homopolymer MFF with simulation. A-C) Comparison of R_g, ν, and R_{ee} between calculated from coordinates of HP-model simulations and fit with our published MFF(R_g, ν) (46) to SAXS profiles with randomly added experimental errors.

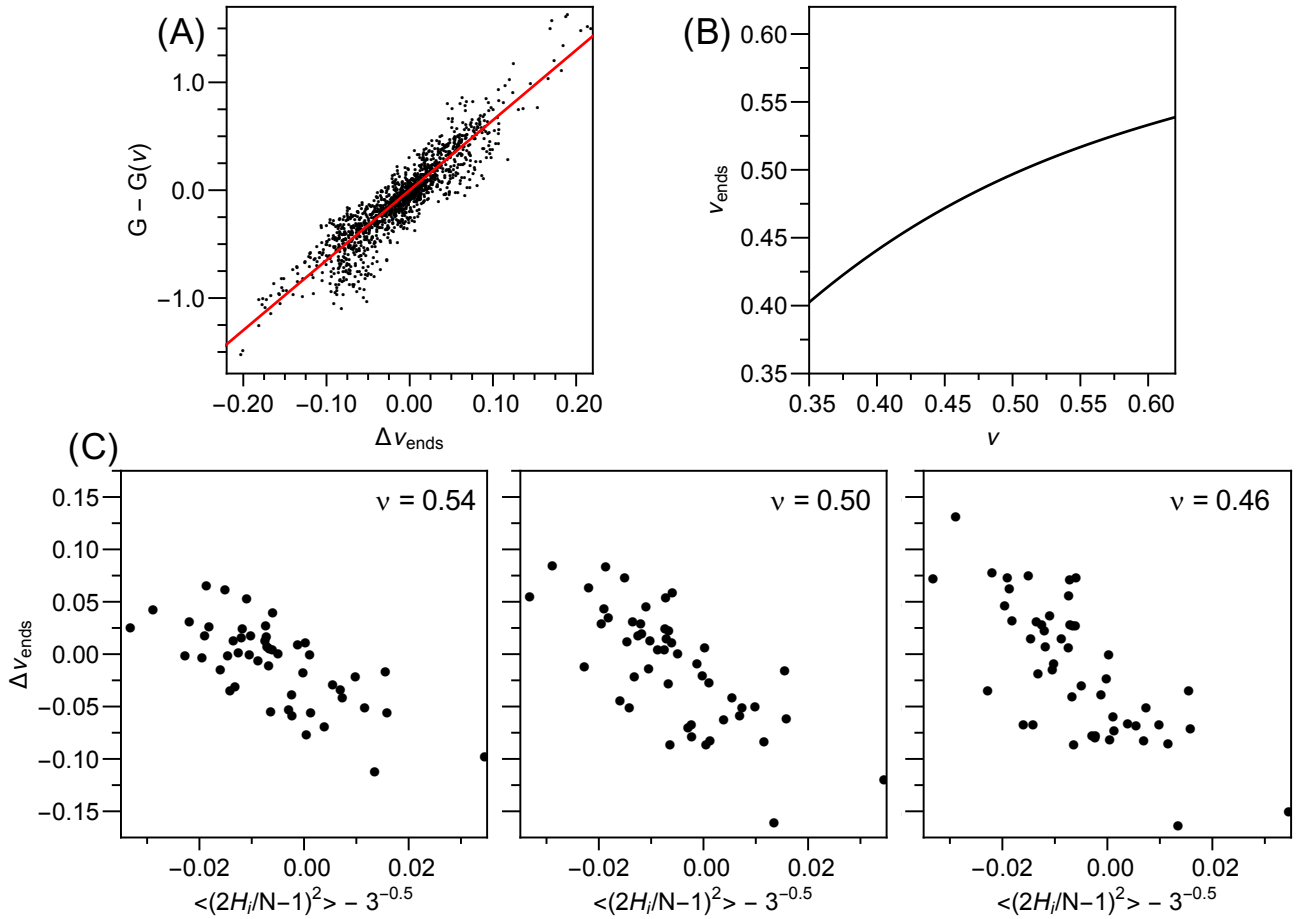


Figure S4: Heteropolymer simulations. (**A**) Correlation between $\Delta\nu_{end}$ with G-G(ν) in HP-model simulations ($(G=R_{ee}/R_g)^2$) where G(ν) is the G value expected for a homopolymer at that value of ν. Red line is the best linear fit. **(B)** Dependence on ν_{ends} with ν extrapolated to simulations where G=G(ν). (**C**) Dependence on $\Delta\nu_{end}$ with the average mean squared deviation of hydrophobic residues from the center of the chain. The x-axis is referenced such that an evenly spaced heteropolymer has a zero x value. The ν values for the three conditions are shown.

Table S1: SAXS on PNt data fit to our $MFF_{homopolymer}$.

		0.150 M KCl	2 M Gdn	4 M Gdn
PNt	R_g (Å):	51.1 ± 0.1	58.7 ± 0.1	62.2 ± 0.5
	ν:	0.542 ± 0.002	0.584 ± 0.001	0.601 ±0.008
	X^2_r:	1.11	0.986	0.946
PNtCC-Alkd	R_g (Å):	51.0 ± 0.2	59.8 ± 0.2	62.2 ± 1.0
	ν:	0.541 ± 0.003	0.593 ± 0.002	0.603 ± N/A
	X^2_r:	1.00	1.04	1.10
PNtC-Alkd	R_g (Å):	51.8 ± 0.2	60.1 ± 0.5	62.0 ± 0.8
	ν:	0.527 ± 0.002	0.588 ± 0.004	0.603 ± N/A
	X^2_r:	1.11	0.977	0.996
PNtCC-Alexa488	R_g (Å):	45.9 ± 0.7	59.6 ± 0.4	65.5 ± 2.7
	ν:	0.498 ± 0.012	0.566 ± 0.004	0.603 ± N/A
	X^2_r:	0.995	0.975	0.867
PNtC-Alexa488	R_g (Å):	49.6 ± 0.2	59.8 ± 0.6	66.1 ± 1.2
	ν:	0.521 ± 0.003	0.576 ± 0.006	0.593 ± 0.015
	X^2_r:	0.892	1.03	1.02

Table S2: Alexa488 fluorescence anisotropy and lifetime measurements[1].

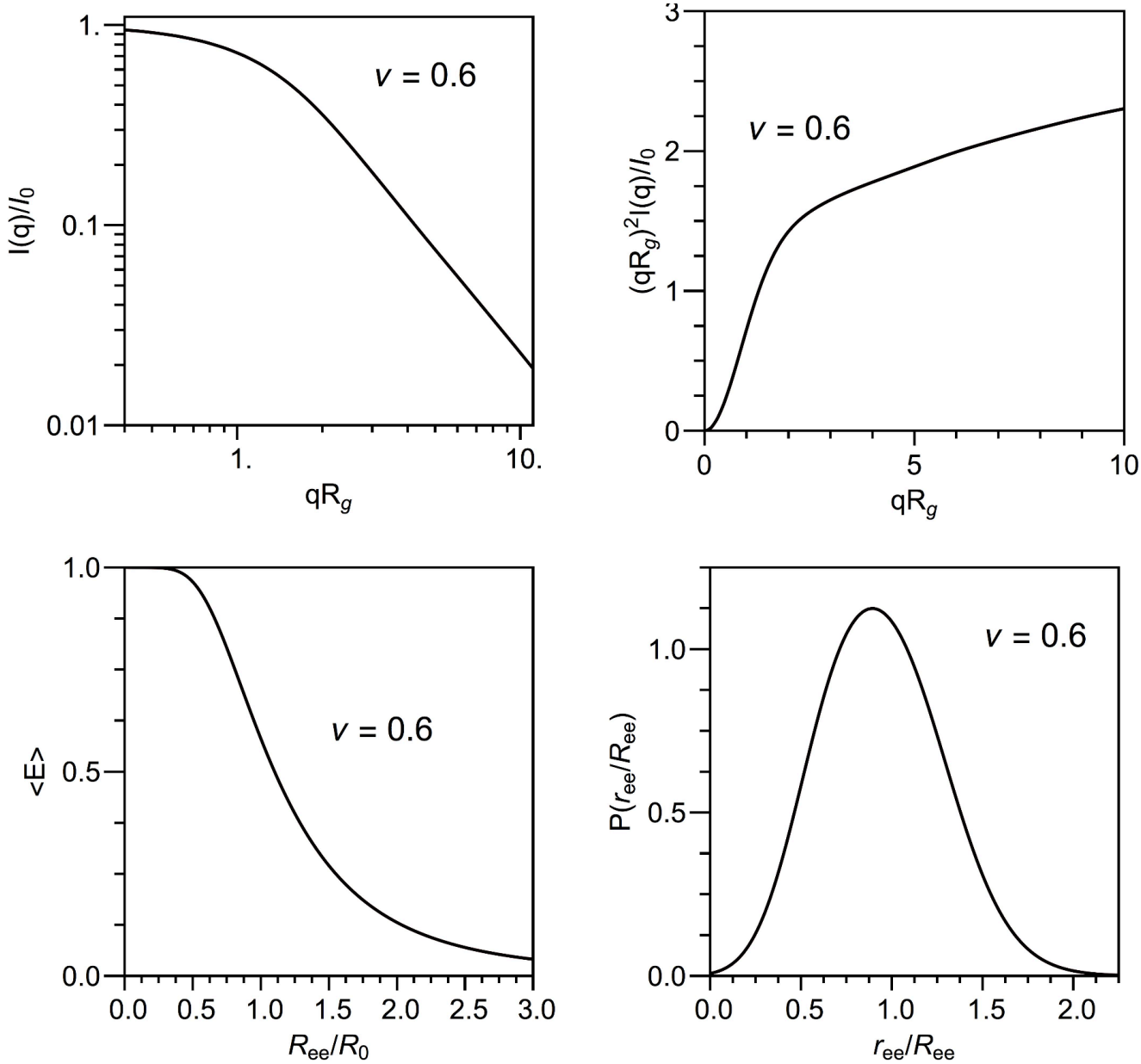
	[Gdn] (M)	Anisotropy	Lifetime (ns)
Free Alexa488	0	0.017 ± 0.001	4.72
	2	0.0304 ± 0.0007	4.82
PNtC-Alexa488	0	0.132 (n=2)	4.67
	2	0.071 (n=2)	4.58
PNtCC-Alexa488	0	0.109 ± 0.005	4.56
	2	0.074 ± 0.001	4.59

[1]Error shown is the SEM of 3 measurements.

Table S3: Analysis of SAXS data on published IDPs with our MFF[1].

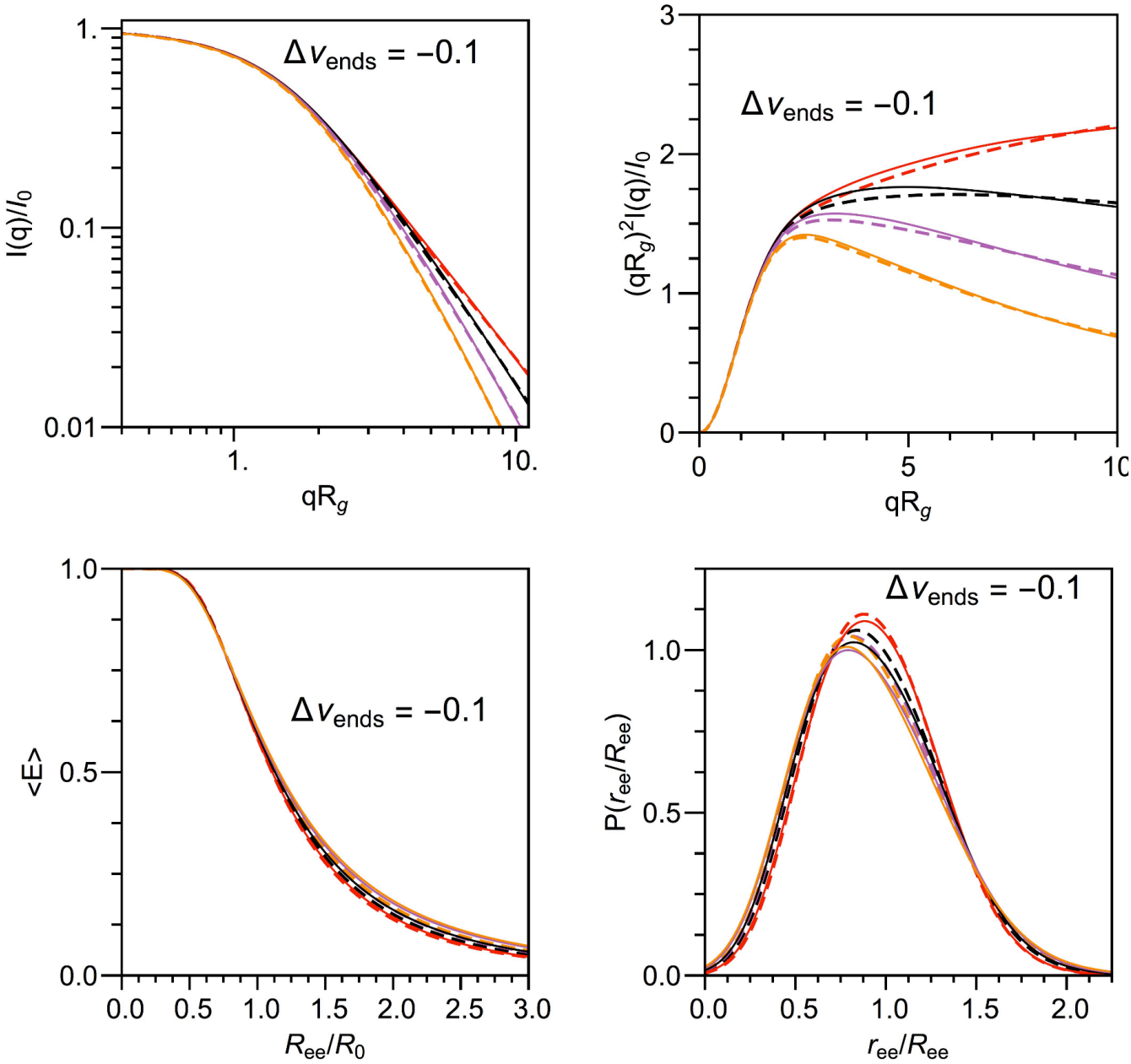
Protein	Citation	R_g(Å)	error	ν	errors	χ^2_r	q min	q max
ID3 of CBP	53	75.7	0.3	0.586	0.001	2.44	0.01	0.15
Ash1	54	29.2	0.2	0.603	0.007	1.04	0.01	0.15
Met2	55	52.7	0.1	0.542	0.002	0.98	0.01	0.15
NSP	43	37.3	0.2	0.603	0.006	1.15	0.01	0.1
N98	43	34.3	0.1	0.556	0.002	2.62	0.03	0.15
IBB	43	33.2	0.1	0.538	0.002	1.05	0.01	0.15
NUS	43	27.3	0.1	0.578	0.010	1.07	0.01	0.15
NLS	43	24.3	0.1	0.603	0.006	0.59	0.01	0.15
N49	43	16.6	0.1	0.537	0.012	0.98	0.01	0.225
NUL	43	33.3	0.1	0.523	0.004	2.20	0.01	0.15
Msh6 NTR	56	61.4	0.2	0.547	0.001	1.93	0.01	0.15
HMPV (P1-60)	57	26.5	0.1	0.603	0.003	0.80	0.01	0.15
redAFP	52	22.2	0.1	0.550	0.006	1.00	0.01	0.225
MeCP2	58,59	59.7	0.9	0.550	0.002	1.11	0.02	0.15
Ki-1/57	58,61	49.5	0.7	0.539	0.003	0.34	0.01	0.15
CSD1	58,60	32.9	0.5	0.565	0.011	0.50	0.02	0.15
HrpO	58,62	35.3	0.3	0.603	N/A	0.72	0.01	0.15
Il-1	58,64	38.7	0.6	0.525	0.005	0.42	0.013	0.142
ERM (1-122)	58,63	37.5	0.5	0.548	0.002	0.14	0.013	0.147
FEZ1	58,65	35.2	0.1	0.603	N/A	0.16	0.01	0.15
p53	58,66	29.1	0.1	0.569	0.007	0.29	0.01	0.15
PIR	58,67	26.5	0.2	0.553	0.008	0.10	0.01	0.15

[1]q is in units of inverse angstroms. When Ref. 58 is included in citation column, Ref. 58 provided data obtained from other studies.



(Initial picture of movie found as attached file 'MovieS1.avi')

Movie S1: Disordered homopolymer polymer functions for SAXS and FRET. Running video changes ν of black curve. Red curve is held at ν=0.6. Note that the x-axis on the left plot is dimensionless and is thus scaled during fitting process.



(Initial picture of movie found as attached file 'MovieS2.avi')

Movie S2: Disordered heteropolymer polymer functions for SAXS and FRET. Running video changes $\Delta\nu_{ends}$ of solid curve. Dashed curves are held at $\Delta\nu_{ends}$=0.0. Red, black, purple, and orange curves are held at ν=0.6, 0.55, 0.5, and 0.45, respectively. Note that the x-axis on the left plot is dimensionless and is thus scaled during fitting process.