PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

A genetically encoded anomalous SAXS ruler to probe the dimensions of intrinsically disordered proteins

Miao Yu[a,b,c,1], Andrey Yu. Gruzinov[d,1], Hao Ruan[a,b,1], Tom Scheidt[a,b,c,1], Aritra Chowdhury[c], Sabrina Giofrè[a,b], Ahmed S. A. Mohammed[d,2], Joana Caria[a], Paul F. Sauter[c], Dmitri I. Svergun[d,3,4], and Edward A. Lemke[a,c,e,3]

Affiliations are included on p. 9.

Edited by Joseph Puglisi, Stanford University School of Medicine, Stanford, CA; received August 6, 2024; accepted October 21, 2024

Intrinsically disordered proteins (IDPs) adopt ensembles of rapidly fluctuating heterogeneous conformations, influencing their binding capabilities and supramolecular transitions. The primary conformational descriptors for understanding IDP ensembles—the radius of gyration (R_G), measured by small-angle X-ray scattering (SAXS), and the root mean square (rms) end-to-end distance (R_E), probed by fluorescent resonance energy transfer (FRET)—are often reported to produce inconsistent results regarding IDP expansion as a function of denaturant concentration in the buffer. This ongoing debate surrounding the FRET-SAXS discrepancy raises questions about the overall reliability of either method for quantitatively studying IDP properties. To address this discrepancy, we introduce a genetically encoded anomalous SAXS (ASAXS) ruler, enabling simultaneous and direct measurements of R_G and R_E without assuming a specific structural model. This ruler utilizes a genetically encoded noncanonical amino acid with two bromine atoms, providing an anomalous X-ray scattering signal for precise distance measurements. Through this approach, we experimentally demonstrate that the ratio between R_E and R_G varies under different denaturing conditions, highlighting the intrinsic properties of IDPs as the primary source of the observed SAXS-FRET discrepancy rather than shortcomings in either of the two established methods. The developed genetically encoded ASAXS ruler emerges as a versatile tool for both IDPs and folded proteins, providing a unified approach for obtaining complementary and site-specific conformational information in scattering experiments, thereby contributing to a deeper understanding of protein functions.

intrinsically disordered protein | anomalous SAXS | distance ruler | structural biology | FRET

Intrinsically disordered regions or proteins (we here collectively refer to them as IDPs) are a class of proteins that lack stable secondary or tertiary structures, constituting approximately 30% of the eukaryotic proteome. IDPs encode diverse functions in essential cellular processes (1, 2). Unlike their folded counterparts, IDPs exhibit dynamic ensembles of rapidly interconverting conformations (3, 4). Understanding their degree of disorder and intramolecular interactions is crucial for unraveling their biological functions in solution, particularly regarding condensation or aggregation at higher concentrations. The two primary experimental observables for assessing conformational ensembles are the radius of gyration (R_G), which averages over all interresidue distances, and the rms end-to-end distance (R_E), which interrogates a specific distance distribution between two residues. However, for IDPs, R_G measured by small-angle X-ray scattering (SAXS) and R_E inferred from single-molecule fluorescent resonance energy transfer (smFRET) measurements were repeatedly reported to show different dependencies on the solvent conditions. Specifically, when the denaturant concentration was reduced, numerous IDPs were observed in smFRET experiments to collapse from expanded ensembles to more compact dimensions. In contrast, SAXS experiments detected little or no compaction under the same conditions (5–8).

A persistent debate surrounds the decoupling of R_E and R_G for IDPs between SAXS and FRET communities, i.e., a substantial change in $G = R_E^2/R_G^2$ as a function of denaturant concentration. This debate is challenging due to the intrinsic limitations of both methods (6, 9–16). SAXS demands higher protein concentrations susceptible to intermolecular interactions, while FRET requires potentially intrusive fluorescent labeling, criticized for example for inducing protein collapse (17). Additionally, deriving R_E or R_G from FRET efficiency often assumes a specific model for distance distribution. Resolving this discrepancy is crucial for understanding IDP conformations and functions, necessitating novel experimental methods.

Significance

Small-angle X-ray scattering (SAXS) provides globally averaged dimensions like the radius of gyration (R_G) but lacks site-specific details such as the end-to-end distance (R_E). To address this, we enhanced SAXS by incorporating two small noncanonical amino acid labels into intrinsically disordered proteins (IDPs), enabling simultaneous extraction of R_E and R_G through anomalous SAXS (ASAXS). ASAXS results reveal a change in the R_E-to-R_G ratio upon denaturation, consistent with previous fluorescent resonance energy transfer (FRET) and SAXS findings. This suggests that discrepancies between FRET and SAXS in describing IDP behaviors stem from inherent IDP attributes rather than methodological shortcomings. Consistent measurement and reporting of fundamental biophysical parameters are essential for understanding IDP roles in phase separation and neurodegenerative diseases.

This article is a PNAS Direct Submission.

[1]M.Y., A.Y.G., H.R., and T.S. contributed equally to this work.

[2]Present address: Physics Department, Fayoum University, Fayoum 63514, Egypt.

[3]To whom correspondence may be addressed. Email: svergun@biosaxs.com or edlemke@uni-mainz.de.

[4]Present address: BIOSAXS GmbH, Hamburg 22607, Germany.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2415220121/-/DCSupplemental.

Published December 6, 2024.

Here, we present a genetically encoded anomalous SAXS (ASAXS) ruler, enabling simultaneous direct measurement of R_G and R_E without assuming a structural model (Fig. 1). ASAXS leverages changes in scattering at different energies near the absorption edges of specific atoms, offering a powerful biological tool with diverse applications, including obtaining distance information between the heavy labels in biological macromolecules (18–22). To achieve site-specific labeling of target proteins with minimal perturbation on their conformations and functions, we first designed a small ASAXS probe, a bromine-containing noncanonical amino acid (Br-ncAA), in which the bromine atom produces a specific anomalous signal. Using amber suppression technology, we then evolved an engineered tRNA/tRNA synthetase pair to incorporate the Br-ncAA at opposing termini of the proteins site-specifically. Taking two different IDPs, the short disordered importin-β-binding domain of importin-α (IBB) and an elastin-like polypeptide (ELP) as model systems, we directly extracted R_E, i.e., the rms distance between the two terminal Br-ncAAs, and R_G from ASAXS measurements. By measuring the R_G and R_E values under near-native and denaturing conditions, we experimentally demonstrated a decoupling between R_G and R_E for IDPs using ASAXS as a unified method. Further comparison of R_E values probed by ASAXS and smFRET exhibited good agreement. The decoupling between R_E and R_G observed in our experiments reflects the unique behavior of IDPs, where these parameters respond differently to changes in solvent conditions. The inherent flexibility of IDPs allows for significant fluctuation in end-to-end distances, even if the overall size remains relatively constant. Different variations of R_E and R_G values could be observed, for example, if the overall shape of the protein changes between the conformational ensembles at different conditions (see e.g., simulations in ref. 7). This underscores the importance of using complementary methods to study IDPs, as a single experimental technique may not fully capture their conformational diversity. Our present findings suggest that the observed SAXS-FRET discrepancy primarily stems from the inherent conformational flexibility and heterogeneity of IDPs, rather than experimental limitations of the SAXS or FRET methods. The developed ASAXS ruler offers a structural biology tool, providing complementary parameters to study protein conformation–function relationships in solution.

Results

Synthesis of a Br-ncAA and Evolution of the Aminoacyl-tRNA Synthetase. High-precision ASAXS measurements of IDP conformations require introducing a small scattering probe to minimize the impact on the protein's conformations while providing sufficient anomalous X-ray scattering signals. To achieve this, we designed a small ncAA bearing two Br atoms, specifically a dibrominated derivative of lysine, *N*-ε-3,5-dibromobenzyloxycarbonyl-L-lysine (diBrK) (Fig. 1*A*). The synthesis involved activating 3,5-dibromobenzyl alcohol with the coupling reagent 1,1'-carbonyldiimidazole to give the corresponding carbamate. Subsequent coupling with Fmoc-protected L-lysine and

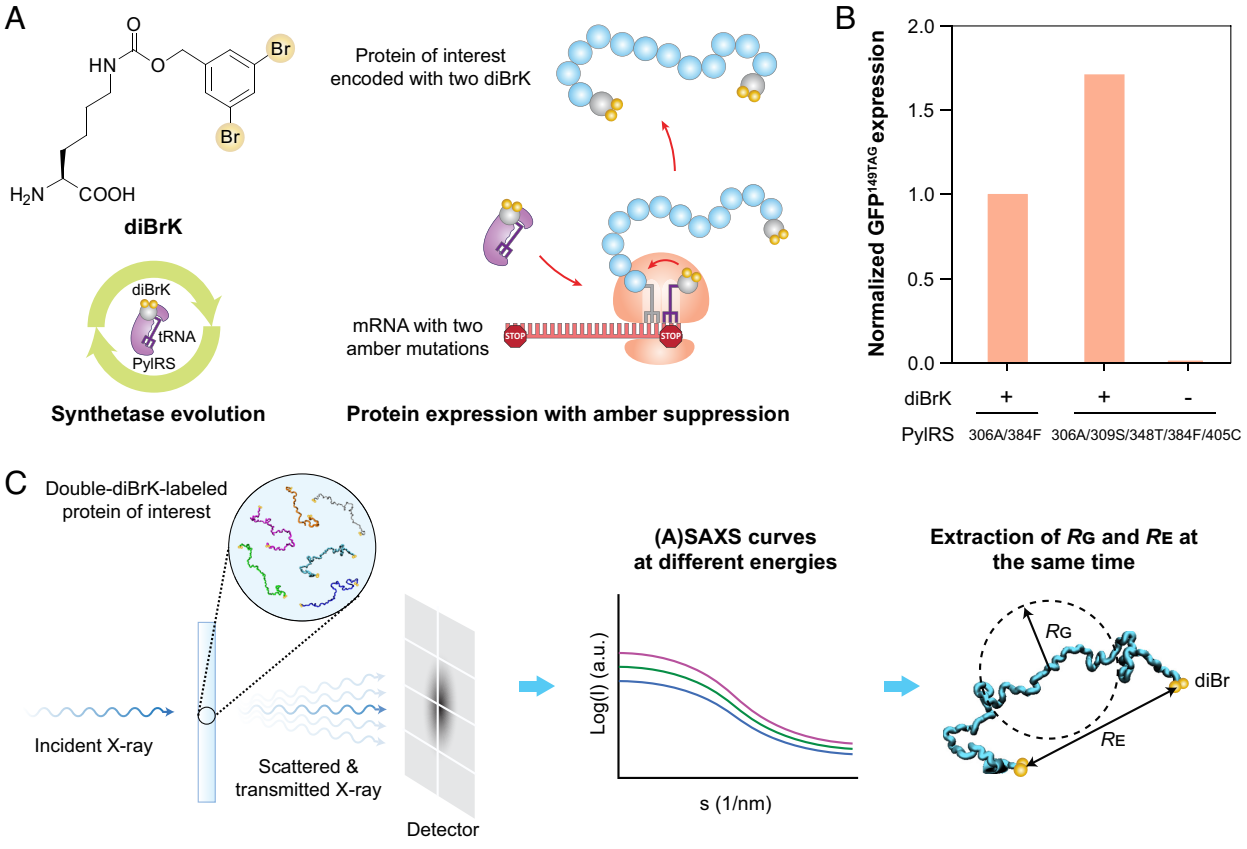

Fig. 1. A genetically encoded ASAXS ruler to directly measure both, the radius of gyration R_G and the rms end-to-end distance R_E for proteins. (*A*) A Br-bearing noncanonical amino acid, *N*-ε-3,5-dibromobenzyloxycarbonyl-L-lysine (diBrK), was designed as a small ASAXS probe. An engineered pyrrolysyl-tRNA synthetase (PylRS) is evolved to incorporate diBrK into the protein of interest at specific sites using amber suppression technology. (*B*) Normalized fluorescence intensities indicating the expression levels of amber codon-disrupted GFP^{149TAG} in *E. coli* modified with the evolved PylRS and tRNA pair in the absence and presence of diBrK. The best-hit $PylRS^{306A/309S/348T/384F/405C}$ gave about 1.7-fold efficiency enhancement compared to $PylRS^{306A/384F}$. (*C*) Schematics showing the analysis pipeline in which R_G and R_E, i.e., the rms distance between the two diBrK labels encoded near the termini of proteins of interest, were directly obtained from the scattering curves by ASAXS measurements at different energies near the absorption edge of the Br atoms.

a final deprotection step in piperidine furnished the diBrK product with 68% yield from a sequential one-pot procedure (*SI Appendix*, Fig. S1; see *Materials and Methods*).

We utilized amber suppression technology to incorporate diBrK into proteins at specific sites (23). To enhance the incorporation efficiency, we conducted evolution of the engineered aminoacyl-tRNA synthetase PylRS$^{306A/384F}$, a variant of pyrrolysyl-tRNA synthetase from *Methanosarcina mazei*, with the mutations 306A/384F, that has previously been used to genetically encode large side chains (24), targeting the diBrK (*SI Appendix*, Fig. S2; see *Materials and Methods*). After five positive and negative selection cycles, we obtained the evolved best-hit synthetase with additional mutations 309S/348T/405C. To assess the efficiency and fidelity of diBrK incorporation into proteins in *E. coli*, an amber mutation (TAG) was introduced at position Asp149 in a C-terminal His-tagged GFP variant (GFP^{149TAG}). Fluorescence intensities of GFP indicated that the best-hit PylRS$^{306A/309S/348T/384F/405C}$ gave ~1.7-fold efficiency enhancement compared to PylRS$^{306A/384F}$ (Fig. 1*B*). Full-length GFP was produced only in the presence of diBrK, demonstrating the high incorporation fidelity of the newly evolved synthetase.

Standard SAXS Measurements of Unlabeled and Double-diBrK-Labeled IBB. We first chose a short 76-amino-acid segment of IBB, which was predicted to lack a secondary structure, as a model protein (Fig. 2*A*) (25–27). This short fragment ensures a comparatively high signal-to-background ratio (the ratio between the anomalous signals from the incorporated diBrK and the scattering from the remaining protein chain) for ASAXS measurements. We expressed double-amber-mutated IBB (IBB$^{2TAG,\ 72TAG}$) in *E. coli* modified with the evolved tRNA synthetase PylRS$^{306A/309S/348T/384F/405C}$, and produced the target protein in an acceptable yield (*Materials and Methods*). We confirmed the incorporation of two diBrK residues by SDS-PAGE analysis and mass spectrometry (*SI Appendix*, Fig. S3 and Tables S1 and S2).

Standard SAXS measurements on Br-labeled IBB under varying solvent conditions revealed an increase in the radius of gyration, $R_{G,L}$, from 2.7 ± 0.1 nm to 3.1 ± 0.1 nm with rising urea concentrations (from 0.5 M to 6 M), (Fig. 2*B* and *SI Appendix*, Tables S3 and S4), indicating more expanded conformations at higher denaturant concentrations, in agreement with previous observations for other proteins (7, 17). Unlabeled IBB (i.e., wildtype) also exhibited increased radius of gyration, $R_{G,U}$, from 2.6 ± 0.1 nm to 2.7 ± 0.1 nm under the same denaturant conditions. A slight but systematic increase in R_G was detected for labeled IBB due to the presence of Br atoms at the termini.

The two dibromobenzyl moieties are more electron-rich than the rest of the protein, which allowed assessment of their distance by comparing the R_G values of the unlabeled and labeled samples. Applying the parallel axis theorem in classical mechanics, $R_{G,L}$ was expressed as (29):

$$I_L(0)R_{G,L}{}^2 = I_U(0)R_{G,U}{}^2 + I_{Br}(0)R_{G,Br}{}^2 + I_U(0)I_{Br}(0)d^2, \quad [1]$$

where $I_L(0)$, $I_U(0)$, and $I_{Br}(0)$ are the forward scattering from the labeled IBB, unlabeled IBB, and the Br labels, respectively. The predicted forward scattering values were computed using the CRYSOL program (*Materials and Methods*) (30). $R_{G,Br}$ is the radius of gyration of Br moiety. The third (cross-) term in Eq. **1** is proportional to the squared distance d between the centers of scattering mass of the protein moiety and that of the Br labels (31). The rms distance between the bromine centers on the termini, R_E is approximately twice the radius of gyration of Br moiety, $R_{G,Br}$, and can thus be estimated as

$$R_E \approx 2 \times R_{G,Br} = 2 \times \sqrt{\left[I_L(0)R_{G,L}{}^2 - I_U(0)R_{G,U}{}^2 - I_U(0)I_{Br}(0)d^2\right]/I_{Br}(0)}. \quad [2]$$

Given that IDPs are disordered, the cross-term contribution is expected to be relatively small. Neglecting the cross-term (i.e., assuming the centers of scattering mass between the protein moiety and Br labels coincide) in Eq. **2**, we directly estimated R_E, which increased from 5.9 ± 0.9 nm in 0.5 M urea to 9.8 ± 0.4 nm in 6 M urea (Fig. 2*C* and *SI Appendix*, Table S5). It is important to note that the cross-term might be bigger than zero for the labeled IBB in different solvents. Therefore, the above calculated R_E from the standard SAXS measurements only provides an upper estimate of the distance between the two diBrK residues, as we denoted as the estimated maximal R_E in Fig. 2*C*.

Extraction of RE from ASAXS Measurements on Double-diBrK-Labeled IBB. To enhance the accuracy of R_E measurement, we conducted ASAXS measurements on Br-labeled IBB. ASAXS has previously demonstrated angstrom-resolution precision in determining distances between gold nanoparticle labels at biomolecular termini (20). Here, we used diBrK as the scattering probe, which has a bromo–bromo distance of 0.58 nm and a molecular weight of 0.44 kDa. In comparison, the gold nanoparticles previously used had a diameter of 1.4 nm and an estimated molecular weight of 17 kDa (20). Our diBrK probe, being an order of magnitude smaller in molecular weight, allows for residue-specific labeling with minimal perturbation to the protein structure, despite the challenge of detecting its much weaker anomalous signal. With its low background, high brilliance, and widely tunable energy range, the BioSAXS beamline P12 at EMBL Hamburg is optimized for fast, high-quality data collection with biological solutions, facilitating overcoming this challenge (21, 32–35).

We collected ASAXS signals of Br-labeled IBB in size exclusion chromatography (SEC) mode at five energy points (13,428, 13,453, 13,468, 13,472, and 13,476 eV, Fig. 3 *A*–*C*) around the Br K-edge (13,474 eV; see *SI Appendix*, Fig. S4). The SAXS intensity I at energy E over the range of scattering vector s reads as (18):

$$I(s,E) = F_0^2(s) + 2f'(E)F_0(s)v_0(s) + \left(f'^2(E) + f''^2(E)\right)v_0^2(s). \quad [3]$$

where $F_0(s)$ is the nonresonant scattering intensity far from the absorption edge, $v_0(s)$ is the resonant scattering intensity from the spatial distribution of anomalous atoms, and $F_0(s)v_0(s)$ is the cross-term represented as a product of nonresonant and resonant scattering amplitudes. f' and f'' are the real and imaginary parts of the X-ray scattering factor that become significant near the energy of the absorption edge of a particular element. The respective scattering factors f' and f'' were referenced from (36) and calculated using the theoretical approximation developed by Cromer and Liberman (37–41) (*SI Appendix*, Table S6). Because $F_0(s)$ corresponds to the Fourier transform of scattering length density, the electron density contribution from the entire particle $v_0(s)$ is related solely to the anomalous scattering length density and reflects the distribution of the anomalous scattering atoms. If the ASAXS measurements are performed at three or more energy points, the system of linear equations in Eq. **3** can be solved to

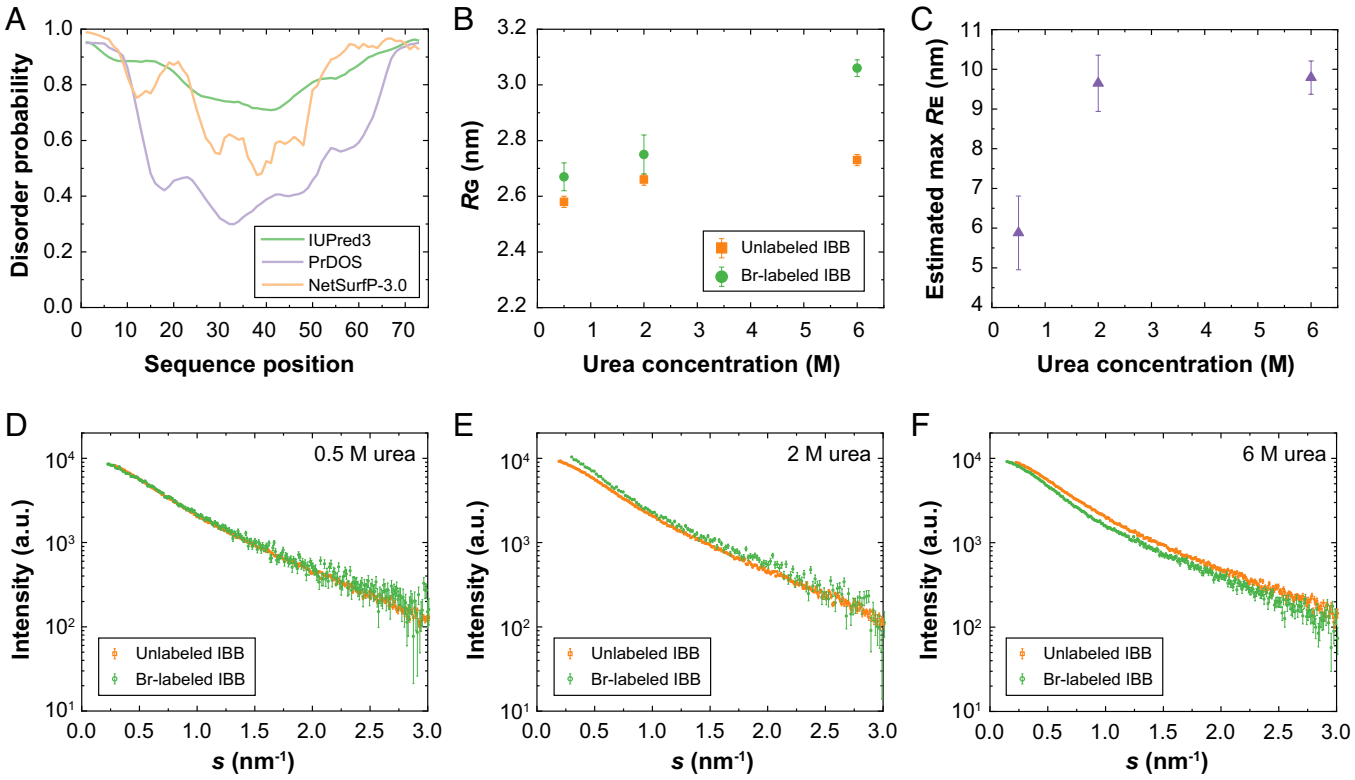

Fig. 2. Standard SAXS measurements for unlabeled and double-diBrK-labeled IBB. (*A*) The probed IBB segment is predicted with a high disorder probability by different predictors (25–27). (*B*) R_G obtained for unlabeled and Br-labeled IBB at different urea concentrations. The symbols and error bars represent the average and SEM, respectively, calculated using PRIMUS from the ATSAS package (28). (*C*) The estimated maximal R_E of Br-labeled IBB at different urea concentrations. The error bars are calculated using error propagation incorporating the respective errors for each term in Eq. **2** and *SI Appendix*, Eqs. **S13–S19**. SAXS curves of unlabeled and Br-labeled IBB in (*D*) 0.5 M urea, (*E*) 2 M urea, and (*F*) 6 M urea. The logarithmic intensity is displayed as a function of momentum transfer $s = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength. The error bars on the SAXS curves in (*D–F*) are due to the standard radial averaging procedure for obtaining 1D scattering curves from 2D scattering data measured by the X-ray detector, estimated based on Poisson counting statistics for photons (28).

extract the resonant scattering intensity term $v_0^2(s)$ (*SI Appendix*, Figs. S5–S8) (42). More energy points provide more stable results of this decomposition. Here, we chose the five energy points at which the variations of f' and f'' are large, so that the nonresonant, resonant, and cross-term intensities could be determined with high confidence. Initially collected SAXS curves were normalized before further processing, using previously reported theoretical values of f' and f'' (20). By global fitting the scattering patterns, we decomposed the signals into energy-independent and anomalous contributions (Fig. 3 *D–I*). The R_G values extracted from the energy-independent terms were consistent with the standard SEC-SAXS results (Fig. 3*J* and *SI Appendix*, Table S7). We then extracted R_E from the anomalous term $v_0(s)$ by doubling the experimental $R_{G,Br}$ value where R_E increased from 6.2 ± 0.7 nm in 0.5 M urea to 8.7 ± 0.8 nm in 6 M urea (Fig. 3*K* and *SI Appendix*, Table S8). These values correlate well with the maximal R_E estimated from the standard SAXS analyses (Fig. 2*C* and *SI Appendix*, Table S5), providing a direct and independent assessment of the interlabel distance.

R_G and R_E Are Decoupled for IBB and ELP. To investigate the decoupling between R_E and R_G for IBB at varying denaturant concentrations, we calculated $G = R_E^2/R_G^2$, a descriptor to assess the overall shape of an IDP. Using the R_E and R_G values from ASAXS measurements, we observed a clear R_E–R_G decoupling, with G increasing from 5.5 ± 1.2 in 0.5 M urea to 7.1 ± 1.6 in 2 M urea to 8.2 ± 1.4 in 6 M urea for IBB (Fig. 3*L* and *SI Appendix*, Table S9). This indicates that the expansion of IBB in denaturants is more pronounced when measured in terms of R_E compared to R_G, aligning with the findings from previous studies that integrated FRET and SAXS (7, 8). Notably, the G obtained in high denaturants exceeds 7.04 [predicted for an infinite-long, swollen homopolymer chain in good solvent condition; $G = 6$ in theta solvent condition (43, 44). Our measured even larger G could be attributed to finite size effects and sequence heterogeneity, considering that IBB is a fairly short heteropolymer chain (7, 45, 46).

Encouraged by the promising outcomes observed in the IBB study, we extended our investigation to include another IDP characterized by a longer sequence—the ELP. Inspired by the intrinsically disordered domain of tropoelastin, ELP features 30 repeats of the VGVPG pentapeptide motif (47). ELP can be viewed as a block copolymer due to the chemical similarity of its repeating units. Compared to IBB, ELP has no charge and higher hydrophobicity. Employing the same ASAXS measurement and analysis procedures as for IBB, we derived R_E and R_G values for double-diBrK-labeled ELP under various solvent conditions (Fig. 4 and *SI Appendix*, Tables S7 and S8). The ASAXS curves for ELP exhibit increased noise due to the protein's lower solubility. We maintained a relatively low concentration to prevent phase separation and aggregation, to maximize the measurement of monomers in SEC-ASAXS. Moreover, given the larger size of ELP compared to IBB (~150 aa versus ~70 aa), the signal-to-background ratio is lower, as the anomalous signals from the incorporated diBrK are less pronounced compared to the scattering from the larger protein chain. Despite the noisier ASAXS curves, we still observed a decoupling trend, with the G value changing from 4.8 ± 0.9 in 0.5 M urea to 4.3 ± 1.6 in 2 M urea to 6.6 ± 1.2 in 6 M urea (Fig. 4*L* and *SI Appendix*, Table S9). Notably, the decoupling was less pronounced for ELP than IBB, consistent with previous studies indicating that this effect is diminished in block

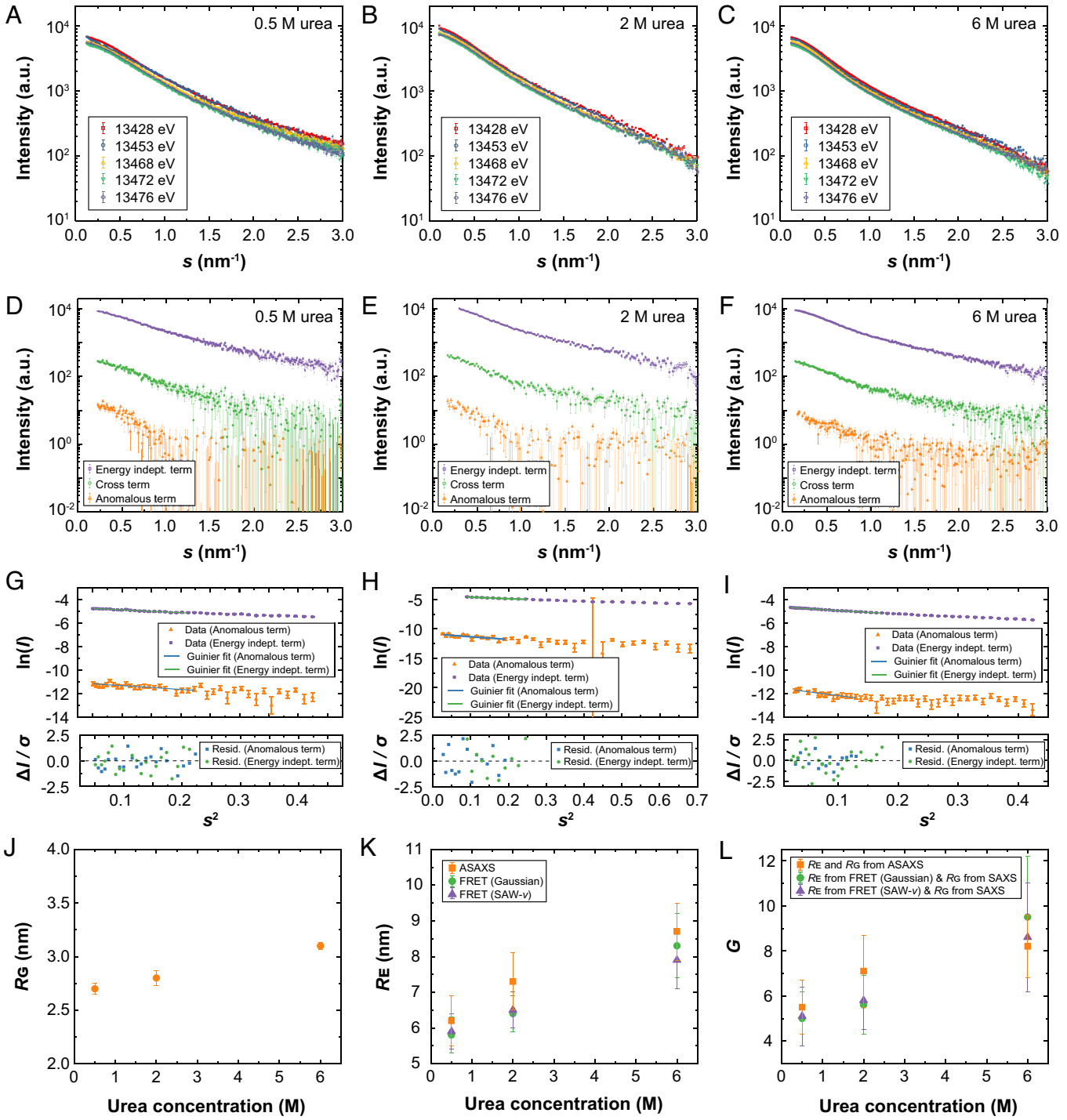

Fig. 3. ASAXS measurements of R_G and R_E for double-diBrK-labeled IBB. ASAXS curves at five different energies (13,428 eV, 13,453 eV, 13,468 eV, 13,472 eV, and 13,476 eV) around the absorption edge of Br for Br-labeled IBB in (*A*) 0.5 M urea, (*B*) 2 M urea, and (*C*) 6 M urea. The error bars on the SAXS curves result from the standard radial averaging procedure for obtaining 1D scattering curves from 2D scattering data measured by the X-ray detector, estimated based on Poisson counting statistics for photons (28). The energy-independent term, the anomalous term, and the cross-term extracted from ASAXS curves in (*D*) 0.5 M urea, (*E*) 2 M urea, and (*F*) 6 M urea. The error bars on the ASAXS curves in (*D–F*) represent the variation in intensity measurements, reflecting the uncertainty range generated through the resampling approach using DATRESAMPLE (28) with $N = 1{,}000$ as discussed in *SI Appendix*. Guinier plots extracted from the energy-independent term and the anomalous term in (*G*) 0.5 M urea, (*H*) 2 M urea, and (*I*) 6 M urea. (*J*) R_G extracted from the energy-independent term, with errors estimated using PRIMUS from the ATSAS package (28). (*K*) R_E derived from ASAXS or smFRET using the Gaussian chain model or SAW-ν model in different urea concentrations. For ASAXS results, errors are given as results of R_G estimation from the anomalous term using PRIMUS from the ATSAS package (28) and are doubled for R_E according to *SI Appendix*, Eq. **S12**. For smFRET results, errors are estimated from the 95% confidence region derived from the Jacobian matrix during the lifetime curve fitting process incorporating the error of Förster distance R_0. (*L*) $G = R_E^2/R_G^2$ obtained directly from ASAXS measurements or by using R_E from FRET and R_G from SAXS changed at different urea concentrations, showing a decoupling effect between R_G and R_E for IBB. The error bars are calculated using error propagation incorporating the respective errors for each term in *SI Appendix*, Eqs. **S7–S11**.

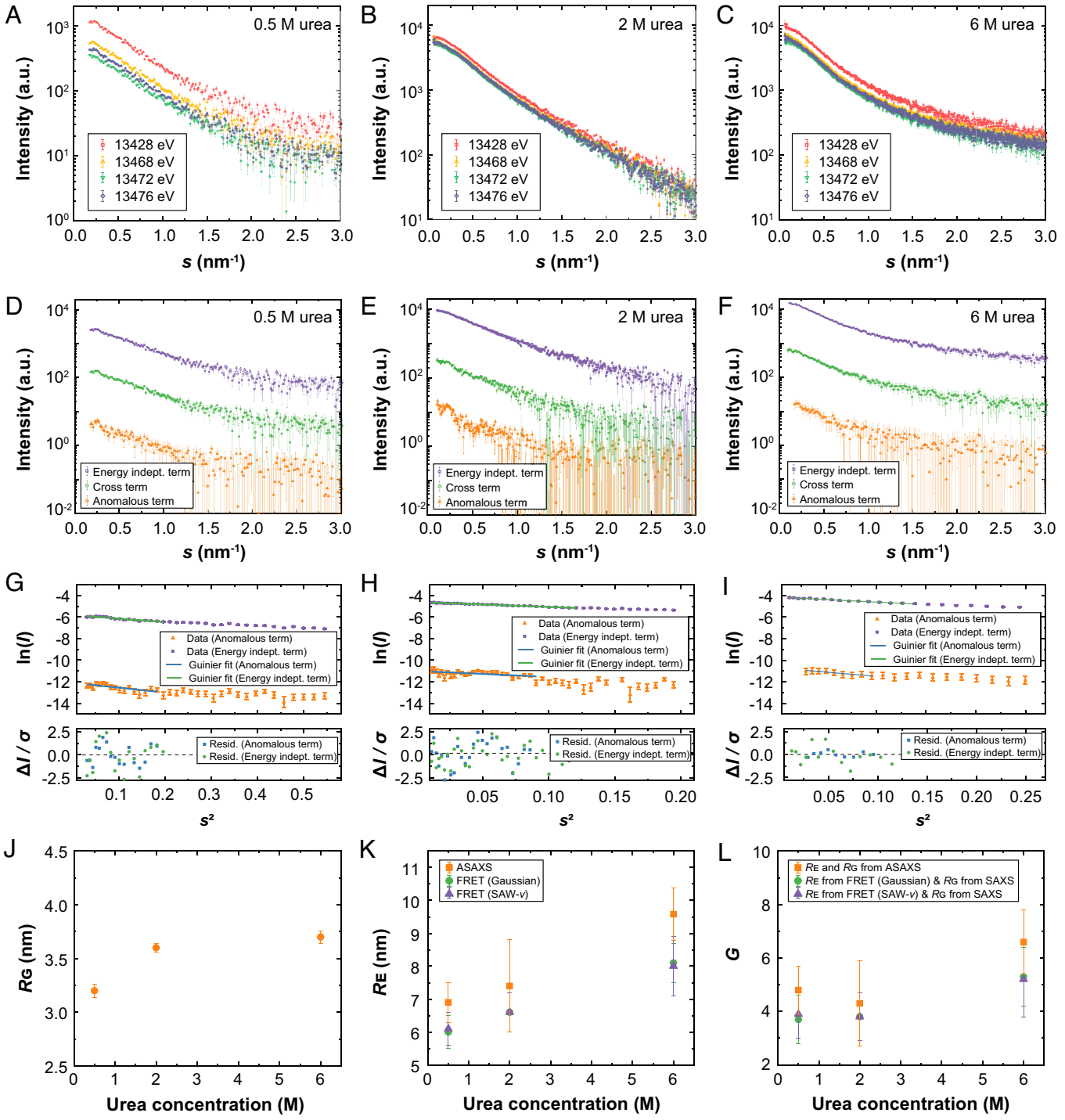

Fig. 4. ASAXS measurements of R_G and R_E for double-diBrK-labeled ELP. ASAXS curves at four different energies (13,428 eV, 13,468 eV, 13,472 eV, and 13,476 eV) around the absorption edge of Br for Br-labeled ELP in (*A*) 0.5 M urea, (*B*) 2 M urea, and (*C*) 6 M urea. The error bars on the SAXS curves result from the standard radial averaging procedure for obtaining 1D scattering curves from 2D scattering data measured by the X-ray detector, estimated based on Poisson counting statistics for photons (28). The energy-independent term, the anomalous term, and the cross-term extracted from ASAXS curves in (*D*) 0.5 M urea, (*E*) 2 M urea, and (*F*) 6 M urea. The error bars on the ASAXS curves in (*D–F*) represent the variation in intensity measurements, reflecting the uncertainty range generated through the resampling approach using DATRESAMPLE (28) with $N = 1{,}000$ as discussed in *SI Appendix*. Guinier plots extracted from the energy-independent term and the anomalous term in (*G*) 0.5 M urea, (*H*) 2 M urea, and (*I*) 6 M urea. (*J*) R_G derived from the energy-independent term, with errors estimated using PRIMUS from the ATSAS package (28). (*K*) R_E derived from ASAXS or from smFRET using the Gaussian chain model or SAW-ν model in different urea concentrations. For ASAXS results, errors are given as results of R_G estimation from the anomalous term using PRIMUS from the ATSAS package (28) and are doubled for R_E according to *SI Appendix*, Eq. **S12**. For smFRET results, errors are estimated from the 95% confidence region derived from the Jacobian matrix during the lifetime curve fitting process incorporating the error of Förster distance R_0. (*L*) $G = R_E^2/R_G^2$ obtained directly from ASAXS measurements or by using R_E from FRET and R_G from SAXS changed at different urea concentrations, showing a decoupling effect between R_G and R_E for ELP. The error bars are calculated using error propagation incorporating the respective errors for each term in *SI Appendix*, Eqs. **S7–S11**.

copolymers and homopolymers (7, 48). The G value for ELP is close to 6, which is similar to the behavior expected for an ideal (Gaussian) chain. The theoretical prediction $G = (2\nu + 1) \times (2\nu + 2)$, where ν is the scaling exponent of the polymer scaling law (43, 44), originates from polymer physics models applied to an ideal chain (49). We compared this theoretical framework with our

experimental results for IBB and ELP. By calculating $(2\nu + 1) \times (2\nu + 2)$ using the measured R_E and the SAW-ν model (49), we observed a trend consistent with the analysis of 26,839 human intrinsically disordered regions from ref. 50 (*SI Appendix*, Fig. S9).

Taken together, our unified scattering method provides direct experimental proof of the decoupling between R_G and R_E for IDPs under different denaturing conditions.

Comparative FRET Measurements for IBB and ELP. To assess the accuracy of ASAXS measurements, we conducted a series of smFRET experiments on both IBB and ELP under varying denaturant conditions. Both proteins were labeled with the FRET dye pair Alexa 488 and Alexa 594 at identical positions as in the ASAXS measurements (*Materials and Methods*). Measurements were taken at a concentration of approximately 50 pM with proteins diffusing freely in solution using a homebuilt confocal microscope. The smFRET experiments utilized pulsed interleaved excitation mode, alternating between exciting the donor dye (Alexa 488) with a 488 nm laser and the acceptor dye (Alexa 594) with a 560 nm laser. This technique enabled discrimination between donor-only (i.e., molecules lacking an active acceptor dye) and FRET populations (i.e., molecules labeled with both donor and acceptor dyes).

Considering IBB and ELP are IDPs with rapidly interconverting conformations, models for distance distribution were necessary to interpret the smFRET data. Therefore, we applied two distinct models: the Gaussian chain model and the self-avoiding walk (SAW)-ν model, with the latter being more suitable for good solvent conditions where polymer chains tend to expand (49). We identified the FRET population based on the fluorescence intensity bursts and fit the fluorescence lifetime decay of the FRET population using the respective models to calculate R_E (*SI Appendix*, Figs. S10 and S11). We incorporated an additional correction that considers distance fluctuations of the dye linkers (*Materials and Methods*).

The R_E results for IBB and ELP are presented in Figs. 3*K* and 4*K*, respectively (*SI Appendix*, Table S10 and Figs. S12–S15). For IBB, R_E values obtained using the Gaussian chain model were 5.8 ± 0.5 nm, 6.4 ± 0.5 nm, and 8.3 ± 0.9 nm in 0.5 M, 2 M, and 6 M urea, respectively, while those obtained using the SAW-ν model were 5.9 ± 0.5 nm, 6.5 ± 0.5 nm, and 7.9 ± 0.8 nm. These closely matched R_E values acquired through ASAXS measurements: 6.2 ± 0.7 nm, 7.3 ± 0.8 nm, and 8.7 ± 0.8 nm, respectively. Similarly, for ELP, the R_E values obtained from both techniques were consistent within error in all tested conditions.

Further, we calculated G using R_E from smFRET and R_G from SAXS. As shown in Figs. 3*L* and 4*L* (*SI Appendix*, Table S11) for IBB and ELP, we observed the same trend of change in G compared to the results obtained from ASAXS measurements. The observed agreement between the smFRET and ASAXS results underscores the reliability and precision of these complementary techniques in probing the structural dynamics of biopolymers. This consistency not only validates the accuracy of ASAXS in capturing conformational changes but also highlights its potential as a tool for advancing structural biology.

Discussion

We developed an ncAA and demonstrated its site-specific genetic encoding into proteins. Using this system, we showcased the effectiveness of our genetically encoded ASAXS ruler, which enables simultaneous extraction of both R_G and R_E values for IDPs from a single set of scattering measurements. Our approach eliminates the need to assume a predefined structural model for deriving R_G or R_E. We revealed a decoupling between R_G and R_E in the studied IDPs by varying urea concentration from near-native to highly denaturing conditions. Notably, R_E exhibited greater sensitivity to solvent conditions compared to R_G, consistent with previous findings from FRET and SAXS studies (7, 8). This decoupling of R_E-to-R_G is expected not only for heteropolymers but also for homopolymers under various solvent conditions, with the effect being less pronounced in long homopolymers and block copolymers (7, 44, 48, 51). Accordingly, we observed a less pronounced change in G for ELP, a longer block copolymer, compared to the shorter IBB heteropolymer. Furthermore, our comparison of R_E values measured by ASAXS and smFRET showed high consistency, with deviations remaining within the measurement uncertainties across all urea concentrations tested. This convergence of results from multiple techniques strengthens the credibility of our observations and contributes to a comprehensive understanding of the conformational dynamics exhibited by IDPs under varying denaturant conditions. As a single mutation can alter a protein's properties, dye attachment in FRET can similarly affect protein behavior in some cases. These effects can be identified through functional controls, changes in labeling sites, and dye types. Similarly, many proteins are aggregation-prone, exacerbated by higher concentrations, which can be tested and controlled. However, our findings challenge the idea that the SAXS–FRET discrepancy arises inherently from dye labels in FRET or high protein concentrations in SAXS measurements; rather, it is due to the intrinsic properties of the IDP, such as protein size, sequence motifs and context, solution conditions, and analysis methods, as we and others have suggested previously (7, 8, 12, 13).

The smFRET dataset presented in this paper is included primarily for information and comparative purposes. Our main conclusion—that R_E and R_G are substantially decoupled in heteropolymers—is drawn from ASAXS data. Importantly, our method allows for the direct measurement of R_E from ASAXS without assuming a priori models for data interpretation, diverging fundamentally from predictions of R_E based on molecular form factors (MFFs) in standard SAXS profiles, where hydrophobic-like interactions were first assumed to give rise to the ensembles of heteropolymer conformations (17). SAXS MFFs do not afford unique determinations of R_E, as the ensemble-to-MFF mapping is practically many-to-one (9, 10, 52). Notably, by measuring R_G and R_E by ASAXS on the identical protein constructs under the same denaturant conditions, we ensure that any minor influences from the small Br labels or the denaturants are consistent across all measurements. This consistency minimizes the possibility that the observed decoupling of R_G and R_E is due to the effects of labeling probes or solvent conditions.

The decoupling of R_G and R_E is not just a methodological detail but a crucial insight into IDP behavior. IDPs, unlike folded proteins, remain in highly dynamic states, constantly transitioning between different conformations. This dynamic nature is what allows them to engage in a wide range of interactions within the cell, often in a transient or conditional manner. Therefore, understanding how R_E and R_G respond to environmental changes provides essential information about the physical mechanisms governing IDP function. Knowing and agreeing on R_G and R_E for an IDP is crucial, as these are the most fundamental biophysical parameters to characterize an IDP ensemble. In the absence of a specific fold, R_G or R_E are the simplest parameters to characterize how inter- and intramolecular interactions of the IDP are balanced with interactions of the solvent. This is particularly important to predict how the protein will behave at higher concentrations; for example, a protein that prefers self-interactions is more likely to phase separate than a well-solvated and expanded protein in buffer (53). Due to the prominent role of IDPs in phase separation mechanisms and many neurodegenerative diseases, where strong self-interactions leading to amyloid formation are of concern,

fundamental biophysical parameters must be consistently measured, reported, and universally accepted (54, 55).

The decoupling of R_G and R_E highlights the caution needed when applying polymer theories to IDPs, as interconverting these two measures may not adhere to assumptions like $G = 6$ for a Gaussian chain. Our results also emphasize the importance of using complementary experimental techniques to capture this complexity fully. For example, descriptors like R_G and R_E can be further integrated with, for example, the hydrodynamic radius R_H obtained from pulsed-field-gradient NMR diffusion, fluorescence correlation spectroscopy, and dynamic light scattering experiments (50, 56–60).

Beyond resolving the long-standing SAXS–FRET debate in the IDP community, our genetically encoded ASAXS ruler holds promise for probing the dimensions of various types of proteins. With the potential to encode many more diBrK, our ASAXS ruler could analyze multiple distances, providing a more comprehensive view of protein conformations. While the current method has limitations in protein expression yield and signal-to-background scattering ratio for even longer proteins, future developments in genetic code expansion (61–65) and beamline brilliance could extend its applicability to measure larger distance distributions in protein dynamics, surpassing the limitations of other techniques like FRET, restricted to a maximum spacing of ~10 nm between fluorophores. In summary, our genetically encoded ASAXS ruler emerges as a tool in integrative structural biology.

Materials and Methods

Synthesis of Noncanonical Amino Acid (ncAA) *N*-ε-3,5-Dibromobenzyloxycarbonyl-L-Lysine (diBrK, *SI Appendix*, Fig. S1). 3,5-dibromobenzyl alcohol was dissolved in dimethylformamide (DMF) and 1,1′-carbonyldiimidazole (CDI) was added. This reaction was stirred for 2 h until the starting material was fully consumed. Fmoc-L-Lys-OH, diisopropylethylamine (DIPEA), and a DMF/H2O mixture were then added, and the mixture was stirred at 90 °C for 3 h. Deprotection was performed by adding piperidine and stirring for 1 h, after which the reaction mixture was neutralized with HCl, and the precipitate was filtered and washed with various solvents. The final product was dried overnight under vacuum and analyzed using LC–MS and ^{1}H NMR, confirming its identity and purity. An extended description of diBrK synthesis is provided in *SI Appendix*.

Evolution of Aminoacyl-tRNA Synthetase for diBrK (*SI Appendix*, Fig. S2). A library with 5x NNK mutations on PylRS$^{306A/384F}$ from *M. mazei* was screened. This process included five cycles of positive and negative selection. Initially, the pBK plasmids containing the PylRS$^{306A/384F}$ library were transformed into *E. coli* along with a positive selection plasmid, and plated on Luria Broth (LB) agar with chloramphenicol and diBrK. Library plasmids from grown colonies were then retransformed with a negative selection plasmid and plated on LB agar with arabinose but without diBrK. This cycle was repeated four more times. Selected library plasmids were finally cotransformed with a pALS-GFP^{149TAG} plasmid, and GFP fluorescence was measured to identify the most active variants, which were then sequenced and further validated for GFP expression. An extended description of synthetase evolution is provided in *SI Appendix*.

Recombinant Expression and Purification of Double-diBrK-Labeled IBB and Wild-Type IBB (*SI Appendix*, Fig. S3 *A* and *B*). IBB fragment was cloned into a pQE-14His-TEV vector, with TAG mutations at positions 2S and 72G for the double-diBrK-labeled version. For expression, plasmids containing the evolved tRNA/tRNA synthetase and the IBB fragment were cotransformed into BL21-AI *E. coli* cells, which were grown in LB medium with antibiotics and diBrK, and induced with arabinose and IPTG. Cells were harvested, lysed, and the protein was purified using Ni–NTA affinity chromatography, followed by chitin bead incubation to remove N-terminal truncations. Further purification involved dialysis, a second round of Ni–NTA affinity chromatography, and TEV protease treatment to remove the His-tag, with the final product purified by reversed-phase C18 chromatography. The purity and mass of the proteins were confirmed by SDS-PAGE and mass spectrometry, yielding approximately 250 mg of double-diBrK-labeled IBB from a 500 L culture. An extended description of IBB expression and purification is provided in *SI Appendix*.

Recombinant Expression and Purification of Double-diBrK-Labeled ELP and Wild-Type ELP (*SI Appendix*, Fig. S3 *C* and *D*). ELP fragment (30 repeats of VGVPG) was cloned into a pQE-intein-CBD-14His vector, with TAG mutations at positions 6 V and 149P for the double-diBrK-labeled version. Plasmids containing the evolved tRNA/tRNA synthetase and the ELP fragment were cotransformed into BL21-AI *E. coli* cells, which were grown, induced, and harvested similarly to the IBB procedure. The cells were lysed, and the protein was purified using Ni–NTA affinity chromatography, followed by overnight incubation with β-mercaptoethanol to remove the intein-CBD-14His tag. Further purification included dialysis, a second round of Ni–NTA affinity chromatography, and SEC. The final product was analyzed by SDS-PAGE and mass spectrometry, confirming its purity and mass. Approximately 100 mg of double-diBrK-labeled ELP was obtained from a 300 L culture. An extended description of ELP expression and purification is provided in *SI Appendix*.

Recombinant Expression, Purification, and Fluorescence Labeling of IBB and ELP. To prepare IBB for FRET labeling, the double-cysteine mutant IBB (pQE-14His-TEV-IBB$^{1-73,2cys,72cys}$-intein-CBD) was purified according to the protocols mentioned above. The proteins were reduced with 10 mM DTT, which was then removed via buffer exchanges to a maleimide labeling buffer. Alexa Fluor 488 maleimide and Alexa Fluor 594 maleimide were added in a 2:1 molar ratio and incubated overnight at 4 °C. The reaction was terminated with 10 mM DTT, and excess dyes were removed using SEC. For smFRET measurements of ELP, a mutant with a stop codon (TAG) and a cysteine mutation (6Cys,149TAG) was expressed and purified using similar methods to the double-diBrK-labeled ELP. Cotransformed BL21-AI *E. coli* cells were grown with 1 mM AcF added at OD600 of 0.2. After purification, ELP was labeled with Alexa488-hydroxylamine and Alexa594-maleimide. Excess dyes were removed via SEC. An extended description of protein labeling is provided in *SI Appendix*.

Solution Composition for Standard SAXS and ASAXS Measurements. For both standard SAXS and ASAXS, the measurement solution contained 0.5 M / 2 M / 6 M urea (MP Biomedicals), 0.3 M KCl (Sigma-Aldrich), 5 mM DTT (Biomol), 1 × PBS (Gibco), pH 7.4. The buffers were freshly prepared, filtered with a Steritop filter unit with 0.22 μm pore size (Merck), and degassed.

Standard SEC-SAXS Measurements. Synchrotron radiation X-ray scattering data were collected on the EMBL P12 beamline at the PETRA III storage ring (DESY, Hamburg). Measurements were carried out at 23 °C. The lyophilized protein samples were first dissolved in 4 M GdmCl (Affymetrix), 10x PBS, pH 7.4 to a final concentration of 20 mg/mL and injected into a Superdex 75 Increase 5/150 column at the flow rate of 0.4 mL/min for 0.5 M urea, 0.35 mL/min for 2 M urea, and 0.25 mL/min for 6 M urea. The flow rates were chosen to be as fast as possible to minimize the potential radiation damage without causing the overpressure of the column. The injected sample volume varied from 20 μL to 90 μL to generate a series of SAXS curves at different concentrations. All the SEC runs were repeated at least twice. The column was equilibrated for at least 30 min in between the buffer changes.

Standard SEC-SAXS was performed at a 10,000 eV energy setting, which is the standard user setting at the P12 beamline. Data were recorded using a 6 M PILATUS detector (DECTRIS, Switzerland) at a sample-detector distance of 3.0 m, with a wavelength of 0.124 nm, covering the range of momentum transfer $0.1 < s < 7.1$ nm^{-1} ($s = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle). Preliminary SAXS data reduction and analysis were performed by automatic SASFLOW pipeline (66). Data treatment was carried out using CHROMIXS (67) from the ATSAS package.

1,800 to 2,880 frames were collected with an exposure time of 250 ms per frame. The frames were then selected based on the intensity plot along the elution profile as represented by CHROMIX. The homogeneity of the sample was checked from the calculated R_G (*SI Appendix*, Fig. S6). The selected sample and buffer frames were averaged separately, and the latter was subtracted from the former. The linearity of the Guinier plot extracted from representative SEC-SAXS data is illustrated in a typical example (*SI Appendix*, Fig. S6*C*).

Anomalous SEC-SAXS Measurements. Potassium bromide solution was first measured to determine the Br absorption edge and the X-ray double crystal monochromator energy offset (21). The inflection point was obtained as 13,488 eV,

which is +14 eV above the theoretical value of 13,474 eV (*SI Appendix*, Fig. S4). The inflection point was then verified again with Br-labeled protein solution. Next, five energy points were selected around the theoretical Br edge, including 13428 eV, 13,453 eV, 13,468 eV, 13,472 eV, and 13,476 eV. The measured energy points were adjusted according to the obtained offset.

Structural Parameters for SAXS Analysis. The single particle intensity after spherical averaging can be expressed using the Guinier approximation (68) for the low s range ($sR_G < 1$) as

$$I(s) \approx I(0)\exp\left(-\frac{R_G^2 s^2}{3}\right), \quad [4]$$

where R_G is the radius of gyration of the particle and $I(0)$ is the forward scattering that depends on the electron density distribution within the particle, it is related to the shape of the particle and provides an indication of its compactness. The R_G can also be assessed from the particle distance distribution function as

$$R_G = \frac{\int_0^{D_{max}} r^2 p(r)\mathrm{d}r}{2\int_0^{D_{max}} p(r)\mathrm{d}r}. \quad [5]$$

The function $p(r)$ is the distribution of distances between volume elements inside the particle weighted by their excess scattering densities and it is related to the scattering intensity $I(s)$ by an inverse Fourier transform:

$$p(r) = \frac{r^2}{2\pi^2}\int_0^{\infty} s^2 I(s)\frac{\sin sr}{sr}\mathrm{d}s. \quad [6]$$

ASAXS Data Analysis. Detailed description of the ASAXS data analysis is extensively discussed in the previously published work (21). Briefly, the calibration of the angular axis involved using the powder diffraction pattern of silver behenate measured at a specific energy, typically at the edge of the element of interest. For each energy, the angular scale was appropriately adjusted to take into account the change in the wavelength and the recorded intensities of the SAXS signal were scaled according to the procedure described in (20). Initial data reduction including the s-range adjustment and normalization to the transmitted beam for absolute intensity at each energy was incorporated into the SASFLOW pipeline used routinely at the P12 beamline (66). Additionally, all the SAXS profiles have been corrected for the fluorescence.

Error Analysis. R_G values were initially calculated using the Guinier approximation with a linear fit. To account for the sensitivity to point selection and intensity values, the DATRESAMPLE software (28) from the ATSAS package was used to generate resampled SAXS curves. The decomposed ASAXS curves were resampled 1,000 times, varying the window of points used to calculate R_G. Histograms of R_G values were generated for each curve and combined to understand the potential variation. Error propagation principles were applied to calculate the overall error in the ratio $G = (R_E/R_G)^2$. An extended description of error analysis is provided in *SI Appendix*.

smFRET Measurements. smFRET measurements were conducted using a custom-built spectroscope with a 60×, 1.27 NA water immersion objective. IBB and ELP proteins, labeled with Alexa 488 and Alexa 594, were measured in native or denaturing buffers containing 10 mM DTT. The detection involved freely diffusing labeled proteins using alternating laser excitations, and photon signals were collected and analyzed using the PAM software package (69). Burst analysis provided uncorrected FRET efficiency and stoichiometry histograms, separating donor-only and FRET populations for further time-resolved lifetime fitting. FRET populations were analyzed using Gaussian chain and SAW-ν models to derive R_E. An extended description of smFRET measurements and analysis is provided in *SI Appendix*.

Data, Materials, and Software Availability. All data are available in the main text or *SI Appendix*. All plasmids are available via a cost-free academic material transfer agreement. The experimental SAXS data are deposited into SASBDB database (70) see deposition codes in *SI Appendix*, Table S3.

ACKNOWLEDGMENTS. We thank Simone Mosna, Christine Köhler, Giorgia Celetti, Sabine Zoernack, and Nike Heinß for their contributions to this work. Furthermore, we thank all the members of the Lemke laboratory and Melissa Gräwert from EMBL Hamburg for helpful discussions. We also thank Matthias Ballauff for critical input on our work. We thank the Proteomics Core Facility at the EMBL Heidelberg for expert assistance. E.A.L. and D.I.S. acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG, project number 432343117). E.A.L. acknowledges funding from CRC1551 "Polymer concepts in cellular function" (project number 464588647) and thanks EU infrastructure project iNEXT-Discovery (871037) funding for beamline access. M.Y. acknowledges an MSCA Individual Fellowship (TFNUP 89410) and M.Y. and G.S. Humboldt Research Fellowships for Postdoctoral Researchers. T.S. acknowledges an EMBO Postdoctoral Fellow ALTF 1020-2020.

Author affiliations: [a]Biocenter, Johannes Gutenberg University Mainz, Mainz 55128, Germany; [b]Institute of Molecular Biology postdoctoral program, Mainz 55128, Germany; [c]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg 69117, Germany; [d]BIOSAXS Group, European Molecular Biology Laboratory, Hamburg Unit, Hamburg 22607, Germany; and [e]Institute of Molecular Biology, Mainz 55128, Germany

Author contributions: M.Y., A.Y.G., H.R., T.S., D.I.S., and E.A.L. designed research; M.Y., A.Y.G., H.R., and T.S. performed research; A.C., S.G., A.S.A.M., J.C., and P.F.S. contributed new reagents/analytic tools; M.Y., A.Y.G., H.R., T.S., D.I.S., and E.A.L. analyzed data; and M.Y., A.Y.G., H.R., T.S., D.I.S., and E.A.L. wrote the paper.

The authors declare no competing interest.

1. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell. Biol.* **6**, 197–208 (2005).
2. C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).
3. R. K. Das, K. M. Ruff, R. V. Pappu, Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol* **32**, 102–112 (2015).
4. Edward A. Lemke *et al.*, Intrinsic disorder: A term to define the specific physicochemical characteristic of protein conformational heterogeneity. *Mol. Cell* **84**, 1188–1190 (2024).
5. D. Johansen, J. Trewhella, D. P. Goldenberg, Fractal dimension of an intrinsically disordered protein: Small-angle X-ray scattering and computational study of the bacteriophage λ N protein. *Protein Sci.* **20**, 1955–1970 (2011).
6. J. A. Riback *et al.*, Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **358**, 238–241 (2017).
7. G. Fuertes *et al.*, Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E6342–E6351 (2017).
8. A. Borgia *et al.*, Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J. Am. Chem. Soc.* **138**, 11714–11726 (2016).
9. J. Song, G.-N. Gomes, T. Shi, C. C. Gradinaru, H. S. Chan, Conformational heterogeneity and FRET data interpretation for dimensions of unfolded proteins. *Biophys. J.* **113**, 1012–1024 (2017).
10. J. Song, J. Li, H. S. Chan, Small-angle X-ray scattering signatures of conformational heterogeneity and homogeneity of disordered protein ensembles. *J. Phys. Chem. B* **125**, 6451–6478 (2021).
11. K. M. Ruff, A. S. Holehouse, SAXS versus FRET: A matter of heterogeneity? *Biophys. J.* **113**, 971–973 (2017).
12. R. B. Best *et al.*, Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water". *Science* **361**, eaar7101 (2018).
13. G. Fuertes *et al.*, Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water". *Science* **361**, eaau8230 (2018).
14. J. A. Riback *et al.*, Response to comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water". *Science* **361**, eaar7949 (2018).
15. R. B. Best, Emerging consensus on the collapse of unfolded and intrinsically disordered proteins in water. *Curr. Opin. Struct. Biol.* **60**, 27–38 (2020).
16. I. Reinartz, M. Weiel, A. Schug, FRET dyes significantly affect SAXS intensities of proteins. *Isr. J. Chem.* **60**, 725–734 (2020).
17. J. A. Riback *et al.*, Commonly used FRET fluorophores promote collapse of an otherwise disordered protein. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8889–8894 (2019).
18. H. B. Stuhrmann, Anomalous small angle scattering. *Q. Rev. Biophys.* **14**, 433–462 (1981).
19. R. C. Miake-Lye, S. Doniach, K. O. Hodgson, Anomalous x-ray scattering from terbium-labeled parvalbumin in solution. *Biophys. J.* **41**, 287–292 (1983).
20. T. Zettl *et al.*, Absolute intramolecular distance measurements with angstrom-resolution using anomalous small-angle X-ray scattering. *Nano Lett.* **16**, 5353–5357 (2016).
21. AYu. Gruzinov *et al.*, Anomalous SAXS at P12 beamline EMBL Hamburg: Instrumentation and applications. *J. Synchrotron Rad.* **28**, 812–823 (2021).
22. N. Dingenouts *et al.*, Counterion distribution around a spherical polyelectrolyte brush probed by anomalous small-angle X-ray scattering. *Macromolecules* **37**, 8152–8159 (2004).
23. I. Nikić *et al.*, Debugging eukaryotic genetic code expansion for site-specific click-PAINT super-resolution microscopy. *Angew. Chem. Int. Ed.* **55**, 16172–16176 (2016).
24. T. Yanagisawa *et al.*, Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode Nε-(o-Azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187–1197 (2008).

25. T. Ishida, K. Kinoshita, PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* **35**, W460–W464 (2007).
26. M. H. Høie *et al.*, NetSurfP-3.0: Accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res.* **50**, W510–W515 (2022).
27. G. Erdős, M. Pajkos, Z. Dosztányi, IUPred3: Prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297–W303 (2021).
28. K. Manalastas-Cantos *et al.*, ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **54**, 343–355 (2021).
29. P. B. Moore, Small-angle scattering: Information content and error analysis. *J. Appl. Crystallogr.* **13**, 168–175 (1980).
30. D. Svergun, C. Barberato, M. H. J. Koch, *CRYSOL* – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
31. D. I. Svergun, M. H. J. Koch, P. A. Timmins, R. P. May, *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules* (Oxford University Press, 2013).
32. M. A. Graewert *et al.*, Adding size exclusion chromatography (SEC) and light scattering (LS) devices to obtain high-quality small angle X-ray scattering (SAXS) data. *Crystals* **10**, 975 (2020).
33. N. R. Hajizadeh, D. Franke, D. I. Svergun, Integrated beamline control and data acquisition for small-angle X-ray scattering at the P12 BioSAXS beamline at PETRAIII storage ring DESY. *J. Synchrotron Rad.* **25**, 906–914 (2018).
34. C. E. Blanchet *et al.*, Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Crystallogr.* **48**, 431–443 (2015).
35. C. M. Jeffries, M. A. Graewert, D. I. Svergun, C. E. Blanchet, Limiting radiation damage for high-brilliance biological solution scattering: Practical experience at the EMBL P12 beamline PETRAIII. *J. Synchrotron Rad.* **22**, 273–279 (2015).
36. E. A. Merritt, Data from "X-ray Absorption Edges", http://skuld.bmsc.washington.edu/scatter/AS_periodic.html. Accessed 15 November 2024.
37. S. Brennan, P. L. Cowan, A suite of programs for calculating x-ray absorption, reflection, and diffraction performance for a variety of materials at arbitrary wavelengths. *Rev. Sci. Instrum.* **63**, 850–853 (1992).
38. D. T. Cromer, D. A. Liberman, Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge. *Acta Crystallogr. Sect. A* **37**, 267–268 (1981).
39. D. T. Cromer, D. Liberman, Relativistic calculation of anomalous scattering factors for X rays. *J. Chem. Phys.* **53**, 1891–1898 (1970).
40. D. T. Cromer, Anomalous dispersion corrections computed from self-consistent field relativistic Dirac-Slater wave functions. *Acta Crystallogr.* **18**, 17–23 (1965).
41. B. H. Toby, R. B. Von Dreele, *GSAS-II*: The genesis of a modern open-source all purpose crystallography software package. *J. Appl. Crystallogr.* **46**, 544–549 (2013).
42. M. Sztucki, E. Di Cola, T. Narayanan, Anomalous small-angle X-ray scattering from charged soft matter. *Eur. Phys. J. Spec. Top.* **208**, 319–331 (2012).
43. P. J. Flory, The configuration of real polymer chains. *J. Chem. Phys.* **17**, 303–310 (1949).
44. H. Hofmann *et al.*, Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16155–16160 (2012).
45. A. N. Rissanou, S. H. Anastasiadis, I. A. Bitsanis, Monte carlo study of the coil-to-globule transition of a model polymeric system. *J. Polym. Sci. B Polym. Phys.* **44**, 3651–3666 (2006).
46. G. Koren *et al.*, Intramolecular structural heterogeneity altered by long-range contacts in an intrinsically disordered protein. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2220180120 (2023).
47. S. Roberts, M. Dzuricky, A. Chilkoti, Elastin-like polypeptides as models of intrinsically disordered proteins. *FEBS Lett.* **589**, 2477–2486 (2015).
48. G. Zhang, C. Wu, "Folding and formation of mesoglobules in dilute copolymer solutions" in *Conformation-Dependent Design of Sequences in Copolymers I, Advances in Polymer Science*, A. R. Khokhlov, Ed. (Springer, Berlin Heidelberg, 2006), pp. 101–176.
49. W. Zheng *et al.*, Inferring properties of disordered chains from FRET transfer efficiencies. *J. Chem. Phys.* **148**, 123329 (2018).
50. G. Tesei *et al.*, Conformational ensembles of the human intrinsically disordered proteome. *Nature* **626**, 897–904 (2024).
51. B. Hammouda, SANS from homogeneous polymer mixtures: A unified overview. *Polymer Characteristics* **106**, 87–133 (1993).
52. K. Ghosh, J. Huihui, M. Phillips, A. Haider, Rules of physical mathematics govern intrinsically disordered proteins. *Annu. Rev. Biophys.* **51**, 355–376 (2022).
53. E. W. Martin *et al.*, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
54. A. Zbinden, M. Pérez-Berlanga, P. De Rossi, M. Polymenidou, Phase separation and neurodegenerative diseases: A disturbance in the force. *Dev. Cell* **55**, 45–68 (2020).
55. S. Alberti, D. Dormann, Liquid–liquid phase separation in disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).
56. N. Kassem *et al.*, Order and disorder–An integrative structure of the full-length human growth hormone receptor. *Sci. Adv.* **7**, eabh3805 (2021).
57. B. Schuler, A. Soranno, H. Hofmann, D. Nettels, Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu. Rev. Biophys.* **45**, 207–231 (2016).
58. M. Kar *et al.*, Phase separating RNA binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2202222119 (2022).
59. G.-N.W. Gomes *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).
60. S. Naudi-Fabra, M. Tengo, M. R. Jensen, M. Blackledge, S. Milles, Quantitative description of intrinsically disordered proteins using single-molecule FRET, NMR, and SAXS. *J. Am. Chem. Soc.* **143**, 20109–20121 (2021).
61. J. Fredens *et al.*, Total synthesis of *Escherichia coli* with a recoded genome. *Nature* **569**, 514–518 (2019).
62. M. J. Lajoie *et al.*, Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
63. F. J. Isaacs *et al.*, Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011).
64. C. D. Reinkemeier, E. A. Lemke, Dual film-like organelles enable spatial separation of orthogonal eukaryotic translation. *Cell* **184**, 4886–4903.e21 (2021).
65. M. Yu *et al.*, Visualizing the disordered nuclear transport machinery in situ. *Nature* **617**, 162–169 (2023).
66. D. Franke, A. G. Kikhney, D. I. Svergun, Automated acquisition and analysis of small angle X-ray scattering data. *Nucl. Instrum. Methods Phys. Res., Sect. A* **689**, 52–59 (2012).
67. A. Panjkovich, D. I. Svergun, CHROMIXS: Automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944–1946 (2018).
68. A. Guinier, G. Fournet, *Small-Angle Scattering of X-rays* (Wiley, 1955).
69. W. Schrimpf, A. Barth, J. Hendrix, D. C. Lamb, PAM: A framework for integrated analysis of imaging, single-molecule, and ensemble fluorescence data. *Biophys. J.* **114**, 1518–1528 (2018).
70. A. G. Kikhney, C. R. Borges, D. S. Molodenskiy, C. M. Jeffries, D. I. Svergun, SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci.* **29**, 66–75 (2020).