Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments

Andrea Soranno[1,2], Iwo Koenig[1], Madeleine B. Borgia, Hagen Hofmann, Franziska Zosel, Daniel Nettels, and Benjamin Schuler[2]

Department of Biochemistry, University of Zurich, 8057 Zurich, Switzerland

Intrinsically disordered proteins (IDPs) are involved in a wide range of regulatory processes in the cell. Owing to their flexibility, their conformations are expected to be particularly sensitive to the crowded cellular environment. Here we use single-molecule Förster resonance energy transfer to quantify the effect of crowding as mimicked by commonly used biocompatible polymers. We observe a compaction of IDPs not only with increasing concentration, but also with increasing size of the crowding agents, at variance with the predictions from scaled-particle theory, the prevalent paradigm in the field. However, the observed behavior can be explained quantitatively if the polymeric nature of both the IDPs and the crowding molecules is taken into account explicitly. Our results suggest that excluded volume interactions between overlapping biopolymers and the resulting criticality of the system can be essential contributions to the physics governing the crowded cellular milieu.

single-molecule FRET | unfolded state collapse |
excluded volume screening | Flory–Huggins theory

A surprisingly large number of eukaryotic proteins either contain substantial unstructured regions or are entirely unfolded under physiological conditions (1, 2). These "intrinsically disordered proteins" (IDPs) are involved in many crucial cellular processes, such as transcription, translation, and signal transduction; their functional and conformational properties are thus of great interest for a wide range of biological questions. Important advances in understanding the structures of IDPs have been made over the past decade, especially with spectroscopic techniques, e.g., NMR (3, 4), single-molecule fluorescence (5–7), and with atomistic and coarse-grained molecular simulations (8–10). In contrast with the stable folded structures we are familiar with from 50 y of structural biology, IDPs comprise highly heterogeneous and dynamic ensembles of conformations, which either lack stable tertiary structure altogether or fold only on binding their cellular targets (4). Important components of the cellular environment that affect IDPs include not only specific cellular ligands, but also pH and the concentration of salts (11, 12). An additional contribution that has been difficult to investigate experimentally comes from the large number of different solutes present in a cell that do not interact with an IDP specifically, but result in an environment that is densely filled with macromolecules and metabolites (12–14). Given their lack of persistent structure, the conformations of IDPs are expected to be particularly sensitive to the effects of such molecular crowding. Indeed, first experiments indicate that some IDPs gain structure upon crowding (15), whereas others do not (16–18), but may change their dimensions (19–21). The question of how the conformational distributions of IDPs respond to crowded environments is of particular current interest because IDPs have a vital role in cellular compartments and regions with very high local concentrations of proteins and RNA, such as RNA granules and nuclear pore complexes (22–25). However, a quantitative comprehension of how the concentrations and sizes of the molecular crowding agents (or "crowders") affect IDPs is currently incomplete (26), especially for polymeric crowders. Here we use single-molecule spectroscopy to investigate the

influence of crowding on the conformational distributions of IDPs, as a step toward a quantitative framework of how the polydisperse cellular environment affects these highly flexible molecules.

Single-molecule fluorescence detection in combination with Förster resonance energy transfer (FRET) is a method highly suited for addressing this question (5–7, 11, 27, 28) because it allows the heterogeneous structural ensemble of suitably labeled IDPs to be probed even in the presence of very large concentrations of unlabeled solutes. To investigate the physical principles underlying the crowding effects on IDPs, we study a selection of IDPs representative of the naturally occurring sequence compositions in combination with a broad range of molecular sizes of crowding agents. We primarily use polyethylene glycol (PEG) as a crowding agent. This uncharged polymer with high solubility in aqueous solution (29) (*SI Appendix*, Fig. S1) is available from monomeric ethylene glycol to degrees of polymerization of almost 1,000 (*SI Appendix*, Table S1) at sufficient purity for single-molecule experiments up to physiologically realistic volume fractions of crowder of ~40% (30). PEG is widely used for biomedical applications (31) and for mimicking inert crowding agents (13, 26). Previous work has shown that the conformational properties of IDPs strongly depend on their amino acid sequence composition and charge patterning (8, 11, 27, 28, 32–34). Here we investigate the effect of crowding on four different IDP sequences that span a broad range of net charge per residue and average hydrophobicity (Fig. 1 and *SI Appendix*, Table S2): the N- and C-terminal segments of human prothymosin-α (ProTα-N and -C), the binding domain of the activator for thyroid hormones and retinoid receptors (ACTR), and the N-terminal domain of HIV-1 integrase (IN). Whereas ProTα is highly charged and does not assume a folded structure under any known conditions, ACTR and IN are

Significance

In the interior of a cell, the volume accessible to each protein molecule is restricted by the presence of the large number of other macromolecules. Such a crowded environment is known to affect the stability and folding rates of proteins. In the case of intrinsically disordered proteins (IDPs), however, a class of proteins that lack stable structure, much less is known about the role of crowding effects. We have quantified the conformational changes occurring in IDPs in the presence of high concentrations of different polymers that act as crowding agents. Using single-molecule spectroscopy, we have identified effects that are typical of polymer solutions and have direct implications for the behavior of IDPs within the cell.

Fig. 1. Mean net charge versus mean hydrophobicity per residue for the four disordered protein sequences used in this study: the C- and N-terminal segments of prothymosin α, ProTα-C (blue) and ProTα-N (green), respectively (complete sequence: black); the activator for thyroid hormones and retinoid receptors, ACTR (orange); and the N-terminal domain of the HIV-1 integrase, IN (red). Folded structures refer to the conformations of ACTR and IN in presence of their ligands, NCBD (gray structure) and Zn^{2+} (light gray sphere). The FRET labeling sites (*SI Appendix*, Table S1) are indicated by colored spheres. The dashed gray line indicates the boundary between intrinsically disordered and folded proteins proposed by Uversky (32). Note that the contributions to the net charge from the fluorescent dyes are included (11).

representatives of the classes of IDPs that fold upon binding a protein or a small ligand, respectively.

Results

Quantifying Crowder-Induced Chain Compaction with Single-Molecule FRET.
To probe the intramolecular distance distributions of the IDPs, we attached Alexa Fluor 488 and Alexa Fluor 594 as donor and acceptor fluorophores via cysteine residues introduced at suitable positions, with sequence separations of 55 (ProTα-N), 54

(ProTα-C), 72 (ACTR), and 49 residues (IN) (*SI Appendix*, Table S1). Fig. 2 shows examples of confocal single-molecule FRET experiments performed at increasing concentrations of PEG 6000 (i.e., PEG with a molecular mass of ~6,000 Da; *SI Appendix*, Table S2). Up to three peaks are observed in the transfer efficiency (E) histograms from measurements of labeled IDPs freely diffusing in solution. The peak at $E \sim 0$ results from molecules lacking an active acceptor dye and is not of interest here. The second peak at intermediate E corresponds to the disordered state. The appearance of a third peak at $E \sim 0.7$ and $E \sim 0.9$ for ACTR and IN, respectively, results from the formation of a folded structure in complex with their ligands, the nuclear coactivator binding domain (NCBD) and a Zn^{2+} ion, respectively (*SI Appendix*). This separation of subpopulations is essential for distinguishing the effects of solutes on the conformational distributions within the disordered state from a cooperative transition to a folded state. In the case of IN, our experiments indicate the formation of a small folded population at high PEG concentrations even in the absence of Zn^{2+} (Fig. 2D and *SI Appendix*, Fig. S2), but for all other proteins, only an unfolded population is present (Fig. 2 and *SI Appendix*, Fig. S2). However, with increasing concentration of PEG 6000, three of the four disordered sequences (ProTα-C, ProTα-N, and ACTR) exhibit a clear shift of the peak corresponding to the disordered state toward higher transfer efficiencies, indicating an overall tendency of these proteins to collapse in the presence of crowding agents. For the case of IN, which has the least charged and most hydrophobic sequence (Fig. 1), only very small changes in transfer efficiency are noticeable, clearly demonstrating that molecular crowding does not affect all IDPs equally. Given the importance of intramolecular electrostatic repulsion for their conformations (11, 33), it may seem surprising that the more highly charged IDPs exhibit a more pronounced collapse.

The changes in transfer efficiency of the IDPs induced by the crowding agents can be used to extract information on the corresponding changes in chain conformations. Following previous work on unfolded proteins (35) and IDPs (27, 28), we use a Flory–Fisk distribution, which provides a description of the underlying distance distributions, to quantify the dimensions of the polypeptide chains in terms of mean-squared intramolecular distances or the effective radii of gyration, R_g, of the segments

Fig. 2. Single-molecule FRET can be used for quantifying the compaction of disordered proteins by molecular crowding. Representative FRET efficiency histograms at different volume fractions of polyethylene glycol (PEG) 6000 for ProTα-C (A, blue), ProTα-N (B, green), ACTR (C, orange), and IN (D, red). Histograms of ACTR and IN in the presence of their respective interaction partners, NCBD (C) and Zn^{2+} (D), are shown for comparison. Gaussian and lognormal distributions are used to fit the data (solid lines). The transfer efficiency peaks from molecules lacking an active acceptor dye are shaded in gray. At the highest volume fractions of PEG, some broadening of the peaks is observed due to the increasing fluorescent background. Only IN exhibits a small crowder-induced population at the transfer efficiency of the folded state (see *SI Appendix*, Fig. S2 for detailed controls). (E) The resulting radii of gyration (R_g) for ProTα-C (blue circles), ProTα-N (green triangles), ACTR (orange rhombi), and IN (red squares) illustrate the PEG-induced compaction. Fits (solid lines) are obtained using scaled-particle theory (*SI Appendix*, Eq. S5) with the size of PEG 6000 as a single, globally adjustable fitting parameter. The precision of the values of R_g as estimated from multiple measurements of selected data points is comparable to or smaller than the size of the symbols.

probed by the FRET pair (*SI Appendix*). Note that the analysis is robust with respect to the polymer-physical model used and that the use of multiparameter detection allows us to exclude possible interfering artifacts, such as insufficient rotational averaging of the fluorophores or quenching of the dyes (*SI Appendix*).

Fig. 2*E* shows examples of the resulting changes in R_g as a function of the volume fraction ϕ of PEG 6000 for the four IDP sequences, all of which exhibit collapse upon crowding. Between 0% and 40% of crowder, the changes in R_g range from 0.2 nm (or ∼10% for IN to ∼1 nm (or ∼30%) for ProTα-C. Qualitatively, this is the behavior expected even from a simple hard-sphere model for a crowding agent whose steric repulsion of the IDP chains leads to their compaction (13, 36). A commonly used quantitative framework for such effects is scaled-particle theory (37), which provides an estimate of the change in free energy required for creating a cavity equivalent to the size of the IDP in a solution of hard spheres with a radius corresponding to the size of the crowding agent, R_g^{crd} (*SI Appendix*). If we apply scaled-particle theory, a remarkably good fit is achieved with R_g^{crd} as a global fit parameter (Fig. 2*E*). However, the resulting value for R_g^{crd} of (6.2 ± 0.1) nm is almost twice the measured radius of gyration of PEG 6000 (*SI Appendix,* Fig. S1), signifying that a hard-sphere description is not adequate for polymeric crowding agents such as PEG (38).

Crowder Size Variation Reveals the Importance of Polymer Effects. To identify the origin of this discrepancy, we choose a strategy orthogonal to varying the volume fractions of crowder and probe the influence of different sizes of crowding agents on the compaction of IDPs. Fig. 3 shows the complete data set for all four IDP sequences with PEGs of 10 different degrees of polymerization, P, at volume fractions from 0% to ∼40%. For all IDPs, we observe the tendency to collapse with increasing crowder concentration, but interestingly, the degree of compaction is highly dependent on crowder size. The characteristic behavior is most apparent if we consider the change in R_g of an IDP as a function of P at a fixed volume fraction of PEG, as illustrated in Fig. 4 for ProTα-C with $\phi = 15\%$. The IDPs collapse monotonically as the crowder size increases, but their R_g reaches a plateau for PEGs of more than ∼100 monomers. Notably, this behavior is the opposite of what we expect from scaled-particle theory because the free energy cost for creating a cavity of given size decreases with increasing crowder size (*SI Appendix*); in other words, larger solid-sphere crowding agents have larger interstitial cavities and would thus accommodate expanded IDPs more easily (Fig. 4*A*). To illustrate the discrepancy, Fig. 4*E* shows the resulting prediction for $R_g(P)$ based on scaled-particle theory (solid black line, Fig. 4*E*).

An obvious deficit of scaled-particle theory for the treatment of unfolded proteins is the assumption that the crowders cannot penetrate the unfolded chain. To address this issue, Minton proposed the "Gaussian cloud" model (37) (Fig. 4*B*), where the unfolded protein is described in terms of a continuous Gaussian distribution of monomer density around the center of mass of the protein (*SI Appendix*, Fig. S3). Small solid-sphere crowders can pervade this protein cloud and thus have little effect on the density distribution of the chain. With increasing crowder size, the probability of accommodating the corresponding spheres without steric clashes with the chain decreases, leading to a compaction of the IDP, in agreement with experimental observation (solid gray line, Fig. 4*E*). For very large crowding agents, however, this penetration probability decreases further, and ultimately the limit of classic scaled-particle theory is recovered, in contrast with the experimental observation.

These results strongly suggest that we need to go a step further and take into account the polymeric nature of both IDP and crowding agent to explain the behavior observed experimentally. The simplest realistic model needs to comprise two polymers of different lengths in good solvent, i.e., a ternary system. Note that both the IDPs (24) and the crowder (*SI Appendix*, Fig. S2) (26)
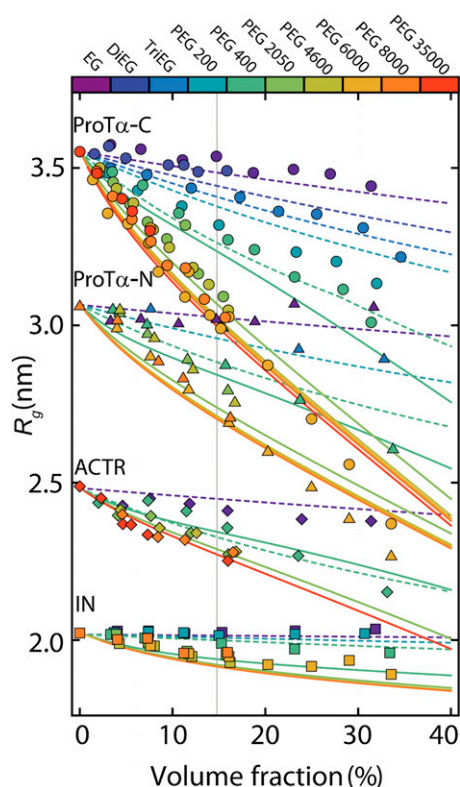

Fig. 3. Both increasing volume fraction and increasing crowder size lead to IDP compaction. Radii of gyration of ProTα-C (circles), ProTα-N (triangles), ACTR (rhombi), and IN (squares) as a function of the volume fraction of PEG obtained from single-molecule FRET experiments. Fits to the data corresponding to the short-chain regime (dashed lines, Eq. **1a**) and the long-chain regime (solid lines, Eq. **1b**) are shown. For the case of PEG 400, both types of fits are reported to illustrate the cross-over between the two regimes. The vertical dashed line indicates the volume fraction of 15% PEG used in Fig. 4.

exhibit the scaling behavior characteristic of polymers in good solvent, which justifies this assumption.

We also need to take into consideration that, unlike the hard spheres assumed in scaled-particle theory, polymer chains can interpenetrate. This aspect becomes most relevant above a limiting volume fraction, referred to as the overlap concentration ϕ^*, where the solution can be thought of as being filled by nonintersecting spheres of the size of a single polymer chain. For volume fractions greater than ϕ^*, the transition between dilute and semidilute regimes occurs, and the chains start to overlap, which will affect the conformations of the polymers (*SI Appendix,* Fig. S1). ϕ^* depends only on the length P of the polymers and on the scaling exponent in the appropriate solvent regime ($\phi^* = P^{-4/5}$ in good solvent; *SI Appendix*); for long chains, this semidilute regime is reached already at volume fractions of a few percent (*SI Appendix,* Fig. S1) and the interpenetration of the chains must thus be taken into account for the majority of our experimental conditions.

Within the framework of the commonly used Flory–Huggins theories, we therefore need to distinguish two scenarios under our experimental conditions: the short-chain regime (Fig. 4*C*) and the long-chain regime (Fig. 4*D*) (39). In the first case, the crowding polymer chains are short and consequently remain below the overlap concentration. The system can thus be depicted as a dilute ($\phi < \phi^*$) solution of PEG chains of radius R_g^{crd} that do not overlap with each other but are able to pervade the volume explored by the IDP (Fig. 4*C*) (39). Inside this volume, the degrees of freedom of the crowders are reduced by the IDP, and the crowder chains will gain entropy by leaving this volume. A further increase in entropy of the crowder molecules results from reducing the volume occupied by the protein. In
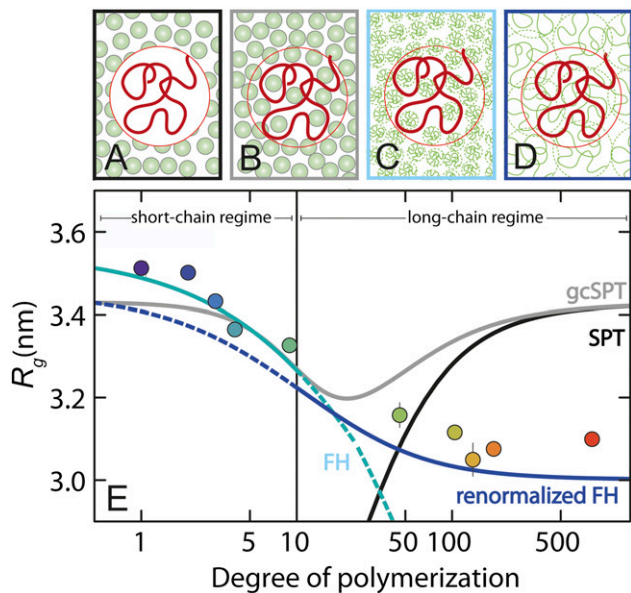
Fig. 4. Polymer concepts explain the compaction of IDPs by crowding agents of increasing size. Graphical representation of (*A*) scaled-particle theory (SPT), (*B*) Gaussian cloud model (gcSPT), (*C*) Flory–Huggins theory (FH) in the short-chain regime, and (*D*) renormalized Flory–Huggins theory (renormalized FH) in the long-chain regime. (*E*) Radius of gyration of ProTα-C as a function of the degree of polymerization of PEG at 15% volume fraction of crowding agent. The data points were obtained from linear interpolation of the volume fraction dependences shown in Fig. 3 (same color code for the PEG size). Fits according to the different theories are shown as black (SPT), gray (gcSPT), cyan (FH theory), and blue (renormalized FH theory) lines. Solid lines indicate the regime for which the respective theories were derived; outside of these regimes, dashed lines are used. Error bars reporting on the precision of the experiments are calculated as 1 SD from the linear fits of data for each PEG series in Fig. 3 (uncertainties are smaller than the size of the symbols unless shown explicitly).

$$R_g(N, P, \phi, a) = R_{g0}\left(\frac{1}{1 + a\phi/\phi^*(P)}\right)^{1/5} \quad \text{for} \ \ P < N^{1/2}, \quad \textbf{[1a]}$$

and

$$R_g(N, P, \phi, s_{NP}) = R_{g0} f(N, P, \phi, s_{NP}) \quad \text{for} \ \ P \geq N^{1/2}, \quad \textbf{[1b]}$$

where R_{g0} is the radius of gyration of the IDP in the absence of crowding; a is an empirical parameter that can account for differences in the solvent quality for the different proteins and interactions between protein and polymer (45) (*SI Appendix*); s_{NP} quantifies the interaction between the protein and the polymer chains; and f is a function that represents the renormalization mapping (*SI Appendix*). It is worth emphasizing that Eqs. **1a** and **1b** contain only a single adjustable parameter each, a and s_{NP}, respectively (*SI Appendix*, Table S4 and Fig. S4). The equations provide a good fit to the experimental data, including the approach to a limiting value of R_g for IDPs in very large crowders (Fig. 4*E*). In fact, the entire data set for all four IDP sequences is described remarkably well by a global fit (Fig. 3). The success of this approach supports the hypothesis that the polymeric properties of both IDP and PEG are essential for understanding the effect of molecular crowding, and that the criticality of the solution cannot be neglected. Considering the highly polydisperse cellular environment, it seems probable that related effects will be prominent in vivo and that mean-field descriptions are insufficient for a quantitative description of crowding in the cell.

The Balance of Hard-Core Repulsion and Other Nonspecific Interactions. Recent experimental results indicate that the presence of weak, nonspecific attractive interactions in the heterogeneous cellular environment can modulate or even dominate the effects of hard-core repulsion that are at the basis of molecular crowding (46–48). The role of such "chemical interactions" is a subject of debate also for proteins and PEG (13, 26). Notably, the approach presented here (Eq. **1b**) allows the relative contributions of hard-core repulsion and other interactions to be quantified in terms of the interaction parameter s_{NP}. In the cases investigated here, the analysis with Eq. **1b** indicates that a small contribution of unfavorable interactions with PEG is present for ProTα and ACTR, and no such interactions are detected in the case of IN (*SI Appendix*, Table S4). We note, however, that even though interactions such as nonspecific attraction between crowder and IDP can modulate the amplitude of the change in R_g with crowder concentration (*SI Appendix*, Fig. S4), the polymeric effects dominate the overall behavior.

An independent means of interrogating the role of nonspecific charge and hydrophobic interactions is to add salt or denaturants to the solution. Fig. 5 shows that neither 1 M KCl nor 4 M guanidinium chloride (GdmCl) nor 4 M urea impedes the collapse of ProTα. The value of R_{g0} depends on ionic strength and denaturant concentration owing to the known charge screening and/or denaturant-induced chain expansion (11). However, the dependence of R_g on the volume fraction of PEG is described by Eq. **1b** with the same values of s_{NP} as in the absence of salt or denaturant, just by rescaling R_{g0} to the value at the corresponding KCl, GdmCl, or urea concentrations without crowder, suggesting that the effect of additional interactions on the compaction of the IDP is small. Finally, we tested the influence of different chemical structures of the crowding polymer in experiments with dextran, polyvinyl alcohol (PVA), and polyvinylpyrrolidone (PVP) (Fig. 5). Even though we could measure these solutions only for volume fractions of crowder of up to 10% owing to fluorescent impurities, in all cases we observed a collapse of ProTα-C similar to that in PEG. The resulting values of s_{NP} for dextran, PVA, and PVP are significantly lower than for PEG (*SI Appendix*, Table S5), indicating better compatibility—or less unfavorable interactions—with ProTα, but the collapse of the IDP is preserved. In summary, the polymeric crowding effects on IDPs observed here are dominated
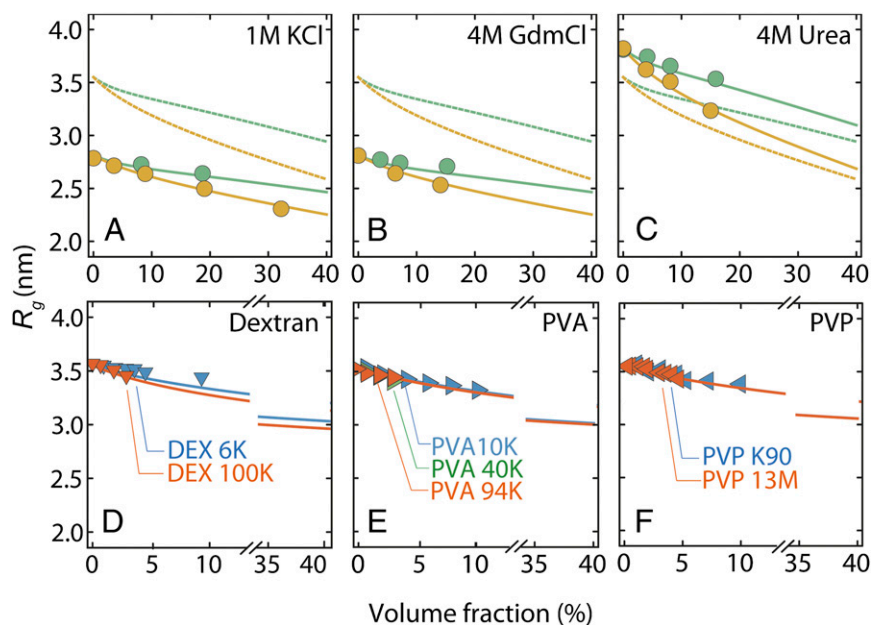
other words, the requisite equality of chemical potentials for crowders inside and outside the volume pervaded by the IDP predicts a collapse of the protein chain (39), similar to the Gaussian cloud model, and in good agreement with the experimental data (cyan line, Fig. 4*E*; also see *SI Appendix*). In the long-chain regime, however, this mean-field theory fails and diverges from the measured results. In this regime, the crowding polymers are often above their overlap concentrations, and their conformations are influenced by mutual interpenetration. In contrast with the case of a single chain in good solvent, where the dimensions are dominated by repulsive interactions between the monomers, the interpenetration by other crowders in the semidilute regime causes a screening of these repulsive interactions within each chain (40, 41). This excluded volume screening will also affect the conformations of the IDP. However, because the polymers have dimensions comparable to or larger than the protein, they will only partially penetrate the IDP. Under these conditions, the ternary system is close to a critical point and can exhibit density fluctuations over a broad range of length scales due to interactions within the protein, within the crowders, and between the crowders and the protein (41, 42). Many critical systems, ranging from the liquid–gas phase transition near the critical point to the magnetization near the Curie point of a ferromagnet and the Kondo effect of electrons in metals, have been successfully described by renormalization group theory (43). The same approach has provided fundamental insights into the scaling invariance for polymer solutions (41). Here we adopt a renormalized Flory–Huggins-type theory developed by Schäfer and Kappeler (44) for a multicomponent system in the long-chain regime. We thus analyzed the data in the short-chain and long-chain regimes according to

Fig. 5. Variation of solution conditions and crowding agents suggest the importance of nonspecific effects on IDP compaction. Radius of gyration of ProTα-C versus the volume fraction of PEG 400 (green circles) and PEG 6000 (yellow circles) in (*A*) 1 M KCl solution, (*B*) 4 M GdmCl, and (*C*) 4 M urea. Fits according to Eq. **1b**, assuming a different value of R_{g0} but the same value of s_{NP} as in Figs. 3 and 4, are shown as green and yellow solid lines for PEG 400 and PEG 6000, respectively. Fits for the same crowding agents in the absence of salt or denaturant (Fig. 3) are included as dashed lines with corresponding colors. The effects of molecular crowding with dextran (*D*), PVA (*E*), and PVP (*F*) on ProTα-C are shown for different sizes of these alternative crowders as indicated. Lines represent the fit to the renormalized FH theory (*SI Appendix*, Table S5) and are extrapolated up to 40% volume fraction for comparison with other polymers.

by hard-core repulsion between the monomers and the resulting excluded volume screening (40, 41), indicating a phenomenon of generic relevance. However, the analysis presented here does allow additional interactions to be included that can modulate the crowding effect.

Discussion

Eqs. **1a** and **1b** can account for the dependence of R_g on crowder concentration and crowder size for all four IDPs investigated (Fig. 3). The question remains, however, why the extent of crowder-induced compaction is so different for the different IDPs. Polymer theory offers an interesting explanation. According to the Flory theorem, the chains in a melt (i.e., in the absence of solvent) of compatible polymers approach their Θ-state. Under these conditions, because of the screening of excluded volume interactions within and between the polymers, the dimensions of the chains scale approximately with the square root of the number of chain segments, and a characteristic radius of gyration $R_{g\Theta}$ is observed (*SI Appendix*). Recent work indicates that $R_{g\Theta}$ for the IDPs investigated here is in the range of ~1.7–2.0 nm (27) (*SI Appendix*). The results in Fig. 3 for the larger PEGs are indeed consistent with asymptotic convergence of R_g for all of the IDP variants toward values in this range in the limit of very high volume fractions of crowder, i.e., under conditions that approach the situation of a melt. In other words, highly expanded IDPs with dimensions much greater than $R_{g\Theta}$ (such as ProTα) are expected to undergo more pronounced compaction on polymeric crowding than those IDPs that are close to $R_{g\Theta}$ already in the absence of crowders (such as IN). Based on the empirical relations between solvent quality and average net charge obtained previously (27), we estimate that ~90% of all IDPs are above the θ-state in the absence of crowding (*SI Appendix*) and should thus be susceptible to compaction by polymeric crowders.

The observations reported here could thus have implications for the functional properties of many IDPs, e.g., for the capture radius for their cellular targets in the framework of a fly-casting mechanism (49, 50) and for the folding propensity of denatured ensembles in the crowded cellular environment (13). However, the balance of the different contributions may be subtle. Whereas a compaction of the chain by crowding will result in a decrease of the capture radius, it will increase the translational diffusion coefficient. These opposing effects will modulate the basic influence of crowding on solution viscosity and the concomitant changes in association rates (51). Similarly, the established effects of crowding

on the stability of the folded and/or bound states of IDPs (13) may be affected by changes in unfolded state dimensions. Single-molecule experiments of the type presented here may help to dissect these contributions quantitatively. Complementary simulations of polymeric crowding could provide valuable insights into the underlying molecular mechanisms.

We note that a substantial fraction of crowding in the cell is due to polymeric molecules such as peptides, nucleic acids, polysaccharides, or other disordered proteins. However, the extent of crowding is strongly affected by the spatial organization of the cell. A remarkable example of very high local concentrations of IDPs are nucleoporins, which line the nuclear pore complexes (25). We estimate the volume fraction occupied by nucleoporins to be between 25% and 55% of the volume available in the pore, about an order of magnitude greater than the overlap concentration (*SI Appendix*). Similarly, IDPs involved in RNA granules (22, 24) or analogous nonmembrane-bound bodies with liquid-like properties (23, 24) are likely to exceed their overlap concentration locally (*SI Appendix*). Under these conditions, polymer effects characteristic of the semidilute regime will be highly relevant for the conformations of IDPs and for the occurrence of possible phase transitions. Interestingly, ProTα often colocalizes with dense speckles such as promyelocytic leukemia bodies (52). Given its abundance in the nucleus of mammalian cells and its high mobility within and near the nucleus (53), we expect that a compaction similar to what we observed here can occur in vivo. According to our results, the dense local environment resulting from liquid–liquid demixing (23, 24) or sol–gel transitions (22) should strongly influence the conformational distributions of IDPs, with consequent impact on the functional properties of the resulting assemblies and their mechanisms of formation. Flory–Huggins theories as used here might thus provide novel insights into the demixing of multicomponent polymeric systems (41). An interesting next step will be a direct comparison of experiments in vitro with intracellular measurements (14, 26), and the required quantitative tools are beginning to emerge (54–56).

Methods

Proteins were expressed, purified, and labeled similar to previous reports (11, 27, 28). Single-molecule measurements were performed using a MicroTime 200 confocal microscope equipped with a HydraHarp 400 counting module (PicoQuant). For details on experiments and theory, see *SI Appendix*.

1. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6(3):197–208.
2. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of in-herently disordered proteins. *Curr Opin Struct Biol* 18(6):756–764.
3. Jensen MR, Ruigrok RW, Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* 23(3):426–435.
4. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19(1):31–38.
5. Ferreon AC, Moran CR, Gambin Y, Deniz AA (2010) Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods Enzymol* 472:179–204.
6. Schuler B, Müller-Späth S, Soranno A, Nettels D (2012) Application of confocal single-molecule FRET to intrinsically disordered proteins. *Methods Mol Biol* 896:21–45.
7. Ferreon AC, Ferreon JC, Wright PE, Deniz AA (2013) Modulation of allostery by pro-tein intrinsic disorder. *Nature* 498(7454):390–394.
8. Mao AH, Lyle N, Pappu RV (2013) Describing sequence-ensemble relationships for intrinsically disordered proteins. *Biochem J* 449(2):307–318.
9. Lyle N, Das RK, Pappu RV (2013) A quantitative measure for protein conformational heterogeneity. *J Chem Phys* 139(12):121907.
10. Fisher CK, Stultz CM (2011) Protein structure along the order-disorder continuum. *J Am Chem Soc* 133(26):10022–10025.
11. Müller-Späth S, et al. (2010) From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107(33):14609–14614.
12. Uversky VN (2009) Intrinsically disordered proteins and their environment: Effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J* 28(7-8):305–325.
13. Zhou HX, Rivas GN, Minton AP (2008) Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annu Rev Bio-phys* 37:375–397.
14. Gershenson A, Gierasch LM (2011) Protein folding in the cell: Challenges and prog-ress. *Curr Opin Struct Biol* 21(1):32–41.
15. Dedmon MM, Patel CN, Young GB, Pielak GJ (2002) FlgM gains structure in living cells. *Proc Natl Acad Sci USA* 99(20):12681–12684.
16. McNulty BC, Young GB, Pielak GJ (2006) Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. *J Mol Biol* 355(5):893–897.
17. Munishkina LA, Cooper EM, Uversky VN, Fink AL (2004) The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. *J Mol Recognit* 17(5):456–464.
18. Szasz CS, et al. (2011) Protein disorder prevails under crowded conditions. *Bio-chemistry* 50(26):5834–5844.
19. Hong JA, Gierasch LM (2010) Macromolecular crowding remodels the energy land-scape of a protein by favoring a more compact unfolded state. *J Am Chem Soc* 132(30):10445–10452.
20. Mikaelsson T, Adén J, Johansson LBA, Wittung-Stafshede P (2013) Direct observation of protein unfolded state compaction in the presence of macromolecular crowding. *Biophys J* 104(3):694–704.
21. Johansen D, Jeffries CM, Hammouda B, Trewhella J, Goldenberg DP (2011) Effects of macromolecular crowding on an intrinsically disordered protein characterized with small-angle neutron scattering with contrast matching. *Biophys J* 100(4):1120–1128.
22. Han TW, et al. (2012) Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell* 149(4):768–779.
23. Li PL, et al. (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483(7389):336–340.
24. Brangwynne CP (2011) Soft active aggregates: Mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. *Soft Matter* 7(7):3052–3059.
25. Rout MP, et al. (2000) The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J Cell Biol* 148(4):635–651.
26. Elcock AH (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr Opin Struct Biol* 20(2):196–206.
27. Hofmann H, et al. (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 109(40):16155–16160.
28. Soranno A, et al. (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc Natl Acad Sci USA* 109(44):17800–17806.
29. Devanand K, Selser JC (1991) Asymptotic-behavior and long-range interactions in aqueous-solutions of poly(ethylene oxide). *Macromolecules* 24(22):5943–5947.
30. Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and ex-cluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol* 222(3):599–620.
31. Harris JM (1992) *Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Ap-plications* (Plenum, New York).
32. Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins un-structured under physiologic conditions? *Proteins* 41(3):415–427.
33. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107(18):8183–8188.
34. Das RK, Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci USA* 110(33):13392–13397.
35. 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.
36. Cheung MS, Klimov D, Thirumalai D (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc Natl Acad Sci USA* 102(13):4753–4758.
37. Minton AP (2005) Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: Macromolecular crowding and protein stability revisited. *Biophys J* 88(2):971–985.
38. Mittal J, Best RB (2010) Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys J* 98(2):315–320.
39. Joanny JF, Grant P, Pincus P, Turkevich LA (1981) Conformations of polydisperse polymer-solutions - bimodal distribution. *J Appl Phys* 52(10):5943–5948.
40. Edwards SF (1966) Theory of polymer solutions at intermediate concentration. *Proc Phys Soc Lond* 88(560P):265–280.
41. Schäfer L (1999) *Excluded Volume Effects in Polymer Solutions as Explained by the Renormalization Group* (Springer, Berlin).
42. Tran HT, Pappu RV (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. *Biophys J* 91(5):1868–1886.
43. Wilson KG (1983) The renormalization-group and critical phenomena. *Rev Mod Phys* 55(3):583–600.
44. Schäfer L, Kappeler C (1993) Interaction effects on the size of a polymer-chain in ternary solutions - a renormalization-group study. *J Chem Phys* 99(8):6135–6154.
45. Nose T (1986) Chain dimension of a guest polymer in the semidilute solution of compatible and incompatible polymers. *J Phys (Paris)* 47(3):517–527.
46. Minton AP (2013) Quantitative assessment of the relative contributions of steric re-pulsion and chemical interactions to macromolecular crowding. *Biopolymers* 99(4):239–244.
47. Sarkar M, Li C, Pielak GJ (2013) Soft interactions and crowding. *Biophys. Rev.* 5(2):187–194.
48. Kim YC, Mittal J (2013) Crowding induced entropy-enthalpy compensation in protein association equilibria. *Phys Rev Lett* 110(20):208102-1–208102-5.
49. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc Natl Acad Sci USA* 97(16):8868–8873.
50. Trizac E, Levy Y, Wolynes PG (2010) Capillarity theory for the fly-casting mechanism. *Proc Natl Acad Sci USA* 107(7):2746–2750.
51. Schreiber G, Haran G, Zhou HX (2009) Fundamental aspects of protein-protein asso-ciation kinetics. *Chem Rev* 109(3):839–860.
52. Vareli K, Frangou-Lazaridis M, van der Kraan I, Tsolas O, van Driel R (2000) Nuclear distribution of prothymosin alpha and parathymosin: Evidence that prothymosin al-pha is associated with RNA synthesis processing and parathymosin with early DNA replication. *Exp Cell Res* 257(1):152–161.
53. Enkemann SA, Ward RD, Berger SL (2000) Mobility within the nucleus and neigh-boring cytosol is a key feature of prothymosin-alpha. *J Histochem Cytochem* 48(10):1341–1355.
54. Gelman H, Platkov M, Gruebele M (2012) Rapid perturbation of free-energy land-scapes: From in vitro to in vivo. *Chemistry* 18(21):6420–6427.
55. Phillip Y, Kiss V, Schreiber G (2012) Protein-binding dynamics imaged in a living cell. *Proc Natl Acad Sci USA* 109(5):1461–1466.
56. Sakon JJ, Weninger KR (2010) Detecting the conformation of individual proteins in live cells. *Nat Methods* 7(3):203–205.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Supporting Information

Material and Methods

Protein preparation and labeling. Cysteine residues for the specific labeling of IDPs using maleimide chemistry were introduced by site-directed mutagenesis at the positions given in Table S1. ProTα (57) variants, IN (58) and ACTR (59) were produced in E. coli BL21 with an N-terminal His-tag for purification. Cells were grown in LB medium and expression was induced with IPTG.

Disruption of harvested cells and ammonium sulfate precipitation were carried out as described previously for ProTα (60). The protein was bound to a gravity flow Ni-NTA column (Thermo Scientific) and eluted with 20 mM Tris, 100 mM sodium chloride, and 500 mM imidazole, pH 7.5. The sample was dialyzed against 20 mM Tris, 100 mM sodium chloride, and 20 mM imidazole, pH 7.5. The His-tag was cleaved off with HRV 3C protease, which also contained a His-tag; a second Ni-NTA chromatography run was used to remove the protease and the His-tag. The flow-through containing ProTα was concentrated with a YM-3 Centriprep centrifugal filter (Merck Millipore), reduced with 10 mM TCEP, and purified by reversed phase (RP) HPLC on a Reprosil Gold 200 column (Dr. Maisch, Germany) and elution with an acetonitrile gradient. Purified ProTα was lyophilized in a SpeedVac concentrator (Thermo Scientific), dissolved in 0.1 M sodium phosphate, pH 7.5, and labeled with Alexa Fluor 488 maleimide (Invitrogen) at a molar ratio of dye to protein of 0.8:1. The reaction mixture was quenched with β-mercaptoethanol, reduced with 10 mM TCEP and purified by RP-HPLC on an XTerra C18 column (Waters). The fraction containing singly labeled ProTα was lyophilized in a SpeedVac concentrator, and labeled and purified analogously with a molar excess of Alexa Fluor 594 maleimide (Invitrogen). The masses of the doubly labeled ProTα-N and ProTα-C were confirmed by electrospray ionization mass spectrometry (ESI-MS).

IN was purified by Ni-NTA chromatography as described for ProTα. The His-tag was removed by enzymatic cleavage with Thrombin protease after dialysis against 50 mM NaHCO₃ pH 9.3, 0.5 M Na₂SO₄, 0.1 M NaCl, 5 mM EDTA, 1 mM DTT. After reduction with 10 mM TCEP and purification by RP-HPLC, IN was lyophilized in a SpeedVac concentrator, dissolved in 50 mM HEPES, 0.5 mM ZnCl₂, and 0.5 M arginine, pH 7.5, and labeled with Alexa Fluor 488 maleimide at a molar ratio of dye to protein of 0.8:1. The reaction mixture was quenched with β-mercaptoethanol, reduced with 10 mM TCEP and purified by RP-HPLC on a XTerra C18 column. The fraction containing singly labeled IN was lyophilized in a SpeedVac concentrator, and labeled and purified analogously with Alexa Fluor 594 maleimide. The doubly labeled IN was purified by RP-HPLC on an XTerra C18 column. The correct mass of the labeled IN was confirmed by ESI-MS.

ACTR was co-expressed with NCBD (nuclear co-activator binding domain of CREB) to improve the stability of ACTR during expression (59). The harvested cells were disrupted with a TS 1.1 cell disruption system (Constant Systems Ltd, England), the protein was bound to a Ni-NTA column, and the His-tag was removed by enzymatic cleavage with HRV 3C protease. After reduction of the sample with β-mercaptoethanol, the protease and the His-tag were removed with a second Ni-NTA column. ACTR was separated from NCBD by RP-HPLC on a Reprosil Gold 200 column by elution with an acetonitrile gradient. The fraction containing ACTR was lyophilized in a SpeedVac concentrator, dissolved in 0.1 M sodium phophate, pH 7.5, and labeled with Alexa Fluor 488 maleimide at a molar ratio of dye to protein of 0.8:1. The reaction mixture was again reduced with β-mercaptoethanol and purified by RP-HPLC on a Reprosil Gold 200 column. The fraction with the singly labeled ACTR was lyophilized in a SpeedVac concentrator,

dissolved in labeling buffer and labeled with a molar excess of Alexa Fluor 594 maleimide. Doubly labeled ACTR was purified by RP-HPLC on a Reprosil Gold 200 column, and the correct mass was confirmed by ESI-MS.

Analogously, the same IDP variants were labeled with a different FRET pair (ATTO 532 and ATTO 647N, Atto-Tec, Germany) whose spectra are shifted towards higher wavelengths where the influence of fluorescent impurities on the transfer efficiency histograms at high PEG concentrations is strongly reduced.

Preparation of crowding solutions. Crowding experiments were carried out in 50 mM sodium phosphate buffer, pH 7.0. Crowding solutions were prepared by mixing acidic (50 mM NaH_2PO_4 + crowding agent) and alkaline (50 mM Na_2HPO_4 + crowding agent) stock solutions to a final pH of 7.0 (+/- 0.2). The concentration of the stock solutions depended on the solubility and purity of the crowding agent used (between 10 % w/w and 40 % w/w crowding agent). Crowding solutions with lower concentrations were prepared by dilution of the corresponding stock solution to the desired concentration with 50 mM sodium phosphate buffer, pH 7.0.

Single-molecule fluorescence spectroscopy. Single-molecule fluorescence measurements were performed with a MicroTime 200 confocal microscope (PicoQuant, Germany) equipped with a diode laser (LDH-D-C-485, PicoQuant, Germany), a 20 MHz supercontinuum laser (SC-450-4, Fianium, UK; wavelength selected with a z582/15 band pass filter (Chroma)) for pulsed interleaved excitation (PIE) (61), and an Olympus UplanApo 60x/1.20W objective (Olympus). Photons emitted from the sample were collected by the same objective. Remaining excitation light was eliminated by a filter (HQ500LP, Chroma Technology) before the emitted photons passed the confocal unit with a 100 μm pinhole. The emitted photons were separated into four channels with a polarizing beam splitter and a dichroic mirror (585DCXR, Chroma). Donor photons were filtered (ET525/50m, Chroma Technology) and then focused on a τ-SPAD avalanche photodiode (PicoQuant). Acceptor photons were filtered (HQ650/100m, Chroma Technology) and detected by a τ-SPAD (PicoQuant). The arrival time of every detected photon was recorded with a HydraHarp 400 counting module (PicoQuant).

All measurements were performed by exciting the donor dye with a laser power of 100 μW at the back aperture of the objective. For PIE measurements, the power used for exciting the acceptor dye was adjusted to match the intensity of the donor emission (between 50 and 70 μW). Single-molecule FRET efficiency histograms were acquired in samples with protein concentrations of about 50 pM to 100 pM. The time between excitation pulse and photon detection was stored with 16 ps resolution, with the lasers pulsed at a repetition rate of 20 MHz. The measurements were performed in 50 mM sodium phosphate buffer, pH 7.0, 200 mM β-mercaptoethanol, and 0.001% Tween-20 with varying concentrations of crowding agents and/or denaturants (guanidinium chloride, urea) or potassium chloride. Each sample was measured for 30 min to 1h at 295 K.

Data Analysis

FRET efficiency histograms. Fluorescence bursts from individual molecules were identified by combining successive photons separated by inter photon times of <100 μs and retaining the burst if the total number of photons detected after donor excitation was >50. Transfer efficiencies for each burst were calculated according to $E=n_A/(n_A+n_D)$, where n_D and n_A are the numbers of donor and acceptor photons, respectively. Corrections for background, acceptor direct excitation, channel crosstalk, differences in detector efficiencies, and quantum yields of the dyes were applied (6). The precision of the measurements as estimated from multiple independent measurements is typically ±0.01 transfer efficiency units and thus comparable to or smaller than

the data points reported in the figures, unless shown explicitly. We estimate a systematic error for R_g of less than ±0.2 nm for the entire dataset. However, a uniform shift of this magnitude for all data toward higher or lower values of R_g does not affect any of our conclusions.

The changes in refractive index caused by increasing concentrations of crowding agents were measured with a digital Abbe refractometer (Krüss, Germany) and were used to recalculate the Förster radius (R_0) for every sample under the assumption that the polymeric crowding agents pervade the solutions uniformly. This assumption does not affect our conclusions, since the narrow range of refractive indices between 1.34 and 1.39 for our experimental conditions has a minor effect on the dimensions of the proteins. Even if we assumed the extreme case that no refractive index change was experienced by the protein locally upon going from pure buffer to the highest volume fraction of PEG, the lack of a refractive index correction would correspond to an apparent compaction of the protein between 0.05 nm (at transfer efficiency of 0.8) and 0.1 nm (at transfer efficiency of 0.3) relative to the values reported here.

Fluorescence lifetimes and anisotropy. Multiparameter detection allows us to exclude possible interfering artifacts, such as insufficient rotational averaging of the fluorophores or quenching of the dyes (62). The dependence of the fluorescence lifetimes on transfer efficiencies determined for each burst (Fig. S5) was compared with the behavior expected for fixed distances and for a chain sampling a broad distribution of distances. For a fixed distance, r, the mean donor lifetime in the presence of acceptor is given by $\tau_{DA}(r) = \tau_D (1-E(r))$, where τ_D is the lifetime in the absence of acceptor, and $E(r) = 1/(1+r^6/R_0^6)$. For a chain with a dye-to-dye distance distribution $P(r)$, the donor lifetime is $\tau_{DA} = \int_0^\infty tI(t)dt / \int_0^\infty I(t)dt$, where $I(t) = I_0\int_0^\infty P(r)e^{-t/\tau_{DA}(r)}dr$ is the time-resolved fluorescence emission intensity following donor excitation. Donor and acceptor lifetimes at different concentrations of crowding agents were analyzed by fitting subpopulation-specific time-correlated photon counting histograms after donor and acceptor excitation, respectively. This allows us to also examine the dependence of donor and acceptor lifetimes on the solution conditions. A systematic decrease of both donor and acceptor lifetimes of up to 10 % was observed with increasing concentrations of crowding agents. These lifetime changes are consistent with the changes expected according to the Strickler-Berg equation (63) for solutions with different refractive indices. Since the changes in donor and acceptor lifetimes are very similar, the contribution of this effect to the observed transfer efficiencies cancels. The variation of the donor lifetime also has no significant impact on R_0, since the donor quantum yield enters into the calculation of R_0 with the power of 1/6, resulting in a maximum change in R_0 of about 2%, less than the statistical experimental uncertainty.

Subpopulation-specific anisotropies were determined for both donor and acceptor, and values were found to vary between 0.03 and 0.08 for the donor and between 0.11 and 0.18 for the acceptor, consistent with values observed in ensemble measurements, and sufficiently low to assume as a good approximation for the orientational factor $\kappa^2 = 2/3$.

Quantifying the radius of gyration from transfer efficiencies. Essentially as described previously (27), FRET efficiencies are converted to radii of gyration according to

$$E = \int_0^{l_c} E(r) \int_{R_c}^{l_c/2} P(r \mid R_g) P_{FF}(R_g) dR_g dr, \qquad \text{(Eq. S1)}$$

3

where l_c is the contour length, and R_c is the radius of the sphere with volume equivalent to the sum of the volumes V_{aa}^i of all the amino acids, $R_c = \left(\dfrac{3}{4\pi} \sum_i V_{aa}^i \right)^{1/3}$. We use the conditional probability density function for a certain end-to-end distance, r, given the radius of gyration, R_g, suggested by Ziv and Haran (35), which describes the distance distribution of two random points inside the sphere of radius $\sqrt{5} \cdot r_g$ ($\sqrt{5}$ is a scaling factor used to satisfy the condition $6\langle R_g^2 \rangle = \langle r^2 \rangle$),

$$P\left(r \mid R_g\right) = \frac{1}{\sqrt{5}R_g} \left[3\left(\frac{r}{\sqrt{5}R_g} \right)^2 - \frac{9}{4}\left(\frac{r}{\sqrt{5}R_g} \right)^3 + \frac{3}{16}\left(\frac{r}{\sqrt{5}R_g} \right)^5 \right], 0 \le r < 2\sqrt{5}R_g \quad \text{(Eq. S2)}$$

$P_{FF}(R_g)$ is the Flory-Fisk distribution for the radius of gyration

$$P_{FF}(R_g) = Z^{-1} R_g^6 e^{-\frac{7R_g^2}{2\langle R_g^2 \rangle}}, \tag{Eq. S3}$$

where $\langle R_g^2 \rangle^{1/2}$ is the root-mean-squared radius of gyration of the chain, and Z is the normalization term. In contrast to previous works, we do not apply corrections to the Flory-Fisk distribution introduced by Sanchez theory (35, 64, 65), since the applied weighting proposed by Sanchez would be valid only for the measurements in absence of crowders and would be inconsistent with the other experimental conditions considered here (see *Scaled-particle theory* section). However, using the Sanchez distribution would result only in a systematic shift of all radii of gyration by approximately 0.1-0.2 nm, which does not affect any conclusions of this work.

Scaled-particle theory (SPT). Following the approach proposed by Minton (37), the effect of macromolecular crowding on the unfolded state can be quantified by weighting the probability density function of the radius of gyration of the disordered ensemble, $P(R_g)$, according to the chemical potential, $\Delta\mu = \Delta\mu(\phi, R_g, R_g^{crd})$, obtained with SPT:

$$\langle R_g^2 \rangle = \frac{\int R_g^2 P(R_g) e^{-\frac{\Delta\mu(\phi, R_g, R_g^{crd})}{k_B T}} dR_g}{\int P(R_g) e^{-\frac{\Delta\mu(\phi, R_g, R_g^{crd})}{k_B T}} dR_g}. \tag{Eq. S4}$$

The simplest choice for $P(R_g)$ is the Flory-Fisk distribution (Eq. S3), where $\langle R_g^2 \rangle^{1/2}$ is substituted by the root-mean-squared radius of gyration of the chain in absence of crowding agents, $\langle R_{g0}^2 \rangle^{1/2}$. If we assume that both the unfolded protein and the crowding agent can be described as rigid spheres (as in the classic SPT), the excess chemical potential of the IDP in the presence of crowding can be written as:

$$\frac{\Delta\mu(\phi, R_g, R_g^{crd})}{k_B T} = -\log(1-\phi) + \left[\overline{R}^3 + 3\overline{R}^2 + 3\overline{R}\right]\frac{\phi}{1-\phi} + \left(2\overline{R}^3 + \frac{9}{2}\overline{R}^2\right)\left(\frac{\phi}{1-\phi}\right)^2 + 3\overline{R}^3\left(\frac{\phi}{1-\phi}\right)^3 , \text{(Eq. S5)}$$

where $\overline{R} = \dfrac{R_g}{R_g^{crd}}$; ϕ is the volume fraction of crowding agent in solution; and R_g^{crd} the radius of gyration of the crowder. Consequently, an increase in ϕ or R_g lead to an increase in the excess chemical potential, whereas an increase in R_g^{crd} causes a decrease in the excess chemical potential. Data reported in Fig. 2e are fitted globally to Eqs. S4 and S5 with a different $\left\langle R_{g0}^2 \right\rangle^{1/2}$ for each IDP and a single effective R_g^{crd} as an adjustable parameter shared by all four proteins (Fig. S6).

A modification of the classic SPT suggested by Minton (37) allows the polymeric nature of the IDP to be taken into account within the SPT framework. Here the unfolded state is represented as a Gaussian cloud where the average number density of residues can be described as a function of the distance from the center of the mass of the protein, r_p, as

$$\rho(r_p) = n\frac{3}{2\pi\left\langle R_g^2 \right\rangle^{1/2}}e^{-\frac{3r_p^2}{2\left\langle R_g^2 \right\rangle}} \tag{Eq. S6}$$

For a system consisting of the protein and a single sphere in solution, r_p can be expressed in terms of the distance between the centers of mass of the two objects, r_{sep}, and the radius of the sphere, which in this case is the radius of the crowding agent, R_g^{crd} (see Fig. S3). The probability, P_0, that no chain segments of the IDP lie within the volume of an arbitrarily placed hard sphere is calculated as a function of R_g of the IDP,

$$\log P_0 = -\left(\frac{6}{\pi}\right)^{1/2}\frac{n}{R_g r_{sep}}e^{-\frac{3r_{sep}^2}{2R_g^2}}\int_0^{R_g^{crd}}e^{-\frac{3r^2}{2R_g^2}}\sinh\frac{3r_{sep}r}{R_g^2}r\,dr . \tag{Eq. S7}$$

Rescaling all distances relative to the size of the crowding agent, R_g^{crd}, yields

$$r_{sep} = f_{sep}R_g^{crd} \tag{Eq. S8}$$
$$R_g = f_g R_g^{crd} \tag{Eq. S9}$$
$$r = xR_g^{crd} . \tag{Eq. S10}$$

P_0 can be written in terms of the rescaled dimensions as

$$\log P_0 = -\left(\frac{6}{\pi}\right)^{1/2}\frac{n}{f_g f_{sep}}e^{-\frac{3f_{sep}^2}{2f_g^2}}\int_0^1 e^{-\frac{3x^2}{2f_g^2}}\sinh\frac{3f_{sep}x}{f_g^2}x\,dx . \tag{Eq. S11}$$

5

By equating the co-volume of the Gaussian cloud and the crowding agent with the co-volume of two hard spheres, an equivalent effective hard-sphere radius, R_g^{eff}, is obtained for each R_g sampled by the IDP, leading to

$$\overline{R}^{eff} = \frac{R_g^{eff}}{R_g^{crd}} = 3\int_0^{\infty}\left(\left(1-P_0\left(f_{sep}\right)\right)f_{sep}^2 df_{sep}\right)^{1/3} - 1 \qquad \text{(Eq. S12)}$$

This rescaled $\overline{R}^{eff}$ can then be inserted in Eq. S5 (Fig. S7).

Further extensions of SPT. In efforts to go beyond the simple description of a fluid of hard spheres, different extensions of the SPT have been implemented. SPT equations have been revised to account for ellipsoidal, cylindrical and infinite rod-like particles (37, 66-68). None of these corrections for different shapes can account for the trends observed in our experimental results. More recently, Qin and Zhou (69) have approached the problem of crowding on IDPs by calculating the co-volume on the basis of explicit simulations of the disordered protein and of the crowding agent in isolation. Even though this post-processing approach captures the effects induced by compact crowders on disordered proteins in their simulations, an extension of this method to polymeric expanded crowders as those used in our experiments has not yet been implemented. Complications in applying this approach come from the difficulty of calculating the correct co-volume between two disordered systems as well as taking into account the change in volume of the polymeric crowder at high concentrations (see the discussion about semidilute regime in the section *Flory-Huggins theories*).

To account for interactions between the crowders and the protein (47), attractive free energy terms have been included in SPT (46, 48), which resulted in the successful description of simulated data (48). A similar approach can be implemented here by adding an attractive interaction between IDP and crowder in Eq. S4. However, even with the functional form suggested by Kim & Mittal (48), the energy parameter would need to be different for each polymer length and possibly for different concentrations to obtain a quantitative fit of our data.

Flory-Huggins theories. A single polymer chain in good solvent adopts swollen conformations and follows a scaling exponent of 3/5, i.e. $R_g^{crd} \sim P^{3/5}$, where P is the number of Kuhn segments of the polymer. Three different concentration regimes need to be distinguished for a polymer in solution: the dilute regime, where the polymer chains are not overlapping; a semidilute regime, where the chains start to overlap and entangle; and a dense regime, where the chains are highly packed (Fig. S1 inset). ϕ^* is the overlap concentration, which separates the semidilute from the dilute regime. ϕ^* can be defined as the concentration of polymer where the volume fraction of the polymer chains in solution is equal to the volume fraction of a single polymer chain, i.e.

$$\phi^* = \frac{Pb^3}{P^{9/5}b^3} = P^{-4/5}. \qquad \text{(Eq. S13)}$$

Flory argued that in concentrated solutions and melts, the polymers exhibit the length scaling of an ideal chain ($R \sim P^{1/2}$) (70). Let us consider the case of one long chain with N segments (the IDP) in a polymer melt of shorter chains with P segments (the crowder). For simplicity, the segment length, b, is assumed to be equal for N-chain and P-chains. By equating the chemical

potentials of crowders inside and outside the volume pervaded by the long chain, a relation between the end-to-end distance, R, of the long N-chain and the number of segments of the P-chains can be obtained (39):

$$\frac{R}{Nb^2} - \frac{N^2 b^3}{PR^4} - \frac{1}{R} = 0 \qquad \text{(Eq. S14)}$$

If $P << N$, the third term in the equation can be neglected, and the size of the long chain can be described by the equation

$$R = N^{3/5} P^{-1/5} b. \qquad \text{(Eq. S15)}$$

In the case of large P, the second term in Eq. S14 can be neglected, and ideal scaling is recovered:

$$R = N^{1/2} b. \qquad \text{(Eq. S16)}$$

The crossover between small and large P is determined by equating Eqs. S15 and S16, which results in the threshold given by the Flory criterion: $P=N^{1/2}$ (39).

In our experiments, we investigate a **ternary system** composed of
- individual test chains with N segments (the IDP),
- a volume fraction, ϕ, of polymer chains with P segments (the crowding agent),
- and the solvent.

Similar to the case of the polymer melt (39), it is possible to describe the interaction between the long chain (the IDP) and the other polymers (the crowding agent) in terms of an effective medium interaction parameter. This term is obtained in the mean field approach of Joanny et al. (39) by equating the chemical potentials of the short chains (crowders) inside and outside the long chain (IDP). The effective medium interaction parameter is then given by:

$$u = \frac{1}{1 + P\phi} \qquad \text{(Eq. S17)}$$

To study the effects of the crowder concentration on the size of the IDP, it is helpful to treat the IDP as a sequence of blobs of size R_P. In dilute solution, on length scales smaller than R_P, the N-chain behaves in the same way as the P-chains, whereas on length scales greater than R_P, the ternary properties of the system become relevant. In the latter case, the long chain formed by N/P blobs experiences the effective medium interaction that in rescaled units is (39)

$$u = \frac{1}{1 + \phi/\phi^*}. \qquad \text{(Eq. S18)}$$

For a chain in good solvent, the radius of gyration is

$$R_g = b N^{3/5} u^{1/5} \qquad \text{(Eq. S19)}$$

And thus, in blob-rescaled units,

$$R_g = R_g^{crd} \left(\frac{N}{P} \right)^{3/5} \left(\frac{1}{1+\phi/\phi^*} \right)^{1/5} = R_{g0} \left(\frac{1}{1+\phi/\phi^*} \right)^{1/5} \qquad \text{(Eq. S20)}$$

The equation can be further extended to the semidilute regime (45). However, since this theory is only valid for the case of a long chain in a solution of shorter chains, this regime will only be approached marginally in our experiments.

For the data analysis presented in Figs. 3 and 4, we modified Eq. S20 by introducing a fitting parameter a, similar to the effective interaction term proposed by Nose (45):

$$R_g = R_{g0} \left(\frac{1}{1+a\phi/\phi^*} \right)^{1/5} \qquad \text{(Eq. 1a, see main text)}$$

All other parameters are not adjustable: ϕ^* is given by Eq. S13, and R_{g0} is obtained from the radius of gyration of each protein in the absence of crowding. a corresponds to a correction of the effective medium interaction term in Eq. S17, which is calculated for a very long chain in a bath of shorter chains, where the long chain is in good solvent and assumed to be large enough to accommodate the small chains (Fig. S8). The degree of expansion of the IDPs investigated here depends on their specific sequence (27), resulting in differences in the interaction term. The values of a obtained from the fits reflect the expected trend and exhibit an increasing deviation from $a = 1$ with increasing compactness of the protein (Table S4).

Extending the classical Flory-Huggins theory to the case of $P > N^{1/2}$ requires a realistic estimate of the density fluctuations in the solution when the solution is no longer dilute. A corresponding quantitative description is provided by the renormalized Flory-Huggins theory derived by Schäfer and Kappeler (44). Renormalization group theory establishes equivalence between microscopically different systems through scaling laws. In doing so, it allows to map a system of long chains (where common perturbation theories break down) onto a system of effectively short chains (where perturbation theories hold).

Here we introduce a length l_0, which defines a sort of lattice unit for the system and will be used to describe both the N-chain (the IDP) and the P-chains (the crowding agent). Following the treatment of Schäfer, we set l_0 equal to the length of the Kuhn segment of the N-chain. Here we adopted the same l_0 for all four proteins, according to the value obtained for ProTα. All the other chemical differences in structure and flexibility between the two chains will be absorbed in specific parameters. The renormalization is then introduced via a renormalized length, $l_R = l_0 / \lambda$, where $\lambda < 1$. Similarly, the number of segments of each chain and the volume fraction are rescaled as a function of the same λ parameter:

$$N \rightarrow N_R(\lambda) \qquad \text{(Eq. S21a)}$$
$$P \rightarrow P_R(\lambda) \qquad \text{(Eq. S21b)}$$
$$\phi \rightarrow \phi_R(\lambda). \qquad \text{(Eq. S21c)}$$

The connection between the microscopic information and the renormalized parameter is given by:

$$1 = \frac{1}{N_R} + 2\left(\hat{u}B_N \phi l_0^{-3} N^{1.76} + \hat{u}\left(\frac{B_P}{B_N}\right)^{\frac{1}{0.588}} B_N^3 \phi l_0^{-3} N^{0.76}\right) N^{-0.76}, \qquad \text{(Eq. S22)}$$

where $\hat{u} = 5.756$ is a parameter connected to the overlap probability of chains in the semidilute regime; $B_i = \dfrac{R_{g0}^{(i)}}{0.8^{0.5} i^{0.588}}$, with $i = \{P, N\}$, contains the structural details of the single N- and P-chains in highly dilute solutions under the assumption that both N- and P-chains are in good solvent; and λ is chosen in order that $l_R = B_N \left(\dfrac{N}{N_R}\right)^{0.588}$. The radius of gyration is then given by

$$R_g(N, P, \phi, s_{NP}) = l_R^2 N_R \left\{ 0.636 + 0.165 N_R^{1/2} - 0.292 N_R^{1/2} f_{NP}^2 G(\phi_R P_R, \frac{N_R}{P_R}) \right\} \qquad \text{(Eq. S23)}$$

where $G(W, y) = W \displaystyle\int_0^\infty z^{-1/2} \frac{B(z)D(zy)}{1 + WD(zy)} dz$ is formally similar to the result obtained with the uniform expansion model (71) when the expansions due to the renormalization are condensed in terms of

$$B(z) = \frac{1}{6}e^{-z} - 2D(z) - \frac{8}{z}(D(z) - 1) - \frac{10}{z^2}(D(z) - 1 + \frac{z}{3}) \qquad \text{(Eq. S24)}$$

$$D(z) = \frac{2}{z^2}(e^{-z} - 1 + z). \qquad \text{(Eq. S25)}$$

The term f_{NP} is obtained under the assumption that interactions between N- and P-chains are small (44), leading to the equation $f_{NP}(s_{NP}) = 1 + \left(\dfrac{l_R}{s_{NP}}\right)^{-0.40}$, where s_{NP} is an invariant parameter with the dimensions of a length that quantifies the interactions between the N- and P-chains. f_{NP} can be connected to the Flory interaction parameter, χ, through the equation $|\chi| = |1 - f_{NP}|$. However, the parameter s_{NP} is preferred to χ since the latter is not invariant in the renormalization flow.

Note that in the global fit of the radii of gyration as a function of volume fraction of PEG for all IDPs (Fig. 3) with the renormalized Flory-Huggins theory (Fig. S9), the only adjustable parameter is s_{NP}, which is taken to be identical for all PEG sizes, but allowed to vary from protein to protein (Table S4); B_N and B_P are well-defined by experimental observables: B_N is calculated from the radius of gyration measured for each protein in the absence of crowding and B_P from the reported values of the radii of gyration for PEG (see Fig. S1 and *PEG scaling law*). If s_{NP} is close to zero, protein and crowding agent are indistinguishable in terms of inter- and intramolecular interactions. In Fig. S4a, the robustness of the functional form of the fitting function at different values of the fitting parameter s_{NP} is illustrated.

s_{NP} provides a new opportunity to quantify the interactions between protein and crowder beyond simple excluded volume effects, sometimes referred to as "chemical interactions". The two variants of ProTα are well fitted with almost identical values of s_{NP} (see Table S4), whereas slightly different values were obtained for ACTR and IN, indicating small variations in the

9

interactions between PEG and proteins. The values obtained are of the same order of magnitude as those previously reported in the literature for ternary solutions of synthetic polymers (72). A conversion of s_{NP} to the more intuitive parameter χ yields repulsive interactions in the range of 0.1-0.3 k_BT per segment between PEG and ProTα or ACTR. Interestingly, the fitted values for the experiments with ProTα and other polymers (see Table S5) suggest lower or even undetectable chemical interactions between the protein and the polymers. The contribution to the overall effects observed here is small (Fig. S4), but this trend is consistent with previous suggestions that PEG may be less inert than other polymers commonly employed in crowding experiments (13). One of the strengths of the approach presented here is the possibility to model a complex polymeric solution including such repulsive (or attractive) interactions.

In the application of the Flory-Huggins theories, we have approximated the length of a Kuhn segment by 0.76 nm for both protein and PEG according to previously reported persistence lengths (27, 73), which is equivalent to two bond segments. Consequently, the number of Kuhn segments of the N- and P-chains is given by half the degree of polymerization. In the case of PEG, the concentration in weight fraction was converted to volume fraction according to $\phi_{v/v} = \phi_{w/w} / (\rho(1-\phi_{w/w})+\phi_{w/w})$ where $\rho = 1.12$ g/cm^3 is an average density for pure solutions of short PEGs. This approach is justified since no significant volume contraction is reported for solutions with the PEG concentrations used here. For the other crowding agents, where specific densities in solution were not reported, a direct correspondence between volume and weight fraction based on the densities of the pure substances was assumed.

PEG scaling law. According to Devanand *et al.* (29), water is a good solvent for PEG. Fig. S1 shows the scaling law obtained in (29) for long PEGs with the radii of gyration of PEGs used in the current study. The radii of gyration of PEGs are taken from (74, 75). In cases where only the hydrodynamic radius was determined experimentally, a conversion between hydrodynamic radius and radius of gyration was applied according to the ratio determined experimentally in (29). A deviation from the predicted scaling behavior (29) is visible for short PEGs, for which finite length effects start to dominate.

Estimation of the radius of gyration of IDPs at the Θ-state. An estimate of the radius of gyration for the four proteins in Θ-state conditions is obtained according to the previous analysis of the scaling exponent of disordered and unfolded sequences presented by Hofmann *et al.* (27), where the radius of gyration is linked to the scaling exponent by (76)

$$R_{g\Theta} = \sqrt{\frac{2l_p^* b}{(2\nu+1)(2\nu+2)}}N^\nu,$$
(Eq. S26)

where l_p^* = 0.4 nm, $b = 0.38$nm, $\nu = 0.5$, and N is the number of amino acids of the respective protein. The resulting radii of gyration in Θ-solvent are about 1.7 nm for IN, 1.8 nm for ProTα−C and ProTα−N, and 2 nm for ACTR.

Estimation of scaling exponents of IDPs. Following the empirical relation for the scaling exponent as a function of the hydrophobicity and net charge of unfolded and disordered proteins obtained previously (27), we estimated the scaling exponent for all full-length disordered sequences longer than 25 amino acids deposited in the Disprot database (v6.01) (77).The exponent ν is calculated according to

$$v(Q) = 1/3 + a\left[1 + \exp(x_0 - Q)/z\right]^{-1} \text{ and } v(H) = 1/3 + a\left[1 + \exp(x_0 + cH - d)/z\right]^{-1}, \quad \text{(Eq. S27)}$$

where H is the hydrophobicity according to the scale of Kyte and Doolittle (78), Q is the mean net charge of the sequence, $a = 0.394$, $z = 0.09$, $x_0 = 0.114$, $c = 1.72$ and $d = 0.9$.
The exponent is determined according to

$$v = \begin{cases} v(Q) & u^* > 0 \vee f = 0 \vee g = 0 \\ v(H) & u^* \le 0 \vee f = 0 \wedge g = 0 \end{cases}, \quad \text{(Eq. S28)}$$

where f and g are the fractions of positive and negative charges in the sequence, respectively, and u^* is calculated according polyampholyte theory (11) as

$$u^* = \frac{4\pi l_b (f - g)^2}{\kappa^2} - \frac{\pi l_b^2 (f + g)^2}{\kappa}, \quad \text{(Eq. S29)}$$

where l_b is the Bjerrum length, and κ^{-1} is the Debye length.

We note that a higher percentage of globule-like IDPs has been estimated from simulations for a different subset of the Disprot database (33, 34). However, taking into account polyampholyte effects and patterning, it has been predicted that the majority of IDPs will maintain coil-like properties (34), and consequently they should be susceptible to the effect of crowding described here.

Physiological concentrations of IDPs in the nuclear pore complex and in RNA granules. A nuclear pore complex contains approximately 200 disordered nucleoporins containing FG-repeats, each with a length of ~600 amino acids (79, 80). From the chain length of these sequences, the overlap concentrations can be estimated to be in the range between of volume fractions between 0.05% and 4%. Assuming the nuclear pore to have a diameter of 30 nm and a height of 40 to 80 nm (25, 80, 81), the volume fraction occupied by the disordered nucleoporins is easily estimated as $\phi_{NP} = \dfrac{v_{NP}}{V_{pore}} = \dfrac{n_{NP} \cdot N_{NP} \cdot \langle v_{aa} \rangle}{\pi r_{pore}^2 h_{pore}}$, where v_{NP} is the volume occupied by the nucleoporins; V_{pore} is the volume available in the pore; n_{NP} is the number of nucleoporins; N_{NP} is the sequence length of nucleoporins; $\langle v_{aa} \rangle$ is the average volume of the amino acid residues (approximately 0.13 nm^3); r_{pore} and h_{pore} are, respectively, the radius and the height of the pore. The volume fraction occupied by disordered nucleoporins is therefore between 25 and 55%, about an order of magnitude higher than the overlap concentration. The effects discussed in the main text are thus highly likely to be of importance for the conformational distributions of nucleoporins *in vivo*.

Various IDPs have been identified in RNA granules (22, 82, 83). Recent work has shown that the multivalency of these proteins can control phase separation and therefore the assembly of RNA granules (23). Even though the mechanism is not fully understood, and the proteins can undergo different conformational changes during the phase-separation process, the mechanism suggested by Li et al. (23) can be taken as an example to test whether the polymeric nature of the proteins is expected to be important in that range of concentrations. For the case of the engineered proline-rich motives (PRM) (23), the sequence is likely to be almost completely disordered. The molar concentration of the protein at which a given volume is occupied entirely by those IDPs is given by $c = 1/(N \langle v_{aa} \rangle \cdot N_A)$, where N is the number of amino acids of the sequence, and N_A is the Avogadro number. A calculation for the case of a sequence with 250 amino acids as those

considered in the work of Li *et al. (23)* result in a local protein concentration in the range of ~50 mM. The corresponding overlap concentration is estimated to be between volume fractions of 1% and 6% or 0.5 and 3 mM. Considering that two different proteins are mixed in similar ratios in these experiments, a concentration of 0.25 to 1.5 mM is sufficient to reach the overlap regime. Phase separation for proteins of this length occurs at concentrations of approximately 50 µM, only 5 times less than the overlap concentration. However, due to phase separation, in the droplets, a concentration of proteins 100 times higher than the bulk solution is reported (23). The confinement in the droplet is therefore plausibly causing an increase of protein concentration significantly higher than the overlap concentration, and the overlap between disordered coils will affect the conformations of the disordered sequences. FUS and hnRNPA2, two disordered proteins identified in RNA granules, have been shown to exhibit a sol-gel transition *in vitro* (82) at a concentration above 1 mM with an overlap concentration ranging between 0.15 and 1 mM. Therefore, these proteins are expected to be in the semidilute regime before gelation.
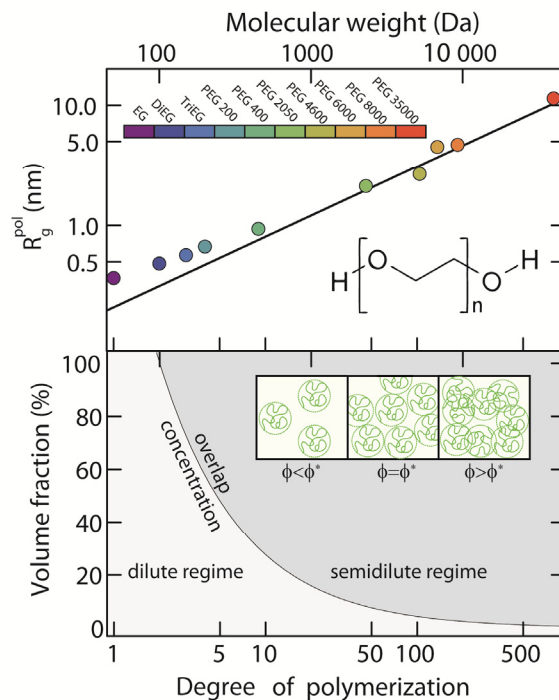
Figure S1. Polymer properties of polyethylene glycol. (a) Radius of gyration of PEG as a function of molecular weight/degree of polymerization (74, 75) with a fit (black line) to the scaling law $R_g^{pol} = 0.21\,\text{nm} \cdot P^{0.583}$ (26). The scaling exponent indicates that water is a good solvent for PEG. Deviations from the fit (which was obtained for PEG molecules over the entire range of lengths originally reported (29)) are due to finite length effects for small values of P. (b) Overlap concentration as obtained from Eq. S10, and schematic representation of a polymer solution in the dilute regime, at the overlap concentration, and in the semidilute regime.
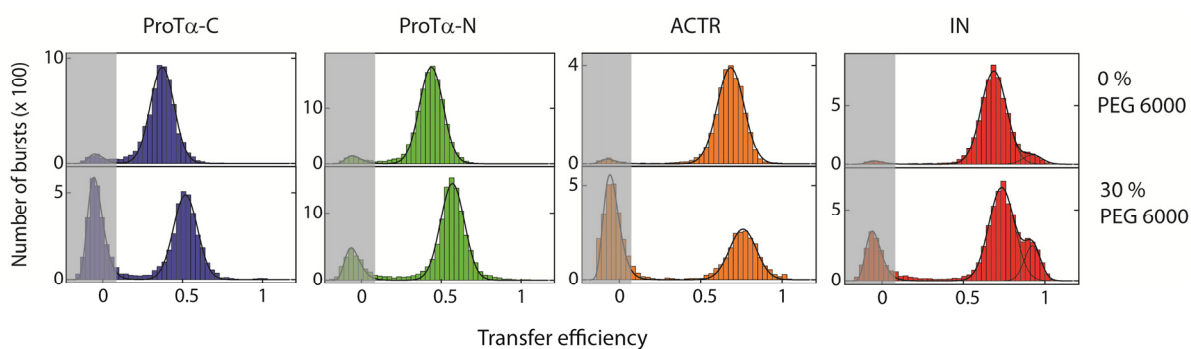
Figure S2. FRET efficiency histograms obtained with FRET dyes shifted toward longer wavelengths compared to Fig. 2. Histograms for ProTα-C, ProTα-N, ACTR, and IN variants labeled with ATTO546 and ATTO647N in the absence and presence of high PEG concentration. The shift towards higher excitation and emission wavelengths reduces the contribution of fluorescence background from impurities in the PEG and provides additional evidence that the peak broadening at high PEG concentration observed in Fig. 2 is mainly due to impurities in the solution. Only in the case of IN, a second peak corresponding to the folded state is detected, consistent with the observations and data analysis in Fig. 2. Gaussian and lognormal distributions were used to fit the peaks (solid lines). The donor-only peaks originating from molecules lacking an active acceptor dye are shaded in grey.
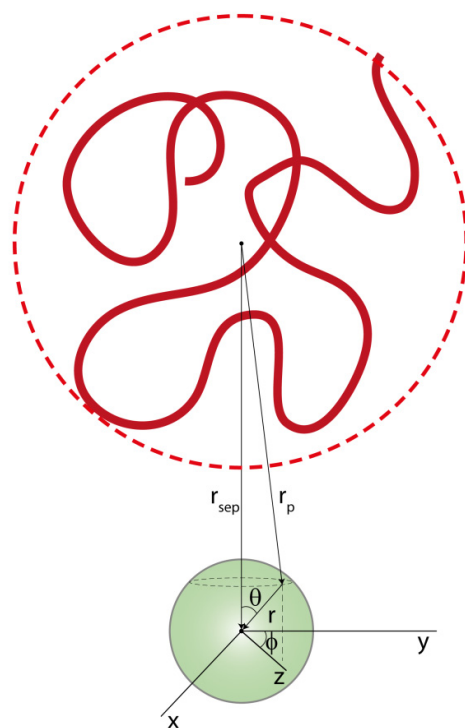
Figure S3. Gaussian cloud model. Representation of the Gaussian Cloud with a disordered conformation of the protein (in red), with the center of mass positioned at distance r_{sep} from a hard sphere (in green) with radius r. Adapted from (37).
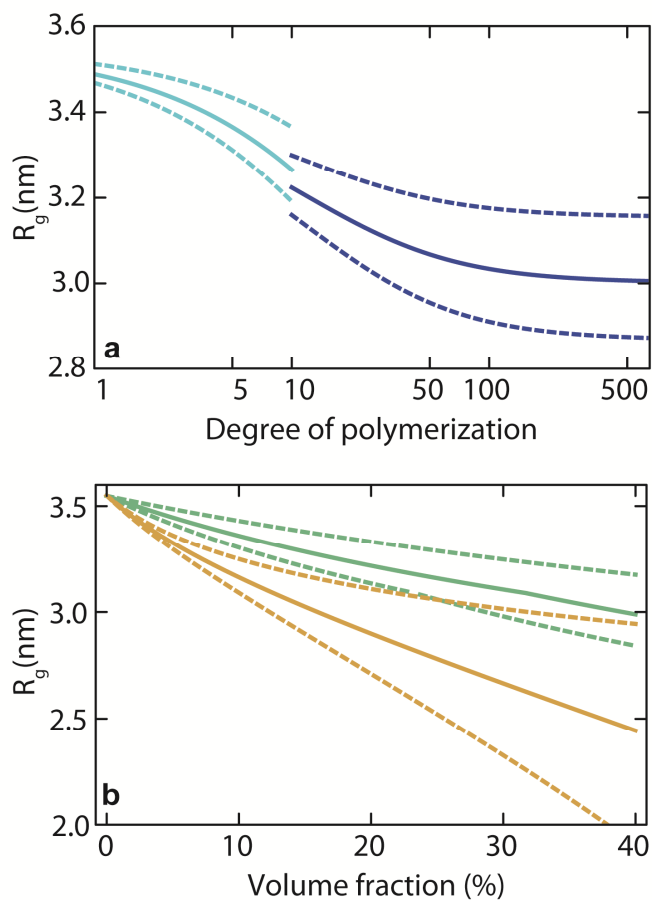
Figure S4. Effects of model parameters in the Flory-Huggins theories. Upper panel: Calculated radius of gyration of ProTα-C as a function of the degree of polymerization of PEG at a volume fraction of 15% according to Flory-Huggins theory (cyan) and renormalized Flory-Huggins theory (blue). Dashed curves show the change in the prediction of Flory-Huggins theory if a deviation of ±50% from the fitted value for the parameter *a* is assumed, and the response of renormalized Flory-Huggins theory to the fitting parameter s_{NP} between 0 (no interactions) and the upper limit of 0.1 nm (strong interactions). Lower Panel: estimation of the radius of gyration of ProTα-C at different volume fractions for PEG 400 (green solid curve) and PEG 6000 (yellow curve). Dashed lines report the response to the fitting parameter of the Flory-Huggins theory for PEG 400 and of the renormalized Flory-Huggins theory for PEG 6000 at the same conditions described above.
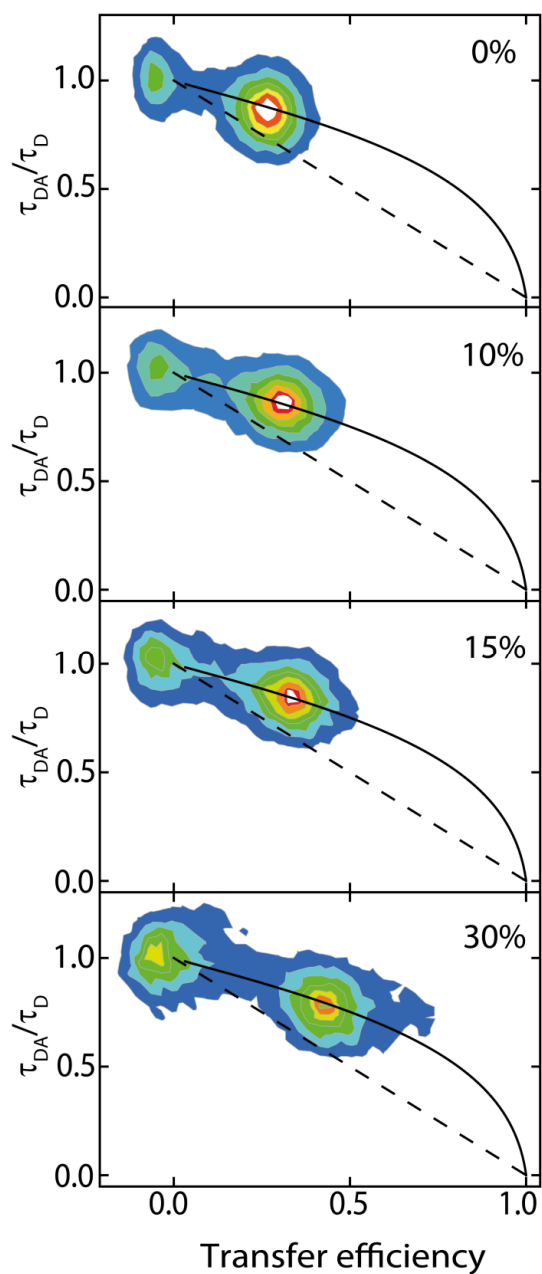
Figure S5. Multiparameter single-molecule fluorescence analysis. Two-dimensional histograms of relative donor τ_{DA}/τ_D lifetime versus FRET efficiency measured between 0 and 30% volume fraction of PEG 6000 compared to the expected trend for a fixed distance (black dashed line) and for a chain reconfiguring over the distribution of distances $P(r)$ given by Eq. S2 (black solid line), as described in detail in the section *Fluorescence lifetimes and anisotropies*.
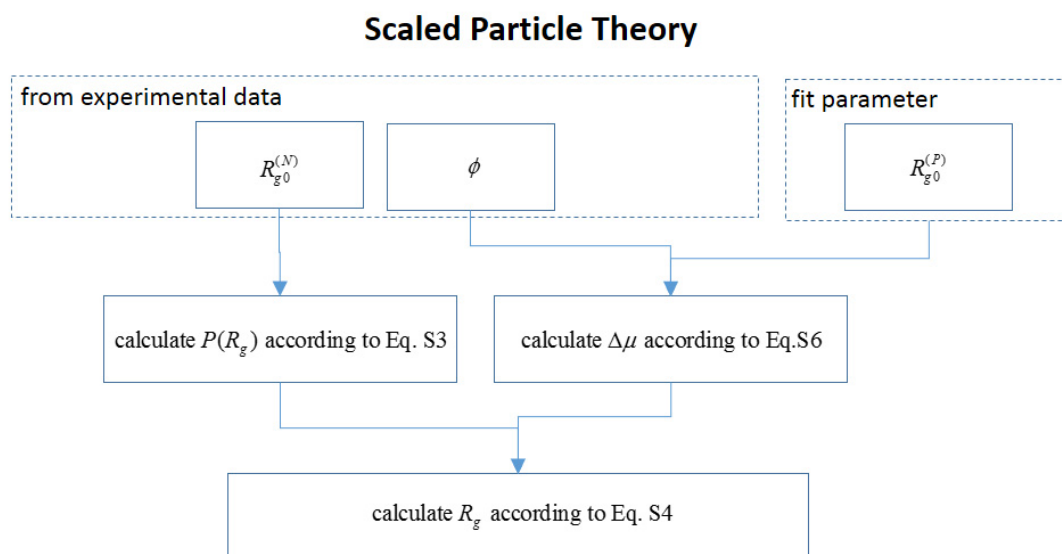
Figure S6. Flowchart for the fitting procedure with scaled-particle theory

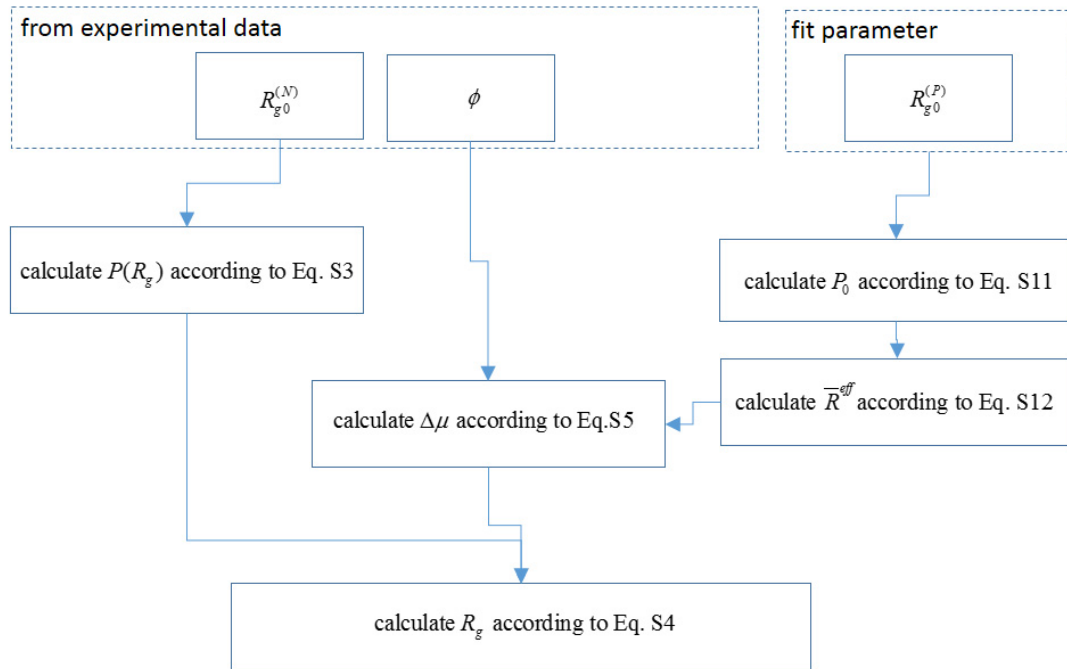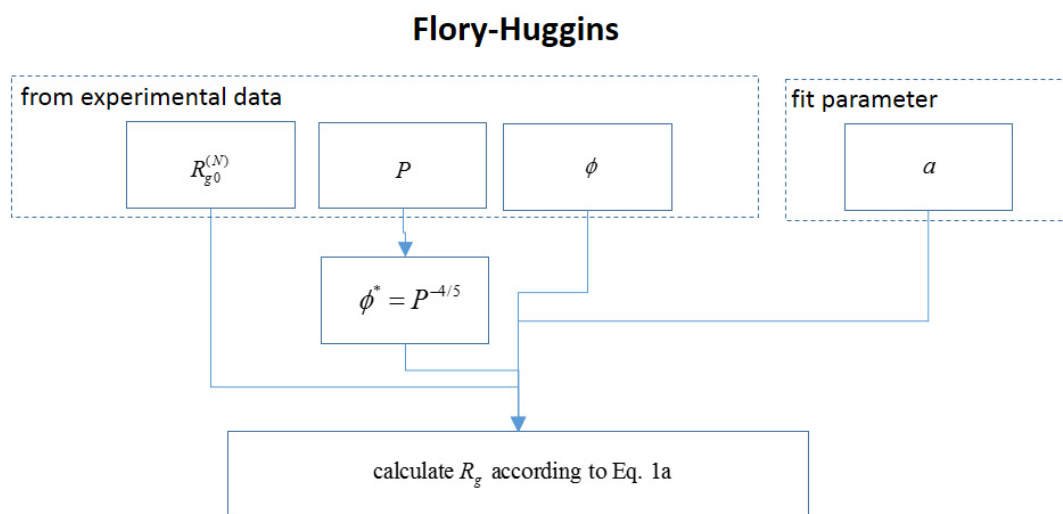**Figure S7.** Flowchart for the fitting procedure with the Gaussian cloud model

Flory-Huggins

Figure S8. Flowchart for the fitting procedure with Flory-Huggins theory.

Renormalized Flory-Huggins

from experimental data

l_0	$R_{g0}^{(N)}, N$	$R_{g0}^{(P)}, P$	ϕ

fit parameter

s_{NP}

$$B_N = \frac{R_{g0}^{(N)}}{0.8^{0.5} N^{0.588}}$$

$$B_P = \frac{R_{g0}^{(P)}}{0.8^{0.5} P^{0.588}}$$

calculate N_R according to Eq. S22

calculate $l_R = B_N \left(\dfrac{N}{N_R} \right)^{0.588}$

$$f_{NP} = 1 + \left(\frac{l_R}{s_{NP}} \right)^{-0.4}$$
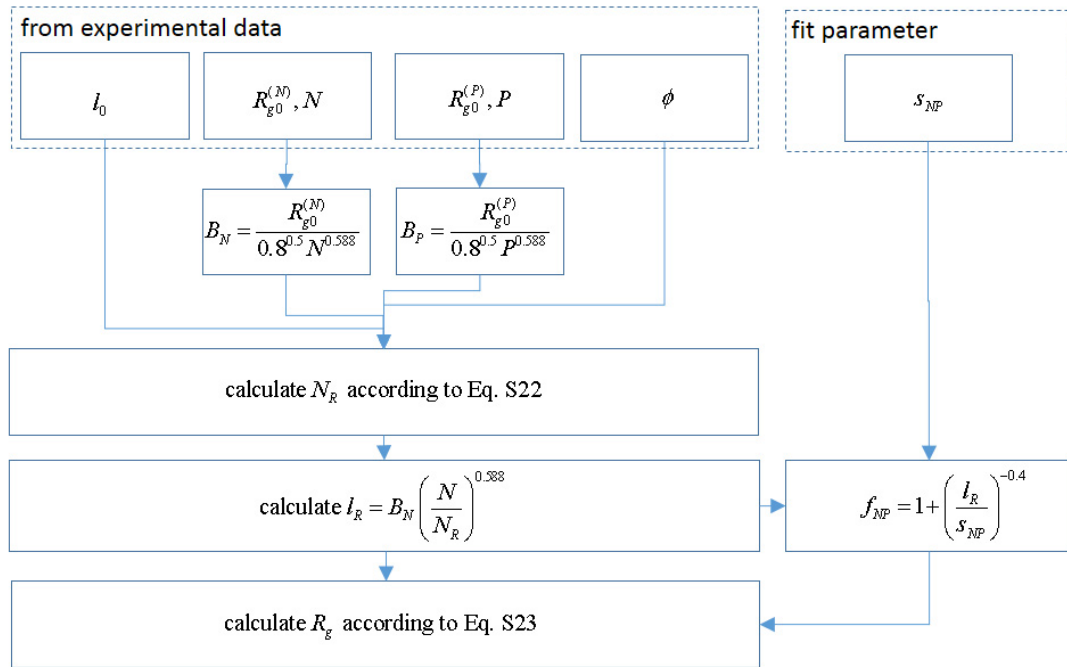
calculate R_g according to Eq. S23

Figure S9. Flowchart for the fitting procedure with the renormalized Flory-Huggins theory.

	Mol. Weight (Da)	Average degree of polymerization
Ethylene glycol[*]	62.07	1
Diethylene glycol[*]	106.12	2
Triethylene glycol[*]	150.17	3
Polyethylene glycol 200[*]	190 - 210	4
Polyethylene glycol 400[*]	380 - 420	9
Polyethylene glycol 2050[*]	1900 - 2200	46
Polyethylene glycol 4600[*]	4400 - 4800	104
Polyethylene glycol 6000[*]	5000 - 7000	136
Polyethylene glycol 8000[Error! Bookmark not defined.]	7300 - 9000	185
Polyethylene glycol 35'000[*]	35000	795
Polyvinyl alcohol 10'000[*]	9000 - 10000	216
Polyvinyl alcohol 40'000[*]	31000 - 50000	920
Polyvinyl alcohol 90'000[*]	89000 - 98000	2125
Polyvinylpyrrolidone K90[*]	360000	3243
Polyvinylpyrrolidone 1.3M[*]	1300000	11712
Dextran 6000[*]	6000	37
Dextran 40'000[*]	40000	247
Dextran 100'000[*]	100000	617

[*]Sigma-Aldrich (Switzerland), [Error! Bookmark not defined.]Carl Roth (Germany)

Table S1. Crowding agents used in this study.

ProTα-C (C56-C110)	1 10	20	30	40	50	56 60	70	80	90	100
	GP SDAAVDTSSE	ITTKDLKEKK	EVVEEAENGR	DAPANGNAEN	EENGEQEADN	EVDEE**C**EEGG	EEEEEEEEGD	GEEEDGDEDE	EAESATGKRA	AEDDEDDDVD
	110 TKKQKTDED**C**									

ProTα-N (C2-C56)	1 10	20	30	40	50	56 60	70	80	90	100
	GP **C**DAAVDTSSE	ITTKDLKEKK	EVVEEAENGR	DAPANGNAEN	EENGEQEADN	EVDEE**C**EEGG	EEEEEEEEGD	GEEEDGDEDE	EAESATGKRA	AEDDEDDDVD
	110 TKKQKTDEDD									

ACTR (C1-C73)	1 10	20	30	40	50	60	70 73
	GP **C**GTQNRPLLR	NSLDDLVGPP	SNLEGQSDER	ALLDQLHTLL	SNTDATGLEE	IDRALGIPEL	VNQGQALEPK QD**C**

IN (C8-C57)	1 10	20	30	40	50	57
	GSH MFLDGID**C**AQ	EEHEKAHSNF	RAMASDFNLP	PVVAKEIVAS	CDKCQLKGEA	MHGQVD**C**

Table S2. Sequences of the proteins used in this study.

	ProTα-C	ProTα-N	ACTR	IN
R_{g0} (nm)	3.46±0.02	3.03±0.02	2.47±0.02	1.95±0.02

	PEG 6000
R_g^{crd} (nm)	5.8±0.1

Table S3. Global fit of PEG 6000 data with scaled-particle theory. Parameters obtained from fitting the data in Fig. 2 with Eq. S4

	$P < N^{1/2}$	$P > N^{1/2}$
	a	s_{NP} (nm)
ProTα-C	1.18±0.05	0.032±0.004
ProTα-N	0.7±0.1	0.024±0.008
ACTR	0.8±0.1	0.08±0.02
IN	0.05±0.02	< 2·10⁻⁶ [*]

[*] s_{NP} is sufficiently close to zero that the ternary system reduces to a binary system where the protein and the crowding agents cannot be distinguished

Table S4. Global fit of Protα-C, Protα-N, ACTR and IN in the presence of PEG with Flory-Huggins theories. Parameters obtained from fitting the complete dataset in Fig. 3 with Eqs. 1a and 1b.

	s_{NP} (nm)
PEG	0.032±0.004
PVA	< $4 \cdot 10^{-5}$ [*]
PVP	< $4 \cdot 10^{-6}$ [*]
Dextran	< $3 \cdot 10^{-5}$ [*]

[*]s_{NP} is sufficiently close to zero that the ternary system reduces to a binary system where the protein and the crowding agents cannot be distinguished

Table S5. Fit results with Flory-Huggins theories of ProTα-C collapse in presence of PVA, PVP and Dextran. Parameters obtained from fitting the datasets of ProTα-C in the presence of PEG (Fig. 3), PVA, PVP and Dextran (Fig. 5) with Eq. 1b.

Supporting Information references

1.   Dyson HJ & Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6(3):197-208.
2.   Dunker AK, Silman I, Uversky VN, & Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18(6):756-764.
3.   Jensen MR, Ruigrok RW, & Blackledge M (2013) Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.*
4.   Wright PE & Dyson HJ (2009) Linking folding and binding. *Curr. Opin. Struct. Biol.*
5.   Ferreon AC, Moran CR, Gambin Y, & Deniz AA (2010) Single-molecule fluorescence studies of intrinsically disordered proteins. *Methods in enzymology* 472:179-204.
6.   Schuler B, Müller-Späth S, Soranno A, & Nettels D (2012) Application of confocal single-molecule FRET to intrinsically disordered proteins. *Methods Mol. Biol.* 896:21-45.
7.   Ferreon AC, Ferreon JC, Wright PE, & Deniz AA (2013) Modulation of allostery by protein intrinsic disorder. *Nature* 498(7454):390-394.
8.   Mao AH, Lyle N, & Pappu RV (2013) Describing sequence-ensemble relationships for intrinsically disordered proteins. *Biochem J* 449(2):307-318.
9.   Lyle N, Das RK, & Pappu RV (2013) A quantitative measure for protein conformational heterogeneity. *J. Chem. Phys.* 139(12).
10.   Fisher CK & Stultz CM (2011) Protein Structure along the Order-Disorder Continuum. *J. Am. Chem. Soc.* 133(26):10022-10025.
11.   Müller-Späth S*, et al.* (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* 107(33):14609-14614.
12.   Uversky VN (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *Protein J.* 28(7-8):305-325.
13.   Zhou HX, Rivas GN, & Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37:375-397.
14.   Gershenson A & Gierasch LM (2011) Protein folding in the cell: challenges and progress. *Curr. Opin. Struct. Biol.* 21(1):32-41.
15.   Dedmon MM, Patel CN, Young GB, & Pielak GJ (2002) FlgM gains structure in living cells. *Proceedings of the National Academy of Sciences of the United States of America* 99(20):12681-12684.
16.   McNulty BC, Young GB, & Pielak GJ (2006) Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder. *Journal of molecular biology* 355(5):893-897.
17.   Munishkina LA, Cooper EM, Uversky VN, & Fink AL (2004) The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. *J Mol Recognit* 17(5):456-464.
18.   Szasz C*, et al.* (2011) Protein Disorder Prevails under Crowded Conditions. *Biochemistry* 50(26):5834-5844.
19.   Hong JA & Gierasch LM (2010) Macromolecular Crowding Remodels the Energy Landscape of a Protein by Favoring a More Compact Unfolded State. *Journal of the American Chemical Society* 132(30):10445-10452.
20.   Mikaelsson T, Aden J, Johansson LBA, & Wittung-Stafshede P (2013) Direct Observation of Protein Unfolded State Compaction in the Presence of Macromolecular Crowding. *Biophysical Journal* 104(3):694-704.

21. Johansen D, Jeffries CM, Hammouda B, Trewhella J, & Goldenberg DP (2011) Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. *Biophysical journal* 100(4):1120-1128.
22. Han TNW*, et al.* (2012) Cell-free Formation of RNA Granules: Bound RNAs Identify Features and Components of Cellular Assemblies. *Cell* 149(4):768-779.
23. Li PL*, et al.* (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483(7389):336-U129.
24. Brangwynne CP (2011) Soft active aggregates: mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. *Soft Matter* 7(7):3052-3059.
25. Rout MP*, et al.* (2000) The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *J. Cell. Biol.* 148(4):635-651.
26. Elcock AH (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr Opin Struc Biol* 20(2):196-206.
27. 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.
28. Soranno A*, et al.* (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single molecule spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 109(44):17800-17806.
29. Devanand K & Selser JC (1991) Asymptotic-Behavior and Long-Range Interactions in Aqueous-Solutions of Poly(Ethylene Oxide). *Macromolecules* 24(22):5943-5947.
30. Zimmerman SB & Trach SO (1991) Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of Escherichia-Coli. *Journal of molecular biology* 222(3):599-620.
31. Harris JM (1992) *Poly(Ethylene Glycol) Chemistry: Biotechnical and Biomedical Applications* (Plenum Press, New York).
32. Uversky VN, Gillespie JR, & Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 41(3):415-427.
33. Mao AH, Crick SL, Vitalis A, Chicoine CL, & Pappu RV (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* 107(18):8183–8188.
34. Das RK & Pappu RV (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U S A* 110(33):13392-13397.
35. 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.
36. Cheung MS, Klimov D, & Thirumalai D (2005) Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proceedings of the National Academy of Sciences of the United States of America* 102(13):4753-4758.
37. Minton AP (2005) Models for excluded volume interaction between an unfolded protein and rigid macromolecular cosolutes: Macromolecular crowding and protein stability revisited. *Biophys. J.* 88(2):971-985.
38. Mittal J & Best RB (2010) Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys. J.* 98(2):315-320.
39. Joanny JF, Grant P, Pincus P, & Turkevich LA (1981) Conformations of Polydisperse Polymer-Solutions - Bimodal Distribution. *J Appl Phys* 52(10):5943-5948.
40. Edwards SF (1966) Theory of Polymer Solutions at Intermediate Concentration. *P Phys Soc Lond* 88(560P):265-&.
41. Schäfer L (1999) *Excluded Volume Effects in Polymer Solutions as Explained by the Renormalization Group* (Springer, Berlin).

42. Tran HT & Pappu RV (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. *Biophysical journal* 91(5):1868-1886.

43. Wilson KG (1983) The Renormalization-Group and Critical Phenomena. *Rev. Mod. Phys.* 55(3):583-600.

44. Schäfer L & Kappeler C (1993) Interaction Effects on the Size of a Polymer-Chain in Ternary Solutions - a Renormalization-Group Study. *Journal of Chemical Physics* 99(8):6135-6154.

45. Nose T (1986) Chain Dimension of a Guest Polymer in the Semidilute Solution of Compatible and Incompatible Polymers. *J Phys-Paris* 47(3):517-527.

46. Minton AP (2013) Quantitative assessment of the relative contributions of steric repulsion and chemical interactions to macromolecular crowding. *Biopolymers* 99(4):239-244.

47. Sarkar M., Li C., & Pielak GJ (2013) Soft interactions and crowding. *Biophys. Rev.* 5(2):187-194.

48. Kim YC & Mittal J (2013) Crowding Induced Entropy-Enthalpy Compensation in Protein Association Equilibria. *Phys. Rev. Lett.* 110(20).

49. Shoemaker BA, Portman JJ, & Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U. S. A.* 97(16):8868-8873.

50. Trizac E, Levy Y, & Wolynes PG (2010) Capillarity theory for the fly-casting mechanism. *Proceedings of the National Academy of Sciences of the United States of America* 107(7):2746-2750.

51. Schreiber G, Haran G, & Zhou HX (2009) Fundamental Aspects of Protein-Protein Association Kinetics. *Chem Rev* 109(3):839-860.

52. Vareli K, Frangou-Lazaridis M, van der Kraan I, Tsolas O, & van Driel R (2000) Nuclear distribution of prothymosin alpha and parathymosin: Evidence that prothymosin alpha is associated with RNA synthesis processing and para-thymosin with early DNA replication. *Exp. Cell Res.* 257(1):152-161.

53. Enkemann SA, Ward RD, & Berger SL (2000) Mobility within the nucleus and neighboring cytosol is a key feature of prothymosin-alpha. *J. Histochem. Cytochem.* 48(10):1341-1355.

54. Gelman H, Platkov M, & Gruebele M (2012) Rapid perturbation of free-energy landscapes: from in vitro to in vivo. *Chemistry* 18(21):6420-6427.

55. Phillip Y, Kiss V, & Schreiber G (2012) Protein-binding dynamics imaged in a living cell. *Proc. Natl. Acad. Sci. U. S. A.* 109(5):1461-1466.

56. Sakon JJ & Weninger KR (2010) Detecting the conformation of individual proteins in live cells. *Nature methods* 7(3):203-205.

57. Gast K, *et al.* (1995) Prothymosin alpha: a biologically active protein with random coil conformation. *Biochemistry* 34(40):13211-13218.

58. Zheng R, Jenkins TM, & Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proceedings of the National Academy of Sciences of the United States of America* 93(24):13659-13664.

59. Demarest SJ, *et al.* (2002) Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* 415(6871):549-553.

60. Yi S, Brickenden A, & Choy WY (2008) A new protocol for high-yield purification of recombinant human prothymosin alpha expressed in Escherichia coli for NMR studies. *Protein expression and purification* 57(1):1-8.

61. Müller BK, Zaychikov E, Bräuchle C, & Lamb DC (2005) Pulsed interleaved excitation. *Biophys. J.* 89(5):3508-3522.

62. Sisamakis E, Valeri A, Kalinin S, Rothwell PJ, & Seidel CAM (2010) Accurate Single-Molecule FRET Studies Using Multiparameter Fluorescence Detection. *Methods Enzymol.* 475:455-514.

63. Strickler SJ & Berg RA (1962) Relationship between Absorption Intensity and Fluorescence Lifetime of Molecules. *Journal of Chemical Physics* 37(4):814-&.

64. 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.

65. Sanchez IC (1979) Phase-Transition Behavior of the Isolated Polymer-Chain. *Macromolecules* 12(5):980-988.

66. Minton AP (1998) Molecular crowding: analysis of effects of high concentrations of inert cosolutes on biochemical equilibria and rates in terms of volume exclusion. *Methods in enzymology* 295:127-149.

67. Boublik T (1974) Statistical Thermodynamics of Convex Molecule Fluids. *Mol. Phys.* 27(5):1415-1427.

68. Oversteegen SM & Roth R (2005) General methods for free-volume theory. *J. Chem. Phys.* 122(21).

69. Qin S & Zhou HX (2013) Effects of Macromolecular Crowding on the Conformational Ensembles of Disordered Proteins. *The journal of physical chemistry letters* 4(20).

70. De Gennes PG (1979) *Scaling concepts in polymer physics* (Cornell University Press, Ithaca, N.Y.).

71. Raphael E, Fredrickson GH, & Pincus P (1992) One Long-Chain among Shorter Chains - the Flory Approach Revisited. *Journal De Physique Ii* 2(10):1811-1823.

72. Kappeler C, Schafer L, & Fukuda T (1991) Light-Scattering from Ternary Polymer-Solutions - an Analysis of the Apparent Radius of Gyration. *Macromolecules* 24(10):2715-2718.

73. Ferry Kienberger, *et al.* (2000) Static and Dynamical Properties of Single Poly(Ethylene Glycol) Molecules Investigated by Force Spectroscopy. *Single Molecules* 1(2):123-128.

74. Kuga S (1981) Pore-Size Distribution Analysis of Gel Substances by Size Exclusion Chromatography. *J. Chromatogr.* 206(3):449-461.

75. Hosek M & Tang JX (2004) Polymer-induced bundling of F actin and the depletion force. *Phys. Rev. E* 69(5).

76. Hammouda B (1993) Sans from Homogeneous Polymer Mixtures - a Unified Overview. *Adv. Polym. Sci.* 106:87-133.

77. Sickmeier M, *et al.* (2007) DisProt: the database of disordered proteins. *Nucleic Acids Res.* 35:D786-D793.

78. Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157(1):105-132.

79. Milles S & Lemke EA (2011) Single molecule study of the intrinsically disordered FG-repeat nucleoporin 153. *Biophys. J.* 101(7):1710-1719.

80. Rout MP, Aitchison JD, Magnasco MO, & Chait BT (2003) Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol.* 13(12):622-628.

81. Alber F, *et al.* (2007) The molecular architecture of the nuclear pore complex. *Nature* 450(7170):695-701.

82. Kato M, *et al.* (2012) Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell* 149(4):753-767.

83. Kato Y & Nakamura A (2012) Roles of cytoplasmic RNP granules in intracellular RNA localization and translational control in the Drosophila oocyte. *Dev. Growth Differ.* 54(1):19-31.