HHS Public Access

Author manuscript

J Phys Chem B. Author manuscript; available in PMC 2019 December 13.

Published in final edited form as:

J Phys Chem B. 2018 December 13; 122(49): 11262–11270. doi:10.1021/acs.jpcb.8b07066.

Atomistic Modeling of Intrinsically Disordered Proteins Under PEG Crowding: Quantitative Comparison with Experimental Data and Implication of Protein-Crowder Attraction

Valery Nguemaha[†], **Sanbo Qin**[†], and **Huan Xiang Zhou**[*,†,‡]

[†]Department of Physics and Institute of Molecular Biophysics, Florida State University, Tallahassee, Florida 32306, United States

[‡]Department of Chemistry and Department of Physics, University of Illinois at Chicago, Chicago, Illinois 60607, United States

Abstract

The malleability of intrinsically disordered proteins (IDPs) has generated great interest in understanding how their conformations respond to crowded cellular environments. Experiments can report gross properties such as fluorescence resonance energy transfer (FRET) efficiency, but cannot resolve the conformational ensembles of IDPs and their interactions with macromolecular crowders. Computation can in principle provide the latter information but in practice has been hampered by the enormous expense for realistic modeling of IDPs and crowders and for sufficient conformational sampling. Here, taking advantage of a powerful method called FMAP (FFT-based Modeling of Atomistic Protein-crowder interactions), we computed how the conformational ensembles of three IDPs are modified in concentrated polyethylene glycol (PEG) 6000 solutions. We represented the IDPs at the all-atom level and the PEG molecules at a coarse-grained level, and calculated the experimental observable, i.e., FRET efficiency. Whereas accounting for only steric repulsion of PEG led to overestimation of crowding effects, quantitative agreement with experimental data was obtained upon including mild IDP-PEG attraction. The present work demonstrates that realistic modeling of IDPs under crowded conditions for direct comparison with experiments is now achievable.

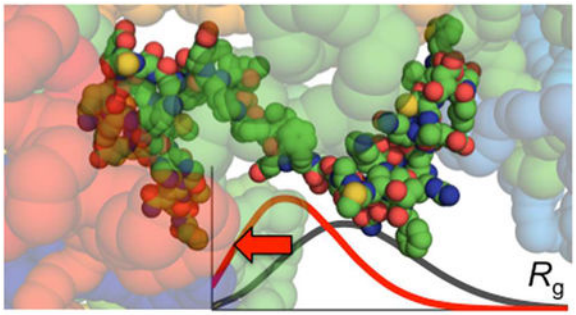
Graphical Abstract

[*]**Corresponding Author:** To whom correspondence should be addressed. Phone: **(**312**)** 355 3102. Fax: **(**312**)** 996-0431. hzhou43@uic.edu.

ASSOCIATED CONTENT

Supporting Information

Supporting Information is available for six additional figures and one additional table (PDF).



Introduction

Over one third of proteins are intrinsically disordered or contain large disordered regions, and the disorder is of critical importance for cellular functions.[1–4] The conformational preferences and the conformational responses upon interacting with other molecules are crucial determinants for the functions of intrinsically disordered proteins (IDPs). Notably, some IDPs undergo disorder-to-order transitions upon binding structured targets,[5–7] or Zn^{+2} ions and drug molecules.[8–11] Other IDPs, upon forming high-affinity complexes, while preserving structural disorder, experience conformational shifts such as compaction.[12] Much less attention has been paid to the fact that, in cells, IDPs encounter a high concentration of bystander macromolecules or crowders, at over 300 g/l or 30% of volume fraction.[13] The malleability of IDPs makes them especially susceptible to the influence by interactions with macromolecular crowders.[14] The aim of this study was to reveal conformational shifts of IDPs produced by realistically modeled protein-crowder interactions.

A number of experimental techniques have been used to characterize IDP conformations under macromolecular crowding. Bulk and single-molecule fluorescence resonance energy transfer (FRET) have indicated conformational compaction of an unfolded protein and IDPs by polymer crowders.[15–16] Using small-angle neutron scattering (SANS), one study found no apparent effect on the radius of gyration (R_g) of an IDP by protein crowders,[17] while another study presented evidence for compaction of the IDP FlgM at moderate concentrations of polymer and protein crowders but recovery in IDP size upon a further increase in crowder concentrations.[18] Moreover, based on fitting scattering profiles with an ensemble of IDP conformations, the latter study suggested that protein crowders can simultaneously induce compaction and expansion of FlgM, leading to two segregated conformational populations. Different NMR experiments have provided residue-specific information on IDPs under crowding. In an early study, backbone amide resonance broadening of FlgM C-terminal residues in the presence of a protein crowder as well as in *Escherichia coli* cells was interpreted as indicating gain of structure [19], although in retrospect the observation is also consistent with nonspecific binding with crowders. For the IDP α-synuclein, backbone chemical shifts have indicated no change in secondary structure under polymer crowding [20] and in *E. coli* cells,[21] whereas paramagnetic resonance enhancement have suggested compaction under polymer and protein crowding as well in mammalian cells.[20, 22] Interestingly, the extent of compaction was more modest for protein crowders relative to polymer crowders. Furthermore, NMR relaxation data suggested weak

interactions of N- and C-terminal residues with cytoplasmic components in mammalian cells, which were partially recapitulated by protein crowders but not polymer crowders.[22] Recently hydrogen-exchange mass spectrometry (HX-MS) showed faster exchange throughout the binding domain of the activator for thyroid hormone and retinoid receptors (ACTR) at 300 g/L Ficoll but slower exchange for one segment at 400 g/L Ficoll.[23] The former observation can be interpreted as indicating ACTR expansion, whereas the latter observation is probably due to Ficoll-induced ACTR aggregation, as found for other IDPs.[18, 20] Techniques such as FRET and SANS report on gross properties (e.g., FRET efficiency and scattering profile) and hence lack the power to resolve conformational ensembles. Although NMR spectroscopy and, to some extent, HX-MS generate residue-level data and may give hints on IDP-crowder interactions, unequivocal, detailed interpretation is extremely challenging.

Computation, in principle, can provide detailed answers on how protein-crowder interactions alter conformational ensembles of IDPs. However, due to the enormous expense for realistic modeling of IDPs and crowders and for sufficient conformational sampling, most computational studies have been limited to simplified models of IDPs and crowders, which nonetheless have generated valuable insight. Computer simulations using a coarse-grained representation (one bead per residue) have shown that IDPs always exhibit compaction in the presence of steric spherical crowders, which exert only steric repulsion, and the extent of compaction depends on the intrinsic flexibility of the IDPs.[24] Similar simulation results have been formulated into scaling relations for the extent of compaction.[25] The effects of steric spherical crowders have been generalized to IDPs modeled as chains of mixed types of beads, to mimic amino-acid compositions of proteins.[26] The extent of compaction was found to depend on bead composition, which affects intrinsic flexibility; hence the dependence on composition recapitulates the dependence on intrinsic flexibility noted in the earlier study.[24] Very recently, Banks at el.[18] reported simulation results for coarse-grained IDPs in the presence of attractive spherical crowders. Adding weak protein-crowder attraction led to moderation of the crowder-induced compaction; with stronger protein-crowder attraction, IDPs exhibit compaction at moderate crowder concentrations but subsequent expansion at higher crowder concentrations. The latter result is qualitatively consistent with the SANS data reported in the same study. In the past few years, all-atom explicit solvent molecular dynamics simulations have shown promise in studying some aspects of proteins under crowded cell-like conditions.[27] For IDPs in particular, though, inaccuracy of force fields and under sampling of conformations still present major obstacles even in dilute solutions,[28–31] let alone in crowded solutions. The problem is underscored by a recent study using all-atom explicit solvent simulations,[32] where a 47-residue C-terminal fragment of ACTR exhibited significant expansion in the presence of 200 g/l of PEG 500, in contrast to the compaction of ACTR by PEG revealed by single molecule FRET.[16]

It is highly desirable to make direct comparison between experimental observables and computational results based on realistic modeling of IDPs and crowders. As emphasized above, interpreting experimental data alone to gain detailed information on crowding-induced shifts in IDP conformations or the underlying IDP-crowder interactions is extremely challenging. For example, to interpret the FRET efficiency measured in a single-molecule FRET experiment in terms of the IDP size, specifically the mean R_g, one has to assume both

a certain distribution for R_g and a certain conditional probability for the FRET donor-acceptor distance at a given R_g value.[16] SANS can yield the mean R_g of an IDP but it is silent on IDP-crowder interactions. In particular, when the mean R_g does not change under crowding, it does not necessarily mean absence of crowding effects, because the observation is also compatible with near cancelation of IDP compaction due to crowder steric repulsion and IDP expansion due to weak IDP-crowder attraction.[17–18] In the latter study, scattering profiles of FlgM under protein crowding were interpreted as indicating segregation of compacted and expanded conformations. Simulations of coarse-grained IDPs in the presence of attractive spherical crowders then suggested that compacted conformations fit into a void and wrap around a single bordering crowder, whereas expanded conformations snake through interstitial crevices and bind multiple crowders simultaneously. However, the crudeness of the IDP and crowder models precluded a quantitative comparison with the experimental data, therefore blocking an opportunity to test the conformational segregation hypothesis and assess the IDP-crowder interaction strengths of the experimental systems.

In the past few years we have developed a computational approach that opens the door to realistic modeling of IDPs under crowded conditions for direct comparison with experiments.[33–38] The basic idea is to determine how crowding biases the statistical weight of each conformation of the protein; in particular, these biases alter the conformational ensemble of an IDP. To implement this approach, we sample the conformations of the protein in the absence of crowders; each crowder-free conformation is then “postprocessed”, to determine the crowding bias, which is the Boltzmann factor of the chemical potential, $\Delta\mu$, of the protein inside the crowders. Our latest method implementing this approach is called FMAP, or FFT-based Modeling of Atomistic Protein-crowder interactions, in which protein-crowder interactions are expressed as correlation functions and evaluated via fast Fourier transform in order to calculate $\Delta\mu$.[37–38] FMAP has been applied to determine the conformational shift of coarse-grained IDPs by repulsive spherical crowders, resulting in IDP compaction identical to that obtained by direct simulations of the IDPs inside the crowders.[24] Most recently we have used FMAP to determine the chemical potential of an atomistic protein at a wide range of concentrations, leading to a phase diagram for liquid-liquid phase separation.[39]

Here we applied FMAP to make quantitative comparison with the FRET efficiency data of Soranno et al.[16] for four fluorescence donor-acceptor pairs in three IDPs under PEG 6000 crowding. We represented the IDPs at the all-atom level and PEG 6000 at a coarse-grained level. Accounting for steric repulsion of PEG produced stronger compaction than experimentally observed, but quantitative agreement was obtained upon including mild IDP-PEG attraction.

Methods

The three IDPs studied here are the 56-residue N-terminal domain of HIV-1 integrase (IN), the 71-residue ACTR, and the 110-residue prothymosin α (ProTα) (Fig. 1). In the experimental study,[16] Cys residues were appended at the N- or C-termini or substituted into internal or terminal positions in order to attach fluorescence donors and acceptors (Alexa Fluor 488 and Alexa Fluor 594, respectively). Specifically, four fluorescence donor-acceptor

pairs were positioned at residue 8 and the C-terminal extra residue in IN, at the N- and C-terminal extra residues in ACTR, and at either residues 1 and 56 or residues 56 and 110 in ProTα (referred to as ProTαN or ProTαC). The amino-acid sequences with these Cys additions and substitutions can be found in Table S2 of ref [16]. We modeled these Cys residues but not the fluorescence donors and acceptors. Below we detail the procedures for generating IDP conformations and calculating FRET efficiencies in the absence and presence of crowders.

Generation of IDP Conformations.

For each IDP, with the amino acid sequence as input, we used two methods, flexible-meccano (http://www.ibs.fr/research/scientific-output/software/flexible-meccano/)[40] and TRaDES, (http://trades.blueprint.org/)[41] to generate 100,000 all-atom conformations to represent the conformational ensemble (Fig. 2A). Similar to our earlier work,[42] both methods randomly choose conformers for each residue in Ramachandran space (i.e., with backbone ϕ, ψ torsion angles as coordinates), and eliminate steric clashes. TRaDES is slightly more elaborate, including a bias in the (ϕ, ψ) probability density of each residue by secondary structure prediction and with clash detection checked at the atom level rather than residue level.

Calculation of FRET Efficiencies.

At a given donor-acceptor distance r, the FRET efficiency is $E(r) = 1/\left(1 + r^6/R_0^6\right)$, where R_0 is the Foster radius of the donor-acceptor pair (approximately 54 Å for the experimental systems). An IDP samples a range of r values, and the observed FRET efficiency is an average over the r values:

$$\bar{E} = \int dr E(r) P(r) \quad (1)$$

where $P(r)$ is the probability density of r.

We took r to be the $C_\alpha - C_\alpha$ distance between the two Cys residues to which the fluorescence donor and accepters were attached. For each IDP, the average over r was done by assigning conformation n out of the $N = 100{,}000$ conformations a statistical weight w_n. Then

$$\bar{E} = \frac{\sum_{n=1}^{N} E(r_n) w_n}{\sum_{n=1}^{N} w_n} \quad (2)$$

If flexible-meccano or TRaDES conformations were used directly, we would have equal weights for all the conformations. The resulting $\bar{E}$ values disagreed with the experimental values for the IDPs in the absence of crowders. To remove this discrepancy, we reweighted the conformational ensemble by an r−dependent statistical weight,

$$w_{0n} = \exp(br_n) \quad (3)$$

where "0" in the subscript signifies that the weight is for the crowder-free case, and the parameter b was tuned for each IDP to reach agreement between predicted and experimental crowder-free $\bar{E}$ values. The resulting b values are listed in Table 1 and Table S1 flexible-meccano and TRaDES conformations, respectively. The raw and reweighted r distributions are compared in Fig. S1.

We also explored other forms of reweighting, such as one that depended on the radius of gyration of the protein. The latter reweighting led to a substantial population of conformations with R_g values much too high for the sequence length of a protein.

The crowding-induced shifts in protein conformations can be accounted for by a biasing weight,[24, 34–35]

$$w_{cn} = \exp(-\beta \Delta \mu_n) \quad (4)$$

where $\Delta\mu_n$ is the chemical potential for inserting a protein molecule into the crowder solution, and β is the inverse of the product between the Boltzmann constant and absolute temperature. All our calculations were for room temperature. Note that w_{cn} is to be multiplied with w_{0n} so that the total statistical weight in the presence of crowders is $w_{0n}w_{cn}$.

The calculation of $\Delta\mu_n$ involved: (1) preparing PEG 6000 solutions by coarse-grained (CG) simulations; (2) selecting an appropriate IDP-PEG interaction energy function; and (3) implementation of FMAP for the IDP-PEG systems. Next we describe these steps.

Coarse-Grained Simulations of PEG 6000 Solutions.

The trajectories of PEG 6000 solutions were kindly provided to us by Dr. Tamio Yamazaki, simulated as in his published work.[43] In brief, each oxyethylene unit of PEG and every three water molecules were represented by a single particle. The energy function for each PEG chain has a bonded term and an angle term; and all particles interact via a Lennard-Jones potential. The parameterization was validated against experimental data on diffusion constant and shear viscosity.

The CG simulations were run for 1 microsecond each, in triplicates for PEG 6000 at 10%, 20%, and 30% weight fractions. The simulation boxes were cubic with side length at approximately 150 Å. The final snapshots of the triplicates were used as representatives of each PEG 6000 solution (Fig. 2B).

IDP-PEG Interaction Energy Function.

We modeled the IDP-PEG interaction energy as pairwise additive. The interaction between IDP atom i and PEG particle j consisted of two parts. Steric repulsion operated when the pair distance r_{ij} was below a nominal contact distance d_{ij}:

$$u_{st}\left(r_{ij}\right) = \infty \text{ if } r_{ij} < d_{ij} \quad (5)$$

Weak attraction, modeled by a Lennard-Jones potential, was then present beyond the hard core:

$$u_{\text{at}}\left(r_{ij}\right) = \varepsilon_{ij}\left[\left(\sigma_{ij}/r_{ij}\right)^{12} - \left(\sigma_{ij}/r_{ij}\right)^{6}\right] \text{ if } r_{ij} > d_{ij} \quad (6)$$

The total IDP-PEG interaction is then

$$U_{\text{int}} = \Sigma_{ij}\, u_{\text{st}}\left(r_{ij}\right) + \alpha\, \Sigma_{ij}\, u_{\text{at}}\left(r_{ij}\right) \equiv U_{\text{st}} + U_{\text{at}} \quad (7)$$

For the FMAP implementation of the steric term, we took d_{ij} as the algebraic average of σ_{ii} and σ_{jj}:[37–38]

$$d_{ij} = \left(\sigma_{ii} + \sigma_{jj}\right)/2 \quad (8)$$

On the other hand, for the FMAP implementation of the attractive term, a combination rule for both ε_{ij} and σ_{ij} based on the geometric mean had to be used:

$$\varepsilon_{ij} = \left(\varepsilon_{ii}\varepsilon_{jj}\right)^{1/2} \quad (9)$$

$$\sigma_{ij} = \left(\sigma_{ii}\sigma_{jj}\right)^{1/2} \quad (10)$$

Initial values of Lennard-Jones parameters (ε and σ) were taken from the Amber force field for protein atoms[44] and from ref [43] for PEG CG particles. All ε_{ij} values were then scaled by a uniform factor α to adjust the magnitude of the IDP-PEG attraction.

FMAP Implementation.

The starting point of FMAP is the Widom expression[45] for the chemical potential for inserting a protein molecule into the crowder solution:

$$\exp\left(-\beta\, \Delta\, \mu_n\right) = \left\langle \exp\left[-\beta \mathrm{U}_{\text{int};n}(\mathbf{R})\right]\right\rangle_{\mathbf{R},\,\mathrm{c}} \quad (11)$$

where $\mathrm{U}_{\text{int};n}$ is the energy of the IDP in the nth conformation interacting with the crowders, and $\mathbf{R}$ is the position of the IDP inside the crowder solution, and $\langle\cdots\rangle_{\mathbf{R},\mathrm{c}}$ signifies averaging

over IDP position and crowder configuration (Fig. 2B,C). The averaging over **R** was achieved by using fast Fourier transform (FFT). In essence, we expressed $\exp[-\beta U_{st}(\mathbf{R})]$ and $U_{st}(\mathbf{R})$ as correlation functions and evaluated them via FFT. Details are reported in our previous work. [37–38]

The discretization required for FFT calculations introduces numerical errors, which can be eliminated by using smaller and smaller grid spacings but result in significant increases in computational times. Note that the averaging over **R** can be split,

$$\left\langle \exp\left[-\beta U_{\mathrm{int};n}(\mathbf{R})\right]\right\rangle_{\mathbf{R}} = \left\langle \exp\left[-\beta U_{\mathrm{st};n}(\mathbf{R})\right]\right\rangle_{0}\left\langle \exp\left[-\beta U_{\mathrm{at};n}(\mathbf{R})\right]\right\rangle_{1} \quad (12)$$

where $\langle\cdots\rangle_0$ and $\langle\cdots\rangle_1$ mean averaging over all grid points and over only grid points that are free of steric clashes between the IDP and crowders. If only a single crowder configuration is used for $\Delta\mu_n$, then eq 12 corresponds to

$$\Delta\mu_n = \Delta\mu_{\mathrm{st},n} + \Delta\mu_{\mathrm{at},n} \quad (13)$$

That is, $\Delta\mu_n$ would be decomposable into steric and attractive contributions. As we found previously,[38] the numerical errors due to discretization are systematic and can be compensated by artificially inflating the nominal contact distances d_{ij}. Here we found that, at a grid spacing of 0.6 Å, the errors in both $\Delta\mu_{\mathrm{st},n}$ and $\Delta\mu_{\mathrm{at},n}$ are nearly compensated by a 6% inflation in d_{ij} (Figs. S2 and S3).

With averaging over crowder configuration, $\Delta\mu_n$ is no longer decomposable. Here we used the final snapshots of the triplicate simulations for each PEG 6000 solution for averaging over crowder configuration. Specifically, we first determined the chemical potential, denoted as $\Delta\mu_{n,\mathrm{I}}$, $\Delta\mu_{n,\mathrm{II}}$, or $\Delta\mu_{n,\mathrm{III}}$, in each of the three crowder configurations, and then obtained the average chemical potential according to

$$\exp(-\beta\Delta\mu_n) = \left[\exp\left(-\beta\Delta\mu_{n,\mathrm{I}}\right) + \exp\left(-\beta\Delta\mu_{n,\mathrm{II}}\right) + \exp\left(-\beta\Delta\mu_{n,\mathrm{III}}\right)\right]/3 \quad (14)$$

For comparison, we also determined the chemical potential when only the steric contribution was accounted for (equivalent to setting the scaling factor α on ε_{ij} to 0).

Results

Assessment of Crowder-Free Conformational Ensembles.

For IN, the mean FRET efficiency calculated from the conformational ensemble generated by flexible-meccano is 0.634, which is already very close to the experimental value of 0.649. The mean R_g calculated from this ensemble was 21.1 Å (Table 1), which agrees closely with a value of 21.0 Å predicted by a scaling relation, $\bar{R}_g = 2.54N_{aa}^{0.522}$Å, where N_{aa} denotes the number of amino acids, that was compiled from a list of IDPs.[46] A very mild $\exp(br_n)$

reweighting with $b = -0.0029$ Å^{-1} (Fig. S1A) brought the mean FRET efficiency to agreement with the experimental value, and decreased the mean R_g slightly to 20.9 Å. Recently IN was studied by molecular dynamics simulations[29] and by fluorescence correlation spectroscopy and photoinduced electron transfer.[47] While the molecular dynamics simulations suggested transient helix formation, the latter experimental study concluded that local structure formation has relatively little effect on single-molecule FRET measurements because they probe only long chain segments within IDPs.

For ACTR, the flexible-meccano conformational ensemble likewise needed only a mild $\exp(br_n)$ reweighting, with $b = -0.0021$ Å^{-1} (Fig. S1A), in order to reproduce the experimental $\bar{E}$ value of 0.498. The mean R_g values before and after the reweighting are 23.6 and 23.4 Å, respectively, in close agreement with a value of 23.8 Å from the scaling relation and suggesting that ACTR, like IN, in buffer has a size typical of IDPs. Using small-angle X-ray scattering (SAXS), Kjaergaard et al.[48] found a decrease in mean R_g from 26.3 Å at 5 °C to 23.9 Å at 45°C for ACTR. In the same study, NMR spectroscopy detected only low levels of transient helical content. By simultaneous fitting single molecule FRET and SAXS data using ensemble reweighting, Borgia et al.[49] found a mean R_g of 24.8 Å at room temperature. Our mean R_g is consistent with these latter values.

The decrease in mean FRET efficiency from 0.649 for IN to 0.498 for ACTR is commensurate with the increase in donor-acceptor sequence separation from 49 to 72 peptide bonds (Fig. 1). In contrast, the experimental $\bar{E}$ values for ProTαN and ProTαC are much lower, at 0.35 and 0.28, respectively, even though the donor-acceptor sequence separations, at 55 and 54 peptide bonds, are close to that of IN. As a result, significant increases in mean donor-acceptor distance are required in reweighting the flexible-meccano conformational ensembles (Fig. S1A). The resulting b values of the $\exp(br_n)$ weight for ProTαN and ProTαC are 0.045 and 0.0708, respectively, with mean r increasing from 49.3 to 64.6 Å and from 46.9 to 68.9 Å. The corresponding increases in mean R_g are more modest, from 28.7 Å to 31.4 Å and from 28.6 to 32.6 Å, respectively, for the two variants of ProTα. For reference, the afore-mentioned scaling relation predicts a $\bar{R}_g$ of 29.5 Å for a 110-residue IDP like ProTα. Using SAXS, Uversky et al.[9] reported a mean R_g of 37.8 Å for this protein. The reweighted flexible-meccano conformational ensemble slightly underestimates the latter value.

While R_g provides a good measure for the size of an IDP, the asphericity provides a very useful measure on the shape of the IDP.[31, 49–50] The latter is defined as

$$\delta = 1 - 3\frac{\lambda_1^2\lambda_2^2 + \lambda_1^2\lambda_3^2 + \lambda_2^2\lambda_3^2}{\left(\lambda_1^2 + \lambda_2^2 + \lambda_3^2\right)^2} \quad (15)$$

where $\lambda_{1,2,3}$ are the eigenvalues of the moment of inertia tensor. The value of δ ranges from 0 to 1, with 0 representing a sphere and 1 representing a thin rod. The mean asphericities calculated from the reweighted flexible-meccano conformational ensembles for IN, ACTR, ProTαN, and ProTαC are 0.44, 0.41, 0.45, and 0.47, respectively. These values are typical

of IDPs.[50] The mean asphericity determined by Borgia et al.[49] using ensemble reweighting was between 0.4 and 0.45 for ACTR.

As expected of IDPs, both R_g and q span a wide range in each protein. Fig. 3 displays the 2-dimensinal probability densities in R_g and δ. All the results by the other IDP conformation-generation method, TRaDES, are very similar to those by flexible-meccano, and hence are presented only in Supporting Information. The r distributions before and after reweighting TRaDES conformations are presented in Fig. S1B, and the corresponding 2-dimensinal probability densities in R_g and δ are presented in Fig. S4.

IDPs Are Over-Compacted If PEG Exerts Only Steric Repulsion.

We now consider the effects of PEG crowding on the conformational ensembles of the four IDPs. Experimentally, the mean FRET efficiencies of the four proteins increase with increasing PEG weight fraction (Fig. 4, solid lines), indicating that the proteins become more and more compact under PEG crowding. PEG has often been thought of as a steric crowder, which exerts only steric repulsion. Qualitatively, steric repulsion by crowders is expected to compact IDPs.

However, when only steric repulsion is accounted for (by FMAP), the mean FRET efficiencies are overestimated for all the four proteins in the presence of 10% PEG 6000 and for IN and ACTR in the presence of 20% PEG 6000, and the overestimation is more severe at the higher PEG weigh fraction (Figs. 4 and S5, triangles connected by dash). In other words, with only steric repulsion by PEG, the IDPs conformations are shifted too far into the compact end. For example, the mean R_g of IN decreases from 20.9 Å in the absence of crowders to 20.0 Å in the presence of 10% PEG 6000 and further down to 17.8 Å at a 20% weight fraction (Figs. 5 and S6, green curves).

We were unable to obtain reliable results for ProTαN or ProTαC at 20% PEG 6000, because, in this denser crowder solution, the occurrence of large voids is extremely rare, and hence the probability for successful insertion of these longer IDP chains becomes exceedingly low.

Mild IDP-PEG Attraction Permits Quantitative Agreement With Experiment.

The effect of steric repulsion can be counteracted by mild IDP-crowder attraction. Indeed, with the scaling parameter α (eq 7) at 0.06, the PEG-induced compaction is sufficiently moderated so that the predicted $\bar{E}$ (Figs. 4 and S5, circles connected by dash) comes into reasonable agreement with the experimental values. With both steric repulsion and mild attraction, the compaction by PEG is more modest. For example, the mean R_g of IN decreases from 20.9 Å in the absence of crowders to 20.5 Å in the presence of 10% PEG 6000 and to 19.3 Å at a 20% weight fraction (Figs. 5 and S6, red curves).

Discussion

By modeling IDPs at the all-atom level, we have made quantitative comparison with experimental data on the their conformational shifts under PEG 6000 crowding. We directly computed the experimental observable, namely the mean FRET efficiency, from the IDP

conformations, after accounting for the conformational bias exerted by the crowder. This bias was treated by a unique method called FMAP, which computes IDP-crowder interactions via FFT. Whereas including only steric repulsion between the IDPs and PEG led to overly compact conformations, quantitative agreement with experimental data was obtained upon including mild IDP-PEG attraction.

Interestingly, mild attraction between IDPs and polymer crowders was implicated in a recent experimental study.[18] The SANS data in this study showed reduction in R_g for an IDP at moderate concentrations of dextran and Ficoll, but then an uptick in R_g at higher concentrations of these polymer crowders. The latter observation is a strong indication for the presence of IDP-crowder attraction. This attraction is much milder than that exerted by protein crowders, as the latter were apparently able to induce conformational expansion to such an extent as to produce segregation between the expanded and compacted populations.

In addition to PEG 6000, Soranno et al.[16] measured the effects of crowding over a wide range of PEG chain length. The degree of IDP compaction was similar for PEG chains longer than PEG 6000, but much less for shorter PEG chains (all at a fixed weight fraction). The latter observation contradicts a classical theory for steric crowders, namely the scaled particle theory, which predicts stronger IDP compaction for smaller crowders. IDP-PEG attraction counteracts the compaction due to steric repulsion, and may provide an explanation for the lesser degree of IDP compaction under crowding by shorter PEG chains. This issue will be addressed in the future.

An underlying assumption of FMAP is that the important conformations of an IDP under crowding are all present in the crowder-free ensemble used for calculation. This assumption is likely valid when IDP-crowder interactions are purely steric or at most mildly attractive. Stronger IDP-crowder attraction may induce conformations, such as an IDP chain wrapping around a crowder, that are unlikely sampled in the absence of crowders. Under those circumstances, FMAP may fail. FMAP may also not work well when inserting large IDPs into dense crowder solutions, as illustrated by inserting ProTαN or ProTαC into 20% PEG 6000.

In cellular environments, crowding-induced conformational shifts may be important for IDP functions. Such conformational shifts may also affect the efficacy of drugs targeting IDPs. There is much still to be learned about the influences of specific and nonspecific interactions on IDPs, nevertheless the present work demonstrates that realistic modeling of IDPs under crowded conditions for direct comparison with experiments is now achievable.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

It gives us great pleasure to contribute to this special issue for Bill Eaton, who has been a mentor and inspiration. This work was supported in part by National Institutes of Health Grant GM118091 and US-Israel Binational Science Foundation Grant 2015376. Dr. Tamio Yamazaki performed coarse-grained molecular dynamics

simulations of PEG 6000 (as described in ref [43]) and provided us with PEG 6000 coordinates. The authors are grateful for his kind assistance.

ABBREVIATIONS

ACTR	binding domain of the activator for thyroid hormone and retinoid receptors
FFT	fast Fourier transform
FMAP	FFT-based modeling of atomistic protein-crowder interactions
FRET	fluorescence resonance energy transfer
HX-MS	hydrogen-exchange mass spectrometry
IDP	intrinsically disordered proteins
IN	N-terminal domain of HIV-1 integrase
PEG	polyethylene glycol
ProTα	prothymosin α
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering

REFERENCES

1. Dunker AK; Cortese MS; Romero P; Iakoucheva LM; Uversky VN, Flexible Nets: The Roles of Intrinsic Disorder in Protein Interaction Networks. FEBS J. 2005, 272, 5129–5148. [PubMed: 16218947]
2. Zhou HX, Intrinsic Disorder: Signaling Via Highly Specific but Short-Lived Association. Trends Biochem. Sci 2012, 37, 43–48. [PubMed: 22154231]
3. Tantos A; Han K-H; Tompa P, Intrinsic Disorder in Cell Signaling and Gene Transcription. Mol. Cell. Endocrinol 2012, 348, 457–65. [PubMed: 21782886]
4. Wright PE; Dyson HJ, Intrinsically Disordered Proteins in Cellular Signalling and Regulation. Nat. Rev. Mol. Cell Biol. 2014, 16, 18–29.
5. Abdul-Manan N; Aghazadeh B; Liu GA; Majumdar A; Ouerfelli O; Siminovitch KA; Rosen MK, Structure of Cdc42 in Complex with the Gtpase-Binding Domain of the 'Wiskott-Aldrich Syndrome' Protein. Nature 1999, 399, 379–383. [PubMed: 10360578]
6. Demarest SJ; Martinez-Yamout M; Chung J; Chen H; Xu W; Dyson HJ; Evans RM; Wright PE, Mutual Synergistic Folding in Recruitment of Cbp/P300 by P160 Nuclear Receptor Coactivators. Nature 2002, 415, 549–553. [PubMed: 11823864]
7. Sorenson MK; Ray SS; Darst SA, Crystal Structure of the Flagellar Sigma/Anti-Sigma Complex Sigma(28)/Flgm Reveals an Intact Sigma Factor in an Inactive Conformation. Mol. Cell 2004, 14, 127–138. [PubMed: 15068809]
8. Cai M; Zheng R; Caffrey M; Craigie R; Clore GM; Gronenborn AM, Solution Structure of the N-Terminal Zinc Binding Domain of Hiv-1 Integrase. Nat. Struct. Biol 1997, 4, 567–577. [PubMed: 9228950]
9. Uversky VN, et al., Zn^{2+}-Mediated Structure Formation and Compaction of the "Natively Unfolded" Human Prothymosin A. Biochem. Biophys. Res. Commun 2000, 267, 663–668. [PubMed: 10631119]

10. Peterson JR; Bickford LC; Morgan D; Kim AS; Ouerfelli O; Kirschner MW; Rosen MK, Chemical Inhibition of N-Wasp by Stabilization of a Native Autoinhibited Conformation. Nat. Struct. Mol. Biol 2004, 11, 747–755. [PubMed: 15235593]
11. De Mol E, et al., Epi-001, a Compound Active against Castration-Resistant Prostate Cancer, Targets Transactivation Unit 5 of the Androgen Receptor. ACS Chem. Biol 2016, 11, 2499–2505. [PubMed: 27356095]
12. Borgia A, et al., Extreme Disorder in an Ultrahigh-Affinity Protein Complex. Nature 2018, 555, 61–66. [PubMed: 29466338]
13. Zimmerman SB; Trach SO, Estimation of Macromolecule Concentrations and Excluded Volume Effects for the Cytoplasm of *Escherichia Coli*. J. Mol. Biol 1991, 222, 599–620. [PubMed: 1748995]
14. Zhou HX, Influence of Crowded Cellular Environments on Protein Folding, Binding, and Oligomerization: Biological Consequences and Potentials of Atomistic Modeling. FEBS Lett. 2013, 587, 1053–1061. [PubMed: 23395796]
15. Mikaelsson T; \AAdén J; Johansson LBA; Wittung-Stafshede P, Direct Observation of Protein Unfolded State Compaction in the Presence of Macromolecular Crowding. Biophys. J 2013, 104, 694–704. [PubMed: 23442920]
16. Soranno A; Koenig I; Borgia MB; Hofmann H; Zosel F; Nettels D; Schuler B, Single-Molecule Spectroscopy Reveals Polymer Effects of Disordered Proteins in Crowded Environments. Proc. Natl. Acad. Sci. USA 2014, 111, 4874–4879. [PubMed: 24639500]
17. Goldenberg DP; Argyle B, Minimal Effects of Macromolecular Crowding on an Intrinsically Disordered Protein: A Small-Angle Neutron Scattering Study. Biophys. J 2014, 106, 905–914. [PubMed: 24559993]
18. Banks A; Qin S; Weiss KL; Stanley CB; Zhou H-X, Intrinsically Disordered Protein Exhibits Both Compaction and Expansion under Macromolecular Crowding. Biophys. J 2018, 114, 1067–1079. [PubMed: 29539394]
19. Dedmon MM; Patel CN; Young GB; Pielak GJ, Flgm Gains Structure in Living Cells. Proc. Natl. Acad. Sci. USA 2002, 99, 12681–12684. [PubMed: 12271132]
20. Bai J; Liu M; Pielak GJ; Li C, Macromolecular and Small Molecular Crowding Have Similar Effects on A-Synuclein Structure. ChemPhysChem 2017, 18, 55–58. [PubMed: 27860069]
21. Waudby CA; Camilloni C; Fitzpatrick AWP; Cabrita LD; Dobson CM; Vendruscolo M; Christodoulou J, In-Cell NMR Characterization of the Secondary Structure Populations of a Disordered Conformation of A-Synuclein within E. Coli Cells. PLoS ONE 2013, 8, e72286. [PubMed: 23991082]
22. Theillet FX, et al., Structural Disorder of Monomeric A-Synuclein Persists in Mammalian Cells. Nature 2016, 530, 45–50. [PubMed: 26808899]
23. Rusinga FI; Weis DD, Soft Interactions and Volume Exclusion by Polymeric Crowders Can Stabilize or Destabilize Transient Structure in Disordered Proteins Depending on Polymer Concentration. Proteins 2017, 85, 1468–1479. [PubMed: 28425679]
24. Qin S; Zhou HX, Effects of Macromolecular Crowding on the Conformational Ensembles of Disordered Proteins. J. Phys. Chem. Lett 2013, 4, 3429–3434.
25. Kang H; Pincus PA; Hyeon C; Thirumalai D, Effects of Macromolecular Crowding on the Collapse of Biopolymers. Phys. Rev. Lett 2015, 114, 068303. [PubMed: 25723249]
26. Miller CM; Kim YC; Mittal J, Protein Composition Determines the Effect of Crowding on the Properties of Disordered Proteins. Biophys. J 2016, 111, 28–37. [PubMed: 27410731]
27. Feig M; Yu I; Wang PH; Nawrocki G; Sugita Y, Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. J. Phys. Chem. B 2017, 121, 8009–8025. [PubMed: 28666087]
28. Best RB; Zheng W; Mittal J, Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J. Chem. Theory Comput 2014, 10, 5113–5124. [PubMed: 25400522]
29. Piana S; Donchev AG; Robustelli P; Shaw DE, Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J. Phys. Chem. B 2015, 119, 5113–5123. [PubMed: 25764013]

30. Rauscher S; Gapsys V; Gajda MJ; Zweckstetter M; de Groot BL; Grubmuller H, Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. J. Chem. Theory Comput 2015, 11, 5513–5524. [PubMed: 26574339]
31. Hicks A; Zhou HX, Temperature-Induced Collapse of a Disordered Peptide Observed by Three Sampling Methods in Molecular Dynamics Simulations. J. Chem. Phys 2018, 149, 072313. [PubMed: 30134733]
32. Candotti M; Orozco M, The Differential Response of Proteins to Macromolecular Crowding. PLoS Comput. Biol 2016, 12, e1005040. [PubMed: 27471851]
33. Qin S; Zhou HX, Atomistic Modeling of Macromolecular Crowding Predicts Modest Increases in Protein Folding and Binding Stability. Biophys. J 2009, 97, 12–19. [PubMed: 19580740]
34. Qin S; Minh DD; McCammon JA; Zhou HX, Method to Predict Crowding Effects by Postprocessing Molecular Dynamics Trajectories: Application to the Flap Dynamics of Hiv-1 Protease. J. Phys. Chem. Lett 2010, 1, 107–110. [PubMed: 20228897]
35. Dong H; Qin S; Zhou HX, Effects of Macromolecular Crowding on Protein Conformational Changes. PLoS Comput. Biol 2010, 6, e1000833. [PubMed: 20617196]
36. Qin S; Zhou HX, Generalized Fundamental Measure Theory for Atomistic Modeling of Macromolecular Crowding. Phys. Rev. E 2010, 81, 031919.
37. Qin S; Zhou HX, FFT-Based Method for Modeling Protein Folding and Binding under Crowding: Benchmarking on Ellipsoidal and All-Atom Crowders. J. Chem. Theory Comput 2013, 9, 4633–4643.
38. Qin S; Zhou HX, Further Development of the FFT-Based Method for Atomistic Modeling of Protein Folding and Binding under Crowding: Optimization of Accuracy and Speed. J. Chem. Theory Comput 2014, 10, 2824–2835. [PubMed: 25061446]
39. Qin S; Zhou HX, Fast Method for Computing Chemical Potentials and Liquid-Liquid Phase Equilibria of Macromolecular Solutions. J. Phys. Chem. B 2016, 120, 8164–8174. [PubMed: 27327881]
40. Ozenne V; Bauer F; Salmon L; Huang JR; Jensen MR; Segard S; Bernadó P; Charavay C; Blackledge M, Flexible-Meccano: A Tool for the Generation of Explicit Ensemble Descriptions of Intrinsically Disordered Proteins and Their Associated Experimental Observables. Bioinformatics 2012, 28, 1463–1470. [PubMed: 22613562]
41. Feldman HJ; Hogue CW, Probabilistic Sampling of Protein Conformations: New Hope for Brute Force? Proteins 2002, 46, 8–23. [PubMed: 11746699]
42. Zhou HX, Dimensions of Denatured Protein Chains from Hydrodynamic Data. J. Phys. Chem. B 2002, 106, 5769–5775.
43. Yamazaki T, Parameterization of Coarse Grained Force Fields for Dynamic Property of Ethylene Glycol Oligomers/Water Binary Mixtures. J. Theor. Comput. Chem 2013, 12, 1350051.
44. Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc 1995, 117, 5179–5197.
45. Widom B, Some Topics in the Theory of Fluids. J. Chem. Phys 1963, 39, 2808–2812.
46. Bernado P; Blackledge M, A Self-Consistent Description of the Conformational Behavior of Chemically Denatured Proteins from NMR and Small Angle Scattering. Biophys J 2009, 97, 2839–2845. [PubMed: 19917239]
47. Zosel F; Haenni D; Soranno A; Nettels D; Schuler B, Combining Short- and Long-Range Fluorescence Reporters with Simulations to Explore the Intramolecular Dynamics of an Intrinsically Disordered Protein. J. Chem. Phys 2017, 147, 152708. [PubMed: 29055320]
48. Kjaergaard M; Norholm AB; Hendus-Altenburger R; Pedersen SF; Poulsen FM; Kragelund BB, Temperature-Dependent Structural Changes in Intrinsically Disordered Proteins: Formation of Alpha-Helices or Loss of Polyproline II? Protein Sci 2010, 19, 1555–1564. [PubMed: 20556825]
49. Borgia A, et al., Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. J. Am. Chem. Soc 2016, 138, 11714–11726. [PubMed: 27583570]

50. Song J; Gomes GN; Gradinaru CC; Chan HS, An Adequate Account of Excluded Volume Is Necessary to Infer Compactness and Asphericity of Disordered Proteins by Forster Resonance Energy Transfer. J. Phys. Chem. B 2015, 119, 15191–15202. [PubMed: 26566073]

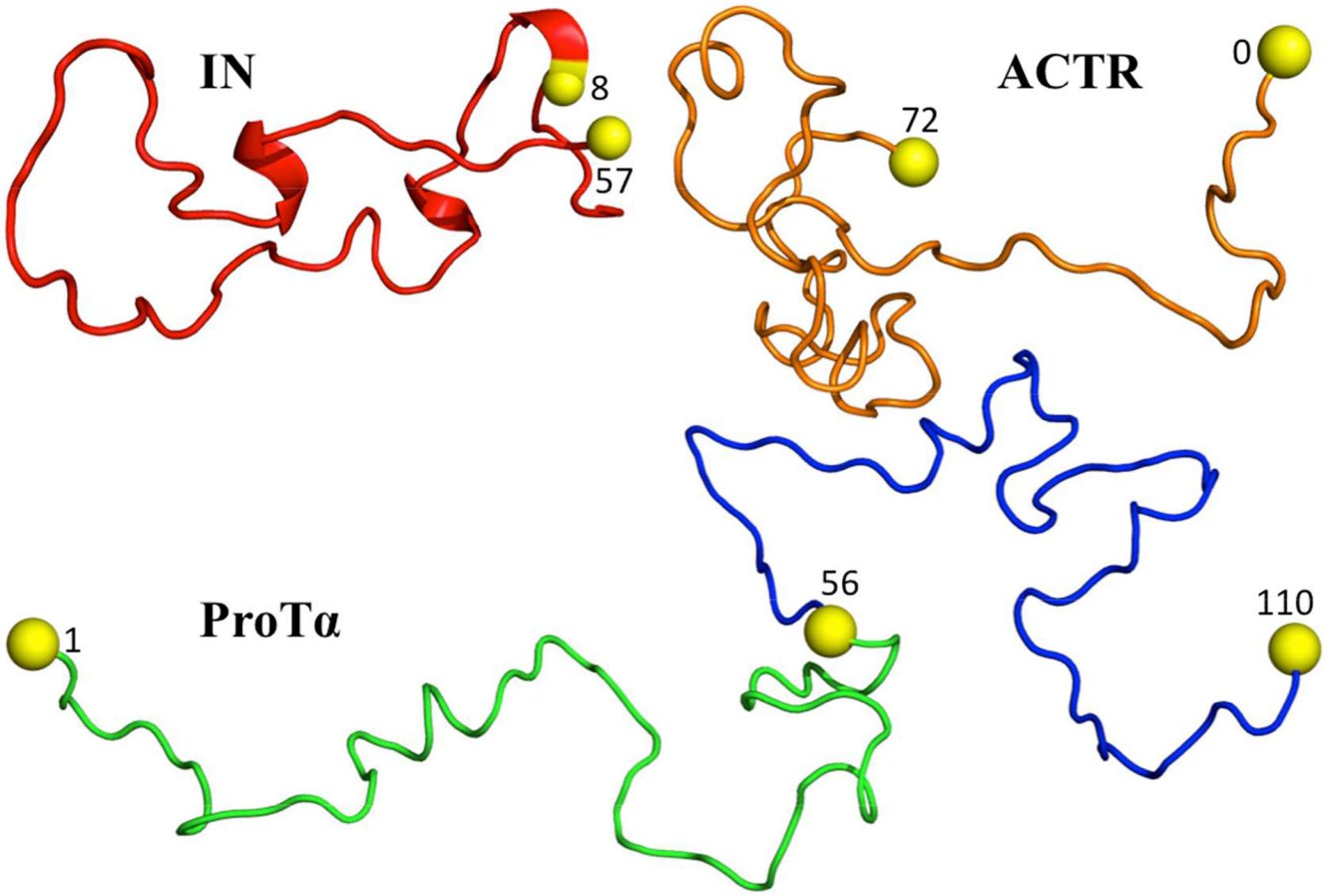


Figure 1.

The three IDPs studied. In each proteins, the Cys residues, with residue number indicated, for attaching fluorescence donors and acceptors are shown as yellow spheres. For ProTα, two fluorescence pairs were studied: one at residues 1 and 56 and the other at residues 56 and 110.

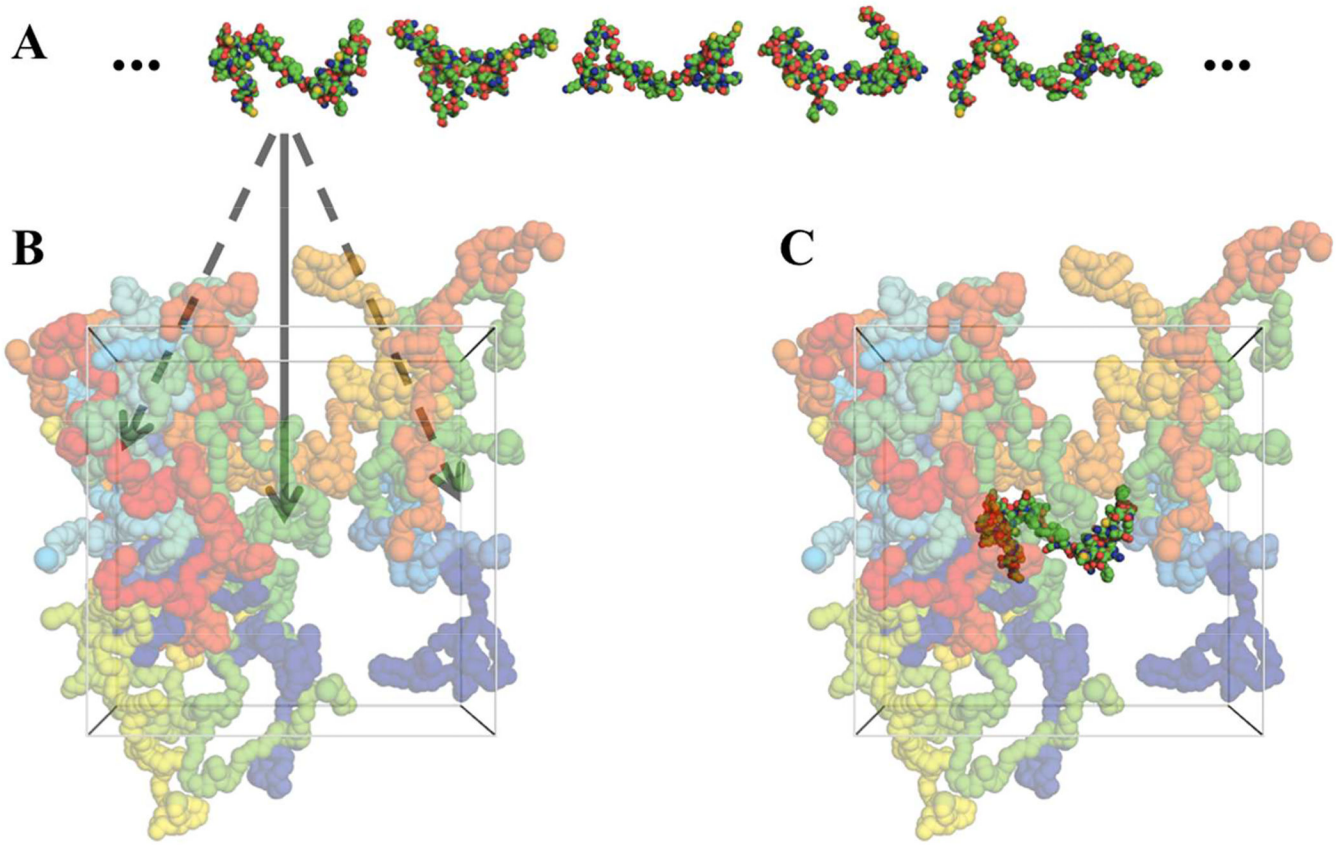


Figure 2.

Representative IDP conformations and PEG configurations and illustration of FMAP. (**A**) Representative conformations of IN. (**B**) A representative configuration of PEG 6000 at a 10% weight fraction. Arrows indicate fictitious insertions of an IN molecule into different positions inside the box of PEG molecules. (**C**) A successful insertion, whereby the inserted IN molecule is free of steric clashes with any PEG molecules.

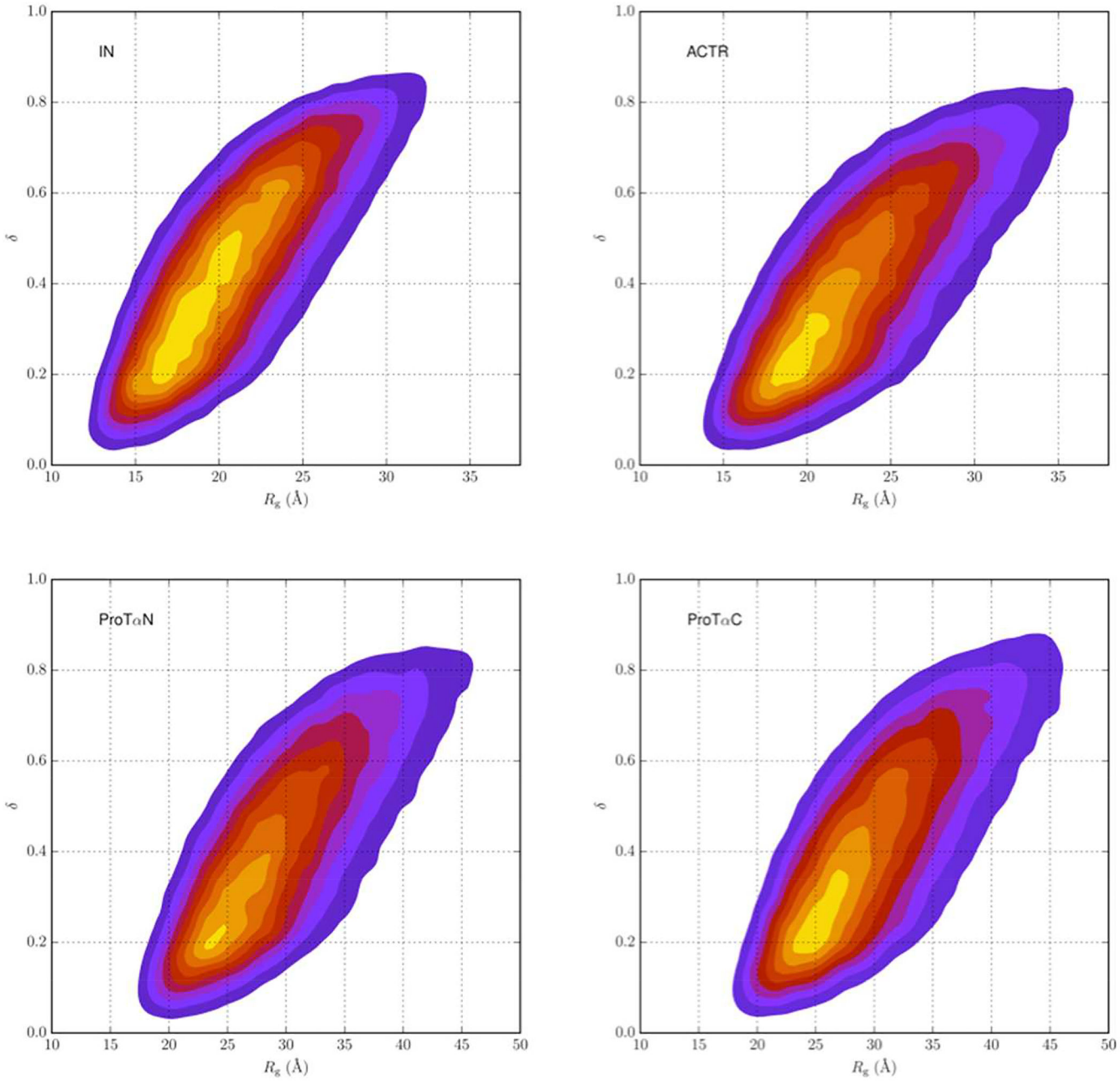


Figure 3.

Two-dimensional probability densities in radius of gyration and asphericity for four IDPs, calculated from reweighted flexible-meccano conformational ensembles.

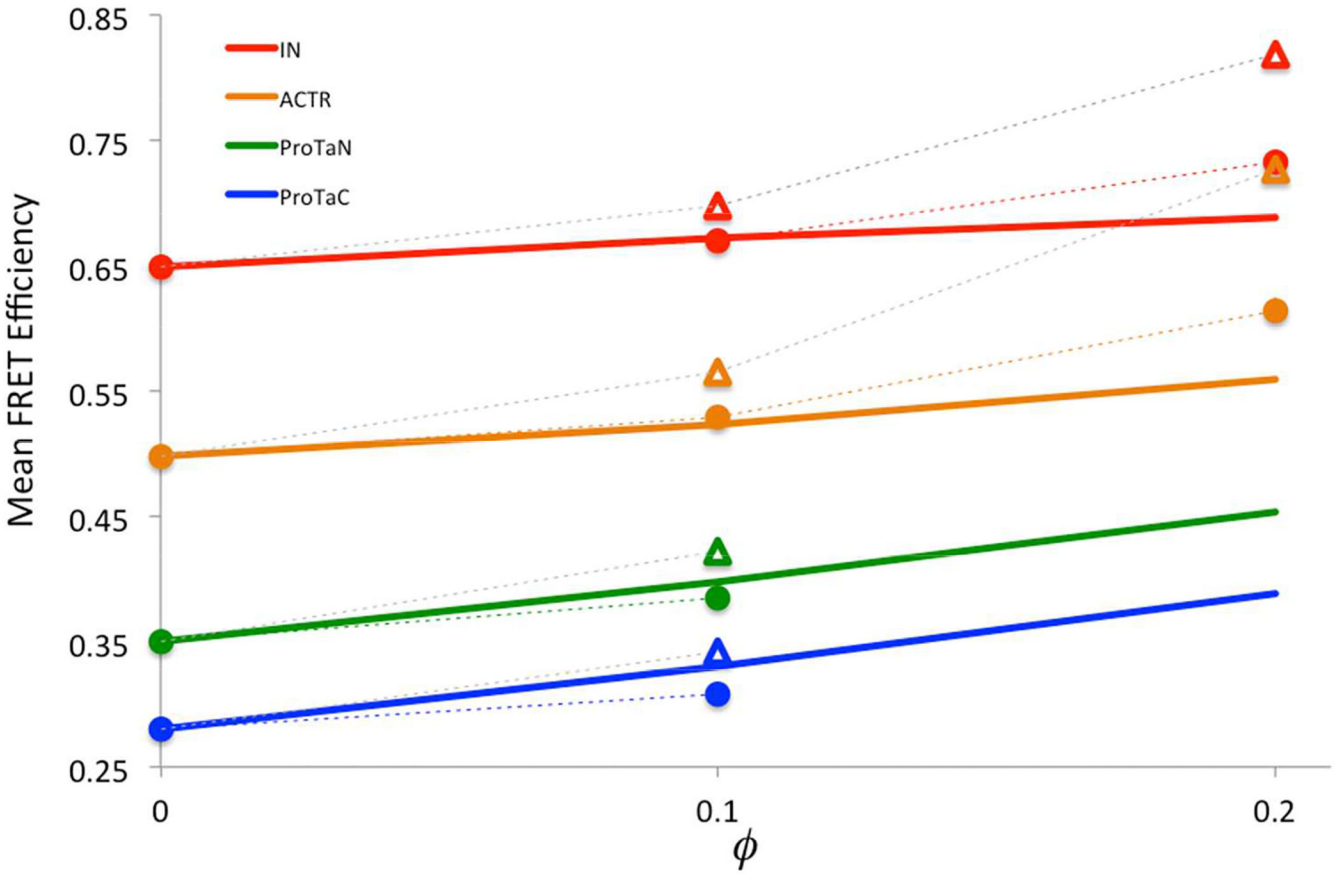


Figure 4.

Experimental and calculated mean FRET efficiencies. Experimental results are shown as solid lines, calculated results are shown as triangles or circles (connected by dash), when only steric repulsion or both steric repulsion and weak attraction between IDPs and PEG are accounted for. Calculated results are from applying FMAP on reweighted flexible-meccano conformational ensembles; errors, as estimated by bootstrapping, are less than the size of the symbols.

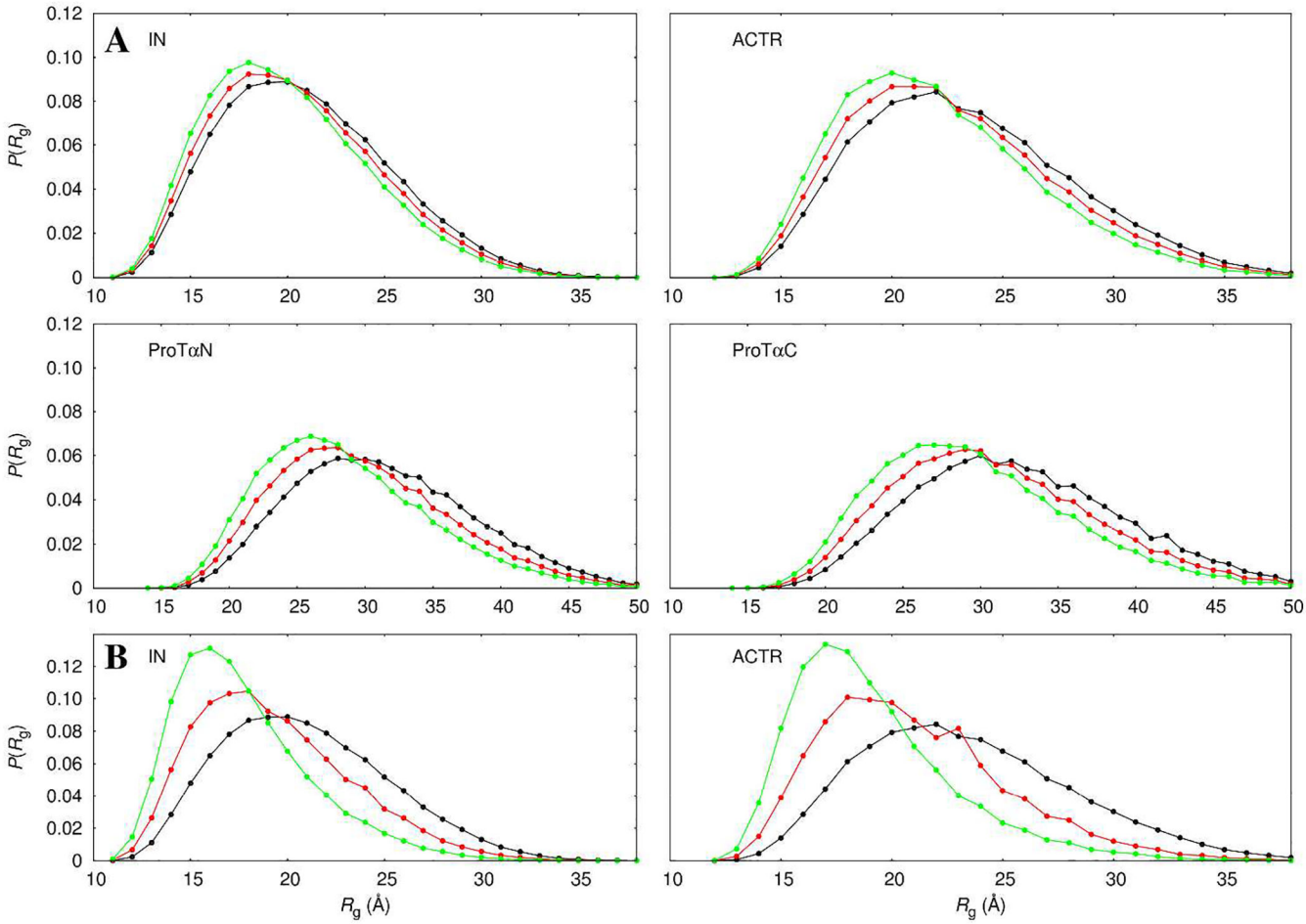


Figure 5.

R_g distributions of the four proteins in the absence and presence of PEG 6000. (**A**) 10% PEG 6000. (**B**) 20% PEG 6000. Results calculated from reweighted flexible-meccano conformational ensembles are in black; those shifted by PEG modeled with steric repulsion only are in green; and those shifted by PEG modeled with both steric repulsion and weak attraction are in red.

Table 1.

Conformational properties of IDPs in the absence and presence of crowders.[a]

		0%		10%			20%		
IDP	***b*** **(Å^{-1})**	**mean R_g (Å)**		$\bar{E}$			$\bar{E}$		
		raw	**reweighted**	**Expt**	**st only**	**st + at**	**Expt**	**st only**	**st + at**
IN	−.0029	21.13	20.94	0.672	0.698	0.670	0.689	0.819	0.733
ACTR	−.0021	23.61	23.43	0.523	0.565	0.529	0.560	0.727	0.614
ProTαN	.0450	28.66	31.42	0.398	0.422	0.385			
ProTαC	.0708	28.62	32.60	0.330	0.342	0.309			

[a]IDP conformations generated by flexible-meccano.