HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2017 September 14.

Published in final edited form as:

J Am Chem Soc. 2016 September 14; 138(36): 11702–11713. doi:10.1021/jacs.6b05443.

Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment

Wenwei Zheng[†,*], **Alessandro Borgia**[‡], **Karin Buholzer**[‡], **Alexander Grishaev**[§], **Benjamin Schuler**[‡], and **Robert B. Best**[†,*]

[†]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA [‡]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland [§]National Institute of Standards and Technology and the Institute for Bioscience and Biotechnology Research, Rockville, Maryland 20850, USA

Abstract

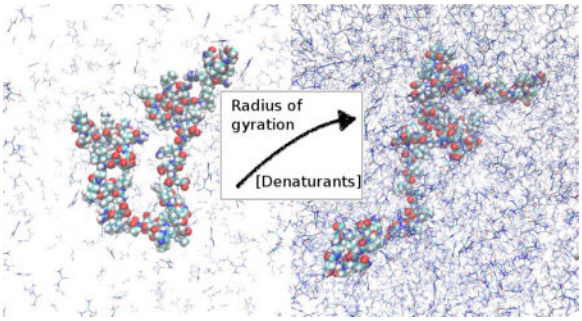
Chemical denaturants are the most commonly used agent for unfolding proteins, and are thought to act by better solvating the unfolded state. Improved solvation is expected to lead to an expansion of unfolded chains with increasing denaturant concentration, providing a sensitive probe of the denaturant action. However, experiments have so far yielded qualitatively different results concerning the effects of chemical denaturation, with studies using Förster resonance energy transfer (FRET) and other methods finding an increase in radius of gyration with denaturant concentration, but with small-angle X-ray scattering (SAXS) studies finding mostly no change. This discrepancy therefore challenges our understanding of denaturation mechanism, and more generally the accuracy of these experiments as applied to unfolded or disordered proteins. Here, we use all-atom molecular simulations to investigate the effect of urea and guanidinium chloride on the structure of the intrinsically disordered protein ACTR, which can be studied by experiment over a wide range of denaturant concentration. Using unbiased molecular simulations with a carefully calibrated denaturant model, we find that the protein chain indeed swells with increasing denaturant concentration. This is due to the favourable association of urea or guanidinium chloride with the backbone of all residues and with the side-chains of almost all residues, with denaturant-water transfer free energies inferred from this association in reasonable accord with experimental estimates. Interactions of the denaturants with the backbone are dominated by hydrogen bonding, while interactions with side-chains include other contributions. By computing FRET transfer efficiencies and SAXS intensities at each denaturant concentration, we show that the simulation trajectories are in accord with both experiments on this protein, demonstrating that there is no fundamental inconsistency between the two types of experiment. Agreement with experiment also supports the picture of chemical denaturation described in our simulations, driven by weak association of denaturant with the protein. Our simulations support some assumptions needed for each experiment to accurately reflect changes in protein size, namely that the commonly used FRET chromophores do not qualitatively alter the results, and that possible effects such as

[*]To whom correspondence should be addressed: wenwei.zheng@nih.gov, robertbe@helix.nih.gov.

Supporting Information: Supporting methods, figures and tables.

preferential solvent partitioning into the interior of the chain do not interfere with the determination of radius of gyration from the SAXS experiments.

Table of Contents



Introduction

Addition of chemical denaturants represents one of the most straightforward and widely used methods of perturbing the stability of proteins, since they are simple to use and rarely involve protein aggregation artifacts which may occur, for example, in thermal denaturation. To a good approximation, their action is the result of differential effects on the folded and unfolded states, for which several models have been proposed[1–3]. In the most accepted model, the denaturant binds weakly to the protein, favouring unfolding due to the greater surface available for binding in the unfolded state[1, 4–10]. At a coarser level, the consequence of such models is that solutions containing higher denaturant concentrations are better able to solvate proteins. Polymer theory predicts that a concomitant effect of this improved solvation should be a further swelling of the unfolded chain with increasing denaturant concentration[11–12], providing a sensitive measure of the denaturant action. Indeed, studies using a large number of experimental techniques have reached this conclusion[13], including ensemble and single-molecule Förster resonance energy transfer (FRET)[14–22], dynamic light-scattering (DLS)[10, 23], nuclear magnetic resonance (NMR)[24–26], and small-angle X-ray scattering (SAXS)[25, 27].

However, a significant number of studies, using SAXS on unfolded proteins[28–32], have reached a qualitatively different conclusion, namely that the radius of gyration (R_g) of the unfolded state of a two-state protein does not vary with denaturant concentration. Also for intrinsically disordered proteins[33], which are easier to study experimentally as the folded and unfolded states do not need to be separated, the results differ according to the method used: FRET experiments suggest an expansion with denaturant concentration for a number of IDPs[34–35], while SAXS experiments on the N protein, an IDP from bacteriophage lambda, were inconclusive regarding the change in R_g with added urea, after considering the errors[36]. The reasons for the differences between the conclusions drawn from these different experiments are unclear, but important to resolve. Firstly, the implication is that at least some of the experimental results, as currently analyzed, are incorrect. This would have wide-ranging implications because both SAXS and FRET experiments are frequently used to characterize IDPs. Secondly, if proteins do not expand with increasing denaturant

concentration, it means we must fundamentally re-evaluate our understanding of the denaturation mechanism. To the best of our knowledge, there is still no accepted solution to this controversy, and it has been identified by two recent reviews as one of the key outstanding problems in protein folding[37–38].

Leaving aside finer details of the experimental analysis, which are not the focus of the present work, it is possible that the apparent collapse or lack thereof is related to the experimental probes. While FRET can detect dimensional changes with unparalleled sensitivity even in heterogeneous samples and at very low sample concentration, it has been suggested that the hydrophobicity of the chromophores used to label the protein may stabilize a collapsed unfolded state at low denaturant concentration, making it difficult to determine the true collapse in the absence of chromophores[31–32]. SAXS requires highly homogeneous samples and relatively high protein concentration, but should provide a robust readout of the protein dimension, because the data at very low scattering angles are a direct measure of the R_g via the Guinier approximation. Assuming perfect experimental data, the only necessary assumption is a lack of systematic variation of solvent structure within the volume of the protein chain, which cannot easily be corrected for by the solvent subtraction procedure routinely used in SAXS data analysis. It is always expected that the composition of the surface layer around the protein will differ from the bulk solvent, but imperfect subtraction of this surface layer should not influence the long-range structural features probed at low scattering angles, since this layer would be highly correlated with the chain locus. The only way in which the inferred R_g could conceivably be altered is a preferential partitioning of denaturant molecules, for example, towards the center of the coil rather than at the periphery, or vice versa. This type of effect could arise, e.g., due to some denaturant molecules binding to multiple sites on the protein, leading to cooperativity and preferred binding where the chain is most dense. Differences in background contrast between denaturant and water molecules could then alter the apparent R_g of the solute. Whether such an effect could be strong enough to alter measurably the R_g, and approximately cancel any expansion with denaturant concentration, is unclear *a priori*.

The interpretation of most experimental data requires a simplified model of some type, for example a polymer model to obtain an average distance or R_g from a FRET experiment[39–40], or a continuum solvent model to interpret SAXS experiments[41], in order to solve the inverse problem of reconstructing molecular properties from a small number of observables (or sub-ensemble averages for FRET). Hence, the results may be sensitive to the specific model chosen. Alternatively, one can use atomistic molecular simulations with an accurate force field as a predictive tool, and employ the experimental data to validate the quality of simulations (since they are themselves based on empirically derived force fields)[42]. While the simulation model is much more complex, it has the advantage of not being fitted to the experiments it is meant to explain, i.e. the parameters are transferable to different proteins. If the simulations are quantitatively comparable to the experiments, they can then be used to provide molecular-scale insights into the observed phenomenon. Simulations have demonstrated their value in interpreting scattering data on unfolded proteins at large scattering angles, where analytical models may be insufficient[43]. Molecular simulations have already been used in a large number of studies to determine the effect of chemical denaturants on protein stability[5, 44–45], on the unfolded state[44, 46–47] and on the mechanism

by which they denature proteins[4–5, 8, 46, 48–53]. However, the rapid denaturation observed in some of these studies suggested that many of the force fields may not quantitatively capture the effect of denaturants on protein stability. Indeed, a recent study by Netz and co-workers found that, while the best combination of protein and urea force fields they tested reproduced well the variation in denaturant affinity from one amino acid to another, the affinity of urea for each residue type was ~0.5 kcal/mol too favourable, per residue[6]. Other studies have also found too strong an association of urea with peptide models using a variety of force fields, and different methods[5, 54]. Clearly, for any study aiming to capture quantitatively the effect of denaturant on unfolded proteins, the simulation model must, at a minimum, reproduce the affinity of the denaturant for the chain.

To address this problem, we have recently parameterized models for urea and guanidinium chloride (GdmCl) in order to achieve a good balance between protein-protein, protein-water and protein-denaturant interactions. Water-protein interactions were first tuned by matching data for a short unfolded peptide, and cross-validated against multiple other experiments[55]. Protein-denaturant interactions were adjusted by scaling these interactions to match experimental solubility data for a tetraglycine peptide. The resulting force field was shown to reproduce denaturant-dependent FRET efficiencies of a fragment of the protein CspTm[35] in which the chromophores were explicitly represented, as well as *m*-values for denaturation of the Trp cage miniprotein[44]. Here, we apply this force field to study the denaturation of the intrinsically disordered protein ACTR (activator for thyroid hormone and retinoid receptors)[56–57] in urea and GdmCl over a range of denaturant concentrations. The advantage of studying an IDP is that in experiments there is no signal from the folded state which needs to be separated, and so, even at equilibrium, data can be recorded down to the lowest denaturant concentrations where the largest change in FRET efficiencies is typically observed. We show first that we are able to capture the experimental FRET efficiencies and X-ray scattering intensities from unbiased simulation trajectories, within the estimated statistical uncertainty of the simulation. The simulations show an expansion of the chain with increasing denaturant concentration, demonstrating that such an expansion can be consistent with both SAXS and FRET experimental results. We have also compared our results with experimental measurements of transfer free energies, and analyzed in detail the contributions made by different groups in the protein to these free energies. We find that urea associates favorably with almost every residue in the protein, explaining the improved solvation implied by chain expansion.

Having shown that we can reproduce the experimental data adequately, the simulation results allow us to test potential molecular scale artefacts which may confound the interpretation of FRET and SAXS data. For SAXS, we investigate the potential effect of solvent structure on the measured SAXS intensity by comparing the signal computed from an all-atom simulation to that calculated using a continuum model for the solvent, as well as by analyzing the first solvation layer in more detail. For FRET, we compare results from simulations with and without explicit representation of the chromophores, to test their influence on the degree of collapse, and some of the approximations which need to be made in order to estimate intramolecular distances from FRET efficiencies.

Methods

Molecular Simulations

All-atom simulations were run using the Gromacs 4.6.7[58] simulation code at a constant temperature of 298 K (maintained by a Langevin thermostat) and pressure of 1 bar (with a Parrinello-Rahman barostat[59]). The time step was 2 fs, electrostatic energies and forces were computed with particle-mesh Ewald[60] using a 0.12 nm grid spacing and real-space cut-off of 0.9 nm. Lennard-Jones interactions were calculated using a twin-range scheme with inner and outer cut-offs of 0.9 and 1.4 nm. The Amber ff03ws force field[55] was used for the protein together with the TIP4P/2005 water model[61] and KBFFs model for urea and GdmCl[44] (i.e. the Kirkwood-Buff force field (KBFF) model[62–63] including scaled denaturant-protein interactions in order to balance protein-denaturant interactions). The dye force field was described in a previous work[64]. Details of system size, composition and run length are summarized in Table S1, and the sequences of the peptides simulated in Table S2.

SAXS Calculations

All-atom SAXS calculations were performed using the algorithm described by Köfinger and Hummer[65], and is briefly described here. The SAXS intensity $I(q)$ is calculated by

$$I(q)=\sum_{i,j} f_i(q) f_j(q)[\Delta I_{\mathrm{ij}}(q)+v I_{\mathrm{ij}}(q)] \quad (1)$$

in which ΔI_{ij} represents the partial intensity difference between protein with solution (foreground) and pure solution (background), and $f_i(q)$ the form factor of species i. The term $vI_{ij}(q)$ adds back the bulk solvent contribution oversubtracted in I_{ij}. This correction is not used in the current work, since it has been shown in the original literature to have limited influence on the SAXS intensity for the range of q we are interested in ($q < 0.5$ Å^{-1}). For each denaturant condition with SAXS calculation, two sets of MD simulations, with and without a protein molecule, were set up with the same number of denaturant molecules and ions. The protein was replaced with additional water molecules in the pure solvent simulation to make the volume of the background simulation the same as the foreground. ΔI_{ij} can then be calculated from all-atom MD data by

$$\Delta I_{\mathrm{ij}}(q)=\delta_{\mathrm{ij}}\Delta N_i+\int_0^{2R}\frac{\Delta H_{\mathrm{ij}}(r)\sin(\mathrm{qr})}{\mathrm{qr}}\mathrm{dr} \quad (2)$$

in which ΔN_i is the difference of the average number of particles of species i between the foreground and background simulations, and ΔH_{ij} are the difference distribution functions of interparticle pair distances between the foreground and background simulations, and R is the radius of the sphere in which foreground observations are made.

SAXS calculations with an implicit solvent model were performed with the programs CRYSOL[41] and FOXS[66]. Here, we briefly describe CRYSOL and refer the readers to the original literature for more details. The scattering intensity is calculated by three terms:

$$I(q)=\langle|A_a(\boldsymbol{q})-\rho_w A_c(\boldsymbol{q})+(\rho_b-\rho_w)A_b(\boldsymbol{q})|^2\rangle_\Omega$$

the first scattering amplitude is that of the protein in *vacuo*, the second is from the volume excluded to the solvent, and the third the one from the surface hydration layer with a higher density than the bulk solution. ρ_b and ρ_w represent the scattering density of surface layer and pure solvent, respectively, and Ω an average over a uniform distribution of macromolecular orientations relative to the incident beam. For a conformational ensemble additional averaging needs to be performed over the $I(q)$ profiles calculated for each ensemble member. Since the hydration layer is empirically estimated, just protein coordinates are required to calculate the SAXS intensity using CRYSOL. Here we use the protein coordinates only from the same all-atom MD simulation data.

FRET Calculations

In most cases, FRET efficiencies E were calculated based on the donor-acceptor distance R, assuming that the orientational dynamics of donor and acceptor chromophores was fast compared to the fluorescence lifetime,[67] so that the orientational factor $\kappa^2 = 2/3$, and that the distance dynamics within the chain are slow relative to fluorescence lifetime of the donor[67–68], i.e.

$$E=\left\langle\frac{1}{1+(\frac{R}{R_0})^6}\right\rangle \quad (3)$$

where the averaging is over all frames of the trajectory, and R_0 is the spectroscopically determined Förster radius[69] for the donor and acceptor dyes used here, AlexaFluor 488 and AlexaFluor 594[14]. R_0 is defined from the refractive index, n, and the R_0 in the absence of denaturant, $R_0(0) = 5.4$ nm, as[69]:

$$R_0(n)=\left(\frac{(n(0))^4\cdot(R_0(0))^6}{n^4}\right)^{\frac{1}{6}}, \quad (4)$$

In this expression, $n(0)$ is the refractive index at 0 M and

$$n([\mathrm{Urea}])=1.3361+0.00841[\mathrm{Urea}],$$

in which [Urea] is the urea concentration (this curve is determined for a solution of 50 mM sodium phosphate, 140 mM β-mercaptoethanol and 0.01% Tween 20)[70]. It is assumed that changes in the donor quantum yield and spectral overlap integral do not significantly change with denaturant concentration. Since the chromophores were not present in most of the simulations, in these cases R was calculated between the C_α of residues X and Y labeled in the experiment. The distance is then rescaled by a factor of

$$\gamma = \left[\frac{N + N_{\text{linker}}}{N}\right]^{\nu} \quad (5)$$

in which N is the number of bonds between the FRET dyes in the experiment, ν is the Flory scaling exponent determined from the scaling of internal chain distances with sequence separation in the simulation (Figure 1),[71] and N_{linker} is a free parameter representing the linker length. It has previously been estimated empirically to be ~9[17, 35].

For the case in which the chromophores were explicitly simulated, the FRET efficiency can be calculated in three different ways. The first is to use the equation and correction factor described above, in which distances between C_α are used. The second is to use the distances between the "C_1" atoms of each chromophore as described previously[64] without a correction factor. A more sophisticated approach, which assumes only that Förster theory is sufficiently accurate, can also be applied to the simulations including explicit donor and acceptor chromophores[64, 72–74]. In this case, the transfer rate $k_{\text{ET}}(x)$ for configuration x in the simulation trajectory is given by

$$k_{\text{ET}}(x) = \frac{3}{2} k_{\text{D}} R_0^6 \frac{\kappa^2(x)}{R^6(x)} \quad (6)$$

where k_{D} is the donor fluorescence decay rate in the absence of an acceptor, and the orientational factor κ is given by

$$\kappa = \hat{\mu}_{\text{D}} \cdot \hat{\mu}_{\text{A}} - 3(\hat{R} \cdot \hat{\mu}_{\text{A}})(\hat{R} \cdot \hat{\mu}_{\text{D}}), \quad (7)$$

where $\hat{\mu}_{\text{D}}$ and $\hat{\mu}_{\text{A}}$ are unit vectors in the direction of the donor and acceptor transition dipoles, respectively, and $\hat{R}$ is a unit vector pointing between donor and acceptor. We assume that the donor and acceptor transition dipole moments are approximately aligned with the long axis of each chromophore system (defined by the vectors between atoms C_{11} and C_{12} within each chromophore), and the distance between the chromophores is taken to be that between the C_1 atoms of each chromophore[64]. The decay in donor fluorescence intensity is evaluated by calculating the survival probability of the excited state with a fluctuating transfer rate, averaged over all possible time origins, t_0, along a simulation trajectory:

$$I(t) = \left\langle \exp\left[-\int_0^t (k_D + k_{\text{ET}}(t_0 + \tau)) d\tau\right] \right\rangle_{t_0} \quad (8)$$

The average FRET efficiency was obtained by integration of the intensity decay (or lifetime distribution)

$$\langle E \rangle = 1 - k_{\text{D}} \int_0^{t_{\max}} I(t) \mathrm{d}t, \quad (9)$$

where the maximum integration time t_{max} was chosen as 20 ns, by which time the fluorescence had essentially decayed to zero for $k_D = 0.238\ ns^{-1}$.

Results and Discussion

ACTR expands in denaturant solution

To sample the configurations of the intrinsically disordered protein ACTR, we ran multiple unbiased, 2 μs-long, equilibrium MD simulations in explicit water and different denaturant concentrations (Table S1). Such extensive trajectories, while still posing a challenge for the large systems considered, are the minimum necessary to obtain a representative sampling, given that the experimental reconfiguration times of unfolded and disordered proteins are typically of the order of 0.05–0.1 μs[17, 75–77]. We use force field models for protein, urea and water which we have recently parameterized to reproduce the balance of interactions between the protein, water, and denaturant components of the system[44, 55]. We note that using such a force field is essential, because recent work has shown that most existing force fields result in too collapsed conformations of proteins even in the absence of denaturant[78–79], with several suggested corrections proposed[55, 80–81]. This would confound any attempt at quantitative comparison with experiment[55]. Although we consider the effects of both urea and GdmCl, in the interest of brevity, we describe only the results for urea in the main text (see Supporting Information for GdmCl).

In Figure 1a, we show the fluctuations in R_g (computed directly from the protein coordinates) over the course of representative simulations at each denaturant concentration. The relatively long time scale of fluctuations necessitates sampling on the microsecond time scale. In Figure 1b, we show the autocorrelation function for the radius of gyration, which yield correlation times ranging from around 40 to 140 ns, comparable to those measured in earlier experiments on other proteins[76–77]. Even though the distributions of R_g are very broad, there is nonetheless a clear increase in its average value as a function of denaturant concentration, illustrated in Figure 1c, as well as in the average distance between the residues labeled with chromophores (Figure 1d). The swelling of the chain is also reflected in an increase in the scaling exponent with denaturant concentration. We have characterized this by means of a power law fit of the dependence of the root-mean-square (RMS) inter-residue distance between pairs of residues on the sequence separation of those residues, Figure 1d[71]. In addition, to obtain better averaging, we have also computed the RMS R_g of the chain segment included between pairs of residues on their sequence separation, Figure 1c. The fits, summarized in Table 1, show an increase in scaling exponent with denaturant concentration, from approximately 0.55 in the absence of denaturant, to a value slightly larger than 0.6 in high denaturant. These exponents are comparable, respectively, to the trend obtained from single-molecule FRET experiments in low and high denaturant, which show a transition from near θ-solvent conditions in water to close to the excluded volume limit in denaturant[35]. The finding of near theta-solvent conditions in water is also consistent with the fractal dimension from SAXS experiments on reduced RNase A in water[82], while the exponent at high denaturant is in accord with the scaling inferred from a comprehensive small-angle X-ray scattering study of a wide range of sequence lengths in high GdmCl concentration[83]. Therefore, our results are consistent, qualitatively, with the expectations of

polymer theory for the changes which occur when the solvent quality is improved[12]. In Figure S1, we show corresponding results for ACTR in GdmCl solutions. Note that although the chain collapses as denaturant is diluted, its most collapsed state, in water, still approximates a θ-state[35]. Thus, in analogy with protein folding[84–85], ACTR does not form a fully collapsed state prior to binding its cognate partner, NCBD. For an intrinsically disordered protein, maintaining an expanded state may have advantages for recognition of binding partners and for binding kinetics[86–87].

Denaturing mechanism of urea

The clear expansion of the chain in urea implies an improvement in solvent quality with increasing denaturant concentration (an alternative explanation for increased R_g might be an increase in chain stiffness, but that would not explain the increase of scaling exponent). The improved solvent quality could be thought of in terms of urea molecules "binding" to the protein, as previously inferred from experimental studies using NMR and X-ray scattering[88]; however, for such weak binding occurring at high denaturant concentration, it is critical to remove the contribution from denaturant molecules which happen to be near the protein but are not necessarily interacting. Therefore, in order to characterize in more detail the weak interactions between the protein chain and denaturant molecules, we use the formalism of preferential interaction coefficients. The preferential interaction coefficient Γ_{UP} is defined experimentally as $\Gamma_{UP} = (\partial m_U/\partial m_P)_{\mu_U}$, where m_U and m_P are the molalities of urea and protein, respectively[89–90]. That is, Γ_{UP} measures how much urea must be added to keep the bulk urea chemical potential μ_U constant when a protein is added to the solution and is expected to be positive if urea interacts favourably with the protein, and vice versa. In simulations, the coefficient can be estimated very simply from the heuristic relation:[89–92]

$$\Gamma_{UP} = \langle n_U^P - n_W^P \left(\frac{n_U^B}{n_W^B} \right) \rangle \quad (10)$$

In this equation, n_U^P and n_W^P are the number of urea and water molecules in a defined volume close to the protein, while n_U^B and n_W^B are the corresponding numbers in the bulk solution away from the protein, i.e. Γ_{UP} is the average number of urea molecules in the volume near the protein, in excess of what would be expected based on the bulk solution composition. We define the volume near to the protein by using a simple cut-off of 0.7 nm between protein heavy atoms and the water oxygen or urea carbon, however the results are fairly insensitive to the choice of cutoff, as long as it is large enough (Figure S2). We can in addition write the total Γ_{UP} as a sum over group contributions, by assigning the water and urea molecules in the protein domain to the group on the protein to which they are closest, corresponding to a Voronoi tessellation of the domain surrounding the protein.[93–94] The groups we have chosen are the backbone and side-chain heavy atoms of each residue.

In Figure 2, we show the decomposition of the preferential interaction for backbone and side-chains for each residue type. We see that urea interacts favourably with the backbone of all residue types, although there is some residue to residue variation. While we have presented the average Γ_{UP} for each residue type, we note that its value is relatively

independent of the sequence context, with similar results being obtained for all residues of a given type (Figure S3). This supports one of the assumptions of the commonly used additive schemes for decomposing protein-denaturant interactions,[1, 9, 95] for example the decomposition of protein folding *m*-values as a sum over independent contributions from different functional groups in the polypeptide chain.[9] On the other hand, the association with glycine, which is often used as a model for the protein backbone in decomposition schemes, is notably higher than the average value (compare dashed and dotted lines in Fig. 2), which may lead such schemes to underestimate the contributions from side-chains. This may reflect some of the known limitations of assuming additivity in calculations of protein-solvent interactions,[96] although we must also concede that our decomposition of space using a Voronoi scheme is certainly not unique. The side-chain contributions show that urea also interacts favorably with almost all side-chains, the only exceptions being the anionic aspartate and glutamate residues, consistent with an earlier study[8]. Our results thus suggest that both backbone and side-chains contribute comparable amounts to the favorable solvation of the unfolded state by urea solutions, in agreement with the results of other recent computational studies[7, 92]. Note that this does not mean they contribute equally to folding *m*-values, which measure how the difference between the folded and unfolded $\Delta\mu_{tr}$ changes with denaturant concentration. A calculation including the folded state (or at least, a fully collapsed state[97]) would be needed to evaluate the relative contribution of the backbone to folding *m*-values.[98]

Since experimental preferential interaction coefficients are not available for all residue types, we have compared our results with per-residue transfer free energies from water to 1 M urea[95]. We estimate transfer free energies $\Delta\mu_{tr}$ from preferential interaction coefficients using the approximate relation $\Delta\mu_{tr} \approx -RT\Gamma_{UP}$. This expression is valid at low denaturant concentrations and for ideal denaturant solutions[90]. A concentration of 1M is sufficiently low for the first assumption to be valid and urea solutions are known to be very close to ideal[99] (also reflected in properties of the KBFF force field[62]). This expression also ignores any systematic changes in protein dimensions with denaturant concentration, which we have just shown to occur. However, given that the percentage increase in protein size is modest, we feel this is also a reasonable first approximation. The Pearson correlation coefficient of the transfer free energies between the simulation and experiment is 0.63 with a p-value of 0.01, suggesting the calculated values capture very well the overall magnitude and sign of the protein-urea interactions, and to a good extent the variation from residue to residue. We note that a direct comparison cannot be made with the charged residues, because their $\Delta\mu_{tr}$ refers to transfer free energies for their Na^+ or Cl^- salts, and so may include significant contributions from the transfer free energies of these ions. The total preferential interaction or transfer free energy depends both on the size of the residue, as well as on its chemical identity. We can approximately normalize for the contribution from size by dividing by the average Connolly solvent-accessible surface area, shown in Figure 2. After this correction for size, the interaction coefficients from simulation are quite similar for most of the residues, the remaining outliers being the charged residues, and glutamine.

We have analyzed further the mechanism of action of urea with both backbone and side-chains, starting with hydrogen bonding, which is the easiest type of interaction to single out (Figure 3). We find that for most residue types, the average number of hydrogen bonds

between the backbone and urea at 1 M denaturant is very close to the number of excess urea molecules, relative to bulk (given by the preferential interaction coefficient). This strongly suggests that hydrogen bonding is the main mode of interaction with the backbone, with many of the side-chains also making hydrogen-bonded interactions with urea. In addition, however, it is clear that the hydrophobic side-chains interact favorably without forming any hydrogen bonds. An apparently more puzzling result is the negative preferential interaction of some of the charged side-chains with urea, despite the number of hydrogen bonds to water being similar for analogous charged and uncharged side-chains (e.g. Asp and Asn). The most likely explanation is an enhanced local water density in the vicinity of the ionic side-chain, such that the average number of urea molecules per water molecule is still lower than in bulk. A high local density of water dipoles helps to solvate the charged side-chains, and indeed we observe a very large first peak in the water g(r) around Asp, relative to Asn (Figure S4): This higher water density is a manifestation of the well-known electrostriction effect of ions. Since there are no residues with aromatic side chains in ACTR, yet these usually have the most favorable water-urea transfer free energies in experiment[95], we have calculated the preferential interaction coefficients of the unfolded Trp cage mini protein using published simulations with the same force field in 3M urea[44] (Figure S5). We find qualitatively that Trp and Tyr have much larger preferential interaction coefficients and therefore more favorable transfer free energy in Trp cage, consistent with experiment. Based on this, one would expect denaturant to lead to a larger expansion in denaturant for sequences containing also aromatic residues.

Although in the main text we focus on urea, in Figure S6, we show the corresponding results for 1 M GdmCl. Similar to urea, we find that both backbone and side-chains make comparable contributions to transfer free energies: while only the Gdm^+ ions have a significant preferential interaction with the backbone, both Gdm^+ and Cl^- ions associate favorably with side-chains in simulation (Figure S7). The major difference is the stronger interaction of the cationic guanidinium ion with the anionic residues. Quantitatively, the magnitude of the interaction coefficients and transfer free energies for GdmCl are about twice the values for urea, consistent with the stronger effect of this denaturant, as well as with experimental transfer free energies reported by Nozaki and Tanford[100]. The denaturant-dependence of protein stability (m-value) is also usually a factor of ~2 larger in GdmCl than in urea, consistent with the dominant role of the unfolded state in determining m[97], although the native state must also contribute[98]. The backbone preferential coefficient is still comparable to the number of hydrogen bonds formed per residue, similar to the urea case (Figure S8), indicating that, according to our simulations, both Gdm^+ and urea interact by hydrogen bonding with the backbone. However, the type of hydrogen bonds formed is different, with Gdm^+ hydrogen bonding exclusively to the CO group of the amide bond (as may be expected from its lack of hydrogen bond acceptors), and urea to both the NH and CO groups (Figure S9). We can compare these results with earlier hydrogen exchange experiments in the presence of urea or GdmCl[101]. Base-catalyzed exchange was found to be blocked by urea and unaffected by GdmCl, which is expected as the base would attack the NH, which can only be blocked by urea hydrogen bonding. The results for acid-catalyzed exchange indicate that urea accelerates exchange while GdmCl has little effect, which was interpreted to mean that Gdm^+ does not hydrogen bond to the CO group either[101], in

contrast to what we find. However, a definitive conclusion based on experiment would require a quantitative model for the *expected* effect of the denaturant on the rate of acid catalysis, which has a more complex mechanism than base catalysis[102]. Overall, our analysis suggests that the effect of both urea and GdmCl can be explained in terms of preferential solvent partitioning, which essentially describes a weak binding of the denaturant to the protein[1], the model favored by most recent studies[4–8, 49–50, 52, 103], and consistent with our results.

The stronger interactions of the protein with the solvent imply relatively weaker protein-protein interactions, which should disrupt any local structure (native or non-native) formed at low denaturant concentration. ACTR is known to be quite unstructured in water, however it does have some residual helical structure which is lost at high urea concentration, as probed by ultraviolet circular dichroism (CD), as well as nuclear magnetic resonance spectroscopy[56, 104]. We have computed average helix fraction as a function of denaturation concentration, and while the data exhibit considerable noise, there does appear to be a modest decrease in helix fraction with increasing denaturant concentration, in good agreement with the helix fraction inferred from CD, considering the statistical error in the simulation (Figure S10), and in accord with the finding that proteins populate more extended structure in denaturant than in water.[105] This loss of helix when the protein expands at higher denaturant concentration is in contrast with the situation when the temperature is raised, which causes ACTR to collapse (due to strengthened hydrophobic effect[106]), but also an apparent reduction of helix content[104], as has also been observed in all-atom simulations of unfolded proteins[107]. Thus there is not a simple connection between the collapse and the formation of helical structure, and collapse can be driven by the different types of interaction, depending on the conditions.

Comparison of Simulations with FRET and SAXS

To validate the results of our simulations, we compare the raw experimental data[70] with that calculated from the all-atom simulations. In Figure 4, we show the mean FRET efficiency computed from the simulations using the Förster equation (Equation 3) as a function of urea concentration, together with the experimental results, for three different pairs of residues labeled with FRET donor and acceptor chromophores. There is naturally a considerable statistical uncertainty in our estimates, given the quantity of data available. Especially at high denaturant concentration, there is a deviation of FRET efficiency between the simulation and experiment. This is probably due to the limited box size affecting end-end distance of more expanded configuration and significantly lower viscosity of the solution at high denaturant concentration. We note that the simulations with a larger solvent box do agree better with experiment at high urea concentration. This may reflect an absence of interactions with the periodic image, but with the caveat that the simulations with the larger boxes are only 0.6 μs versus 2 μs for the small box simulations. Even with the deviation at high urea concentration, the Pearson correlation coefficient of the FRET efficiencies between the simulation and experiment is 0.91 with a p-value in the order of 10^{-7}, suggesting the agreement between experiment and simulation is overall quite good. In Table 2, an all-vs-all comparison of simulation and experimental efficiencies at different concentrations shows that the best agreement is obtained when the concentrations in

simulation and experiment are the same, or nearly the same, implying that the expansion we observed in simulation is also present in experiment. We note that, since the chromophores were not explicitly present in the initial set of simulations, we have accounted for the effects of the protein-chromophore linkers by scaling the separation between the C_α of the labeled residues (Equation 5). In addition, we assume that the efficiency is determined only by the donor-acceptor distance and that the FRET orientational factor $\kappa^2 = 2/3$[108]. We will revisit and justify both of these assumptions in the next section.

We have also computed SAXS scattering profiles $I(q)$ using the all-atom coordinates via an established procedure[65]. To do this, we compute the atom-atom pair distance distribution functions (PDDFs) within a spherical volume around the protein centre of mass, whose summed Fourier transforms yield the scattering intensities. The background is computed from a large simulation box of denaturant solution with a similar concentration, as in the experiment. Contrast matching is performed by comparing the average electron density in a shell outside of the primary sphere in the protein simulation with the electron density in the reference (background) simulation. This calculation, therefore, exactly mimics the experiment, and includes any possible contributions due to cooperative solvent structuring around the protein. We found that the essential parameters in this calculation are the radius of the primary sphere, and the thickness of the surrounding solvent shell used for contrast matching. As we discuss in the supporting text, and show in Figure S11, choosing a radius for the primary sphere which does not completely contain the vast majority (i.e. ~99%) of the disordered protein configurations, distorts the results, for example giving an underestimation of the simulated R_g based on a Guinier approximation. A second requirement is that the solvent shell for contrast matching must be thick enough to be representative of the solvent background. These requirements led us to adjust the system size for the different denaturant concentrations according to the protein R_g, as shown in Table S1. Note that our observed variation in R_g is not an artefact of confinement due to the smaller system sizes used at low denaturant concentration. Within statistical error, we obtain the same radii of gyration when using the same system size for all systems (see Figure 1), the larger system size at high denaturant concentrations only being required for the explicit SAXS calculation. The computed scattering profiles, $I(q)$, are shown in Figure 4 for different denaturant concentrations together with the experimental data. Although the curves at different denaturant concentrations all appear superficially very similar, we find that the simulations capture the subtle differences between them. An all against all comparison of the simulated curves at different denaturant concentrations with the experimental curves at different concentrations shows that, in most cases, the best agreement of the experimental data with simulation (assessed by the reduced χ^2 parameter) occurs when the denaturant concentrations in experiment and simulation are the same (Table 2), so that again the expansion of the chain seen in simulation is consistent with experiment.

Influence of FRET probes

For FRET to yield an accurate estimate of molecular size it is important that the chromophores do not substantially affect the radius of gyration, or its denaturant dependence – it has been implied that the chromophore labels may somehow influence the denaturant dependent collapse[31–32]. In the results described so far, we have used the same simulations

for both FRET and SAXS calculations in order that the results be as comparable as possible. We have also tested the assumption that chromophores should not noticeably perturb the protein by performing simulations of ACTR in urea with explicit chromophores at two different denaturant concentrations, 1 M and 5 M. The force field for the chromophores has been found to reproduce fairly well a battery of experimental data on the interaction of chromophores with zwitterionic tryptophan and on chromophores attached to proteins and peptides[64]. An initial comparison of the radius of gyration shows that at both 1 M and 5 M urea, the R_g is slightly smaller in the simulations with labeled protein than with unlabeled, although at 5 M the difference is well within the statistical error bars (Figure 5(a)). It is clear, however, that R_g increases with denaturant concentration, both for the labeled and unlabeled systems. In a previous study[109], the protein R_g was shown to be insensitive to whether the protein was labeled or not. However that study used a force field (Amber ff03w[110]) in which the unfolded structure was already somewhat collapsed. Here we obtain the same conclusion, although using an improved force field which reproduces the correct dimensions of the unfolded configurations, and we still find little effect of the labels, strengthening the earlier conclusion.

The results of a simple average FRET calculation using the distance between the chromophores directly rather than an approximate distance based on the separation of C_α atoms are included in Figure 5(b) (details in Methods section), showing very similar results. A second assumption in interpreting the FRET data is that the chromophores reorient rapidly on the time scale of the donor lifetime, so that only an average effect of the relative chromophore orientation factor, κ^2, needs to be considered, i.e. $\langle \kappa^2 \rangle = 2/3$. Since, in the simulations with explicit chromophores, we have a complete, unbiased, trajectory of the chromophore positions, we can calculate directly the time-dependent rate coefficient for resonance energy transfer, the decay of the donor fluorescence intensity and consequently the FRET efficiency. Thus, the only remaining assumptions we make are those included in Förster's original theory (e.g. that the transition densities can be approximated as point dipoles). The donor fluorescence decay is shown in Figure S12, and the transfer efficiencies in Figure 5(b). The consistency of the different calculations provides strong support both for the simple distance-based FRET estimate as well as for the assumption of $\langle \kappa^2 \rangle = 2/3$. Indeed, the equilibrium average κ^2 computed from the simulations is very close to the expected value of 2/3 for an isotropic distribution of chromophore orientations (Figure 5(d)), as seen in an earlier study[109]. The reason for the validity of this assumption is that at least one of the chromophores in each case is reorienting rapidly on the scale of the donor lifetime, with rotational correlation times of ~1 ns, with a similar correlation time for κ^2 itself (Figure S12), compared with donor lifetimes of ~2 ns for molecules labeled with both donor and acceptor under the denaturant conditions used. Even though the correlation time for Alexa 488 reorientation is substantially longer than this in the 1 M urea simulation due to formation of stable contacts with the protein, the free rotation of the Alexa 594 ensures a short correlation time for the overall κ^2. The differences between Alexa 488 and Alexa 594 may relate to differences between the chromophores themselves, to the labeling position (N or C terminal), and to the limited sampling in the simulation – these effects would have to be investigated in future work. Although the average κ^2 at 1 M is slightly less than 2/3, which would tend to increase the apparent efficiency, the consistency of the full calculation

including the relative orientation of the dyes with that based only on distance (Figure 5(b)) indicates that most of the variation in efficiency with denaturant concentration comes from changes in the distance distribution.

A second issue in interpreting FRET experiments is that the distance probed by FRET is that between the chromophores, which are usually attached to the protein by long flexible linkers in order to allow the chromophores to reorient freely. Thus a transformation needs to be made to convert the mean square distance between the chromophores to a distance between protein residues, which is the quantity of interest. One procedure for doing this is to rescale the observed distance R_{obs} by assuming that the linkers effectively add a certain number of extra residues to the length of the chain, so that the distance between protein residues is $R = (N/(N + N_{linker}))^{\nu} R_{obs}$ where the number of extra residues N_{linker} has been chosen to be around 9 from the literature[35] and ν is the polymer scaling exponent (from Table 1). Since in the simulations with attached chromophores, we can measure both distances, we determine N_{linker} by minimizing the difference between the average FRET efficiency computed using the distance between chromophores, and that computed using the distance between residues with the N_{linker}-dependent correction. The χ^2 between these two estimates is shown in Figure 5(c), yielding a minimum at $N_{linker} \approx 10$ residues, very close to the value of 9 estimated from experiment.

SAXS calculations using explicit and implicit solvent models

In the above analysis, we have computed SAXS scattering intensities using an all-atom representation, including all solvent molecules[65, 111]. This is the gold standard, and could be important if there were significant solvent structure around the protein which could even affect the measured radius of gyration if, e.g., the solvent specifically partitioned toward the center of the coil rather than being uniformly distributed along its length. Whether such solvent structuring is significant can be elucidated via a straightforward test: comparison of the scattering curves from the all-atom calculations with those from an implicit uniform model for the surface solvent. Using the same protein configurations as for the atomistic SAXS calculation, we have computed scattering profiles using the programs CRYSOL[41] and FOXS[66].

Note that CRYSOL includes parameters describing the average thickness and background contrast of the solvation layer around the protein which are optimized for folded proteins in water to 0.3 nm and ~10% of the bulk density, respectively, while FOXS is also optimized for aqueous solvent. For non-compact unfolded conformations, errors arising from this assumption are only expected to affect $I(q)$ at larger scattering angles, provided that the solvation layer is strongly correlated with the chain locus. In Figure 6 we show the comparison between the explicit solvent calculation of $I(q)$ and a CRYSOL calculation, in which the background electron density is taken from the all-atom simulations at each urea concentration, and the default hydration shell parameter is used. As is evident, the continuum approximation is very good for $q < 0.3$ Å^{-1}, and excellent for $q < 0.04$ Å^{-1}, which includes the Guinier region used to determine the R_g in experiments. The quality of this agreement is not very sensitive to the solvent model used: we have tried alternative procedures of not adjusting the background electron density (Table S5, Figure S13), and of

leaving out the hydration shell altogether (Figure S14). Although the latter of course leads to larger deviations for $q > 0.08$ Å^{-1}, the difference of SAXS intensity from the explicit solvent calculation is still within 3% in the Guinier region. We have also computed scattering intensities using a different program, FOXS[66]. in which the surface solvent is modeled by adjusting the atomic form factor of the solvent-exposed atoms. With default parameters originally optimized for folded proteins in water, we again obtain a good agreement with the SAXS intensity computed from explicit solvent calculations in Guinier region (Figure S15, Table S5). All of these results suggest that a precise description of the hydration shell is not necessary to estimate the R_g of unfolded proteins and that there is no cooperative solvent structuring around the protein, beyond the first solvation layer. As a final verification of this point, we show in Figure 1 the R_g estimated from Guinier fits to the $I(q)$ in Figure 6, demonstrating that the result is almost identical to that obtained using explicit solvent coordinates. Guinier fits to the other scattering calculations with implicit solvent also yield the same results, after considering the statistical error due to finite sampling in MD. Therefore, the R_g inferred from the Guinier fit accurately reflects the expansion of the chain as urea concentration increases (Figure 1(c)). To provide some intuitive understanding of this observation, we have calculated the average number of urea molecules within 4.5 Å of each residue, showing that the distribution of urea within an approximate first solvation shell is uniform along the sequence (Figure S16). These results effectively rule out the possibility that effects such as preferential partitioning of the solvent towards the centre of the coil could distort the R_g inferred from SAXS.

Conclusions

We have used unbiased microsecond atomistic simulations with a force field carefully calibrated against small-molecule solubility data to investigate the effect of denaturants on an intrinsically disordered protein. We find that increasing only the denaturant concentration causes an increase of radius of gyration, end-to-end distance, and polymer scaling exponent. We further show that the molecular origin of the expansion is preferential association of denaturant molecules with the chain. Careful analysis of the interactions between the protein and urea yields transfer free energies from water into denaturant solution in good accord with experiment. With the new force field we achieve a good match with experimental transfer free energies, as well as with the SAXS and FRET data for ACTR, which is essential for a quantitative understanding of the underlying mechanism, We find that almost all residues have a favorable transfer free energy from water to 1 M urea, the only exceptions being the small anionic residues Asp and Glu, for which water is a better solvent. A more detailed breakdown indicates that the backbone and side-chains make similar contributions to the overall transfer free energy. Interactions with the backbone appear to be dominated by hydrogen bonding, whilst other types of interaction (e.g. hydrophobic interactions) are clearly also important for side-chains. The small amount of helical secondary structure present is progressively lost with increasing denaturant concentration.

The observed chain expansion is validated by comparison with experimental FRET efficiencies and SAXS scattering intensities, where quantitative agreement is obtained – emerging only from the basic intermolecular interactions captured by the force field. Thus, at least for the intrinsically disordered protein ACTR, which we study here, all of the

experimental data is consistent with a scenario in which the protein expands, and with the current understanding of denaturation mechanism, mediated by protein-denaturant binding. We have investigated potential molecular-scale artefacts which have been suggested to explain the discrepancies between experiments. First, we verify the accuracy of assuming that solvent distribution has little impact on the radius of gyration of the protein obtained from SAXS. Second, for FRET we show that the chain collapse is not induced by the FRET labels. Since there is no fundamental inconsistency between the two experiments, and in the absence of the above artefacts, the experimental discrepancy most likely relates to the challenging inverse problem of determining properties of IDPs from limited experimental data.[70]

Overall, our results highlight the potential of unbiased atomistic simulations for providing a molecular interpretation for complex experimental data. The good agreement between our simulation results and the properties of ACTR in both water in and denaturant suggests that the force fields used are reaching the point of being a useful tool for the investigation of intrinsically disordered proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Jürgen Köfinger and Gerhard Hummer for providing their software package to perform the all-atom SAXS intensity calculation. R.B. and W.Z. were supported by the intramural research program of the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health. This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov)

References

1. Nozaki Y, Tanford C. The solubility of amino acids and related compounds in aqueous urea solutions. J. Biol. Chem. 1963; 238:4074–4081. [PubMed: 14086747]
2. Frank HS, Franks F. Structural approach to the solvent power of water for hydrocarbons; urea as a structure breaker. J. Chem. Phys. 1968; 48:4746–4757.
3. Schellman JA. Solvent Denaturation. Biopolymers. 1978; 17(5):1305–1322.
4. Camilloni C, Rocco AG, Eberini I, Gianazza E, Broglia RA, Tiana G. Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. Biophys. J. 2008; 94:4654–4661. [PubMed: 18339753]
5. Canchi DR, Paschek D, García AE. Equilibrium study of protein denaturation by urea. J. Am. Chem. Soc. 2010; 132:2338–2344. [PubMed: 20121105]
6. Horinek D, Netz RR. Can simulations quantitatively predict peptide transfer free energies to urea solutions? Thermodynamic concepts and force field limitations. J. Phys. Chem. A. 2011; 115:6125–6136. [PubMed: 21361327]
7. Moeser B, Horinek D. Unified description of urea denaturation: backbone and side chains contribute equally in the transfer model. J. Phys. Chem. B. 2014; 118:107–114. [PubMed: 24328141]
8. Stumpe MC, Grubmüller H. Interaction of urea with amino acids: implications for urea-induced protein denaturation. J. Am. Chem. Soc. 2007; 129:16126–16131. [PubMed: 18047342]
9. Guinn EJ, Pegram LM, Capp MW, Pollock MN, Record MT. Quantifying why urea is a protein denaturant whereas glycine betaine is a protein stabilizer. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:16932–16937. [PubMed: 21930943]

10. Holthauzen LMF, Rosgen J, Bolen DW. Hydrogen bonding progressively strengthens upon transfer of the protein urea-denatured state to water and protecting osmolytes. Biochemistry. 2010; 49:1310–1318. [PubMed: 20073511]
11. De Gennes, P-G. Scaling Concepts in Polymer Physics. Cornell University Press; Ithaca and London: 1979.
12. Flory, PJ. Principles of Polymer Chemistry. Cornell University Press; Ithaca and London: 1953.
13. Haran G. How, when and why proteins collapse: the relation to folding. Curr. Opin. Struct. Biol. 2012; 22:14–20. [PubMed: 22104965]
14. Schuler B, Lipman EA, Eaton WA. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. Nature. 2002; 419:743–747. [PubMed: 12384704]
15. Sherman E, Haran G. Coil-globule transition in the denatured state of a small protein. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:11539–11543. [PubMed: 16857738]
16. Merchant KA, Best RB, Louis JM, Gopich IV, Eaton WA. Characterizing the unfolded states of proteins using single molecule FRET spectroscopy and molecular simulations. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:1528–1533. [PubMed: 17251351]
17. Hoffmann A, Kane A, Nettels D, Hertzog DE, Baumgärtel P, Lengefeld J, Reichardt G, Horsley DA, Seckler R, Bakajin O, Schuler B. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:105–110. [PubMed: 17185422]
18. Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, Chemla DS, Dawson PE, Shultz PG, Weiss S. Single-molecule protein folding: diffusion fluorescence energy transfer studies of the denaturation of chymotrypsin inhibitor 2. Proc. Natl. Acad. Sci. U. S. A. 2000; 97(10):5179–5184. [PubMed: 10792044]
19. Kuzmenkina EV, Heyes CD, Nienhaus GU. Single molecule FRET study of denaturant induced unfolding of RNase H. J. Mol. Biol. 2006; 357:313–324. [PubMed: 16426636]
20. Huang F, Sato S, Sharpe TD, Ying L, Fersht AR. Distinguishing between cooperative and unimodal downhill protein folding. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:123–127. [PubMed: 17200301]
21. Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:2649–2654. [PubMed: 17299036]
22. Möglich A, Joder K, Kiefhaber T. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:12394–12399. [PubMed: 16894178]
23. Nöppert A, Gast K, Müller-Frohne M, Zirwer D, Damaschun G. Reduced-denatured ribonuclease A is not in a compact state. FEBS Lett. 1996; 380:179–182. [PubMed: 8603733]
24. Mok Y-K, Kay CM, Kay LE, Forman-Kay J. NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. J. Mol. Biol. 1999; 289:619–638. [PubMed: 10356333]
25. Choy W-Y, Mulder FAA, Crowhurst KA, Muhandiram DR, Millet IS, Doniach S, Forman-Kay JD, Kay LE. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J. Mol. Biol. 2002; 316:101–112. [PubMed: 11829506]
26. Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson CM, Poulsen FM, Vendruscolo M. Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. J. Am. Chem. Soc. 2004; 126:3291–3299. [PubMed: 15012160]
27. Konuma T, Kimura T, Matsumoto S, Goto Y, Fujisawa T, Fersht AR, Takahashi S. Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase. J. Mol. Biol. 2011; 405:1284–1294. [PubMed: 21146541]
28. Plaxco KW, Millet IS, Segal DJ, Doniach S, Baker D. Chain collapse can occur concomitantly with the rate-limiting step in protein folding. Nat. Struct. Biol. 1999; 6:554–556. [PubMed: 10360359]
29. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR. Early collapse is not an obligate step in protein folding. J. Mol. Biol. 2004; 338:369–382. [PubMed: 15066438]

30. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR. Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. J. Mol. Biol. 2007; 367:609–615. [PubMed: 17292402]
31. Yoo TY, Meisberger SP, Hinshaw J, Pollack L, Haran G, Sosnick TR, Plaxco K. Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-denaturant unfolded state. J. Mol. Biol. 2012; 418:226–236. [PubMed: 22306460]
32. Watkins HM, Simon AJ, Sosnick TR, Lipman EA, Hjelm RP, Plaxco KW. Random coil negative control reproduces the discrepancy between scattering and FRET measurements of denatured protein dimensions. Proc. Natl. Acad. Sci. U. S. A. 2015; 112:6631–6636. [PubMed: 25964362]
33. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr. Opin. Struct. Biol. 2015; 32:102–112. [PubMed: 25863585]
34. Müller-Späth S, Soranno A, Hirschfeld V, Hofmann H, Rüegger S, Reymond L, Nettels D, Schuler B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:14609–14614. [PubMed: 20639465]
35. Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:16155–16160. [PubMed: 22984159]
36. Johansen D, Trewhella J, Goldenberg DP. Fractal dimension of an intrinsically disordered protein: small-angle X-ray scattering and computational study of the bacteriophage lambda N protein. Protein Sci. 2011; 20:1955–1970. [PubMed: 21936008]
37. Takahashi S, Kamagata K, Oikawa H. Where the complex things are: single molecule and ensemble spectroscopic investigations of protein folding dynamics. Curr Opin Struc Biol. 2016; 36:1–9.
38. Gianni S, Jemth P. Protein folding: Vexing debates on a fundamental problem. Biophys Chem. 2016; 212:17–21. [PubMed: 27018826]
39. Song J, Gomes G-N, Gradinaru CC, Chan H-S. An adequate account of excluded volume is necessary to infer compactness and asphericity of disordered proteins by Förster resonance energy transfer. J. Phys. Chem. B. 2015; 119:15191–15202. [PubMed: 26566073]
40. O'Brien E, Morrison G, Brooks BR, Thirumalai D. How accurate are polymer models in the analysis of Förster resonance energy transfer experiments on proteins? J. Chem. Phys. 2009; 130:124903. [PubMed: 19334885]
41. Svergun D, Barberato C, Koch MHJ. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Cryst. 1995; 28:768–773.
42. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. J. Am. Chem. Soc. 2012; 134(8): 3787–3791. [PubMed: 22339051]
43. Petrescu A-J, Receveur V, Calmettes P, Durand D, Smith JC. Excluded volume in the configurational distribution of a strongly-denatured protein. Protein Sci. 1998; 7:1396–1403. [PubMed: 9655344]
44. Zheng W, Borgia A, Borgia MB, Schuler B, Best RB. Empirical optimization of interactions between proteins and chemical denaturants in molecular simulations. J. Chem. Theor. Comput. 2015; 11:5543–5553.
45. Liu Z, Reddy G, Thirumalai D. Folding PDZ2 domain using the molecular transfer model. J. Phys. Chem. B. 2016; doi: 10.1021/acs.jpcb.6b00327
46. Bennion BJ, Daggett V. The molecular basis for the chemical denaturation of proteins by urea. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:5142–5147. [PubMed: 12702764]
47. Holehouse AS, Garai K, Lyle N, Vitalis A, Pappu RV. Quantitative assessments of the distinct contributions of polypeptide backbone amides versus side chain groups to chain expansion via chemical denaturation. J. Am. Chem. Soc. 2015; 137:2984–2995. [PubMed: 25664638]
48. Wallqvist A, Covell DG, Thirumalai D. Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation. J. Am. Chem. Soc. 1998; 120:427–428.
49. Hua L, Zhou R, Thirumalai D, Berne BJ. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. Proc. Natl. Acad. Sci. U. S. A. 2008; 105:16928–16933. [PubMed: 18957546]

50. O'Brien EP, Dima R, Brooks BR, Thirumalai D. Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. J. Am. Chem. Soc. 2007; 129:7346–7353. [PubMed: 17503819]
51. Yang Z, Xiu P, Shi B, Hua L, Zhou R. Coherent microscopic picture for urea-induced denaturation of proteins. J. Phys. Chem. B. 2012; 116:8856–8862. [PubMed: 22780326]
52. Tirado-Rives J, Orozco M, Jorgensen WL. Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. Biochemistry. 1997; 36:7313–7329. [PubMed: 9200680]
53. Caflisch A, Karplus M. Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. Proc. Natl. Acad. Sci. U. S. A. 1994; 91:1746–1750. [PubMed: 8127876]
54. Shukla D, Shinde C, Trout BL. Molecular computations of preferential interaction coefficients of proteins. J. Phys. Chem. B. 2009; 113:12546–12554. [PubMed: 19697945]
55. Best RB, Zheng W, Mittal J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. J. Chem. Theor. Comput. 2014; 10:5113–5124.
56. 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]
57. 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. U. S. A. 2014; 111:4874–4879. [PubMed: 24639500]
58. Hess B, Kutzner C, Van der Spoel D, Lindahl E. GROMACS4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theor. Comput. 2008; 4(3):435–447.
59. Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. 1981; 52(12):7182–7190.
60. Darden T, York D, Pedersen L. An N-log(N) method for Ewald sums in large systems. J. Chem. Phys. 1993; 103:8577–8592.
61. Abascal JLF, Vega C. A general purpose model for the condensed phases of water: TIP4P/2005. J. Chem. Phys. 2005; 123:234505. [PubMed: 16392929]
62. Weerasinghe S, Smith PE. A Kirkwood-Buff derived force field for mixtures of urea and water. J. Phys. Chem. B. 2003; 107:3891–3898.
63. Weerasinghe S, Smith PE. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. J Chem Phys. 2004; 121(5):2180–2186. [PubMed: 15260772]
64. Best RB, Hofmann H, Nettels D, Schuler B. Quantitative interpretation of FRET experiments via molecular simulation: force field and validation. Biophys. J. 2015; 108:2721–2731. [PubMed: 26039173]
65. Köfinger J, Hummer G. Atomic-resolution structural information from scattering experiments on macromolecules in solution. Phys. Rev. E. 2013; 87:052712.
66. Schneidman-Duhovny D, Hammel M, Tainer JA, Sali A. Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. Biophys J. 2013; 105(4):962–974. [PubMed: 23972848]
67. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA. Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:2754–2759. [PubMed: 15699337]
68. Soranno A, Buchli B, Nettels D, Cheng RR, Müller-Späth S, Pfeil SH, Hoffmann A, Lipman EA, Makarov DE, Schuler B. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. Proc. Natl. Acad. Sci. U. S. A. 2012; 109:17800–17806. [PubMed: 22492978]
69. Förster T. Zwischenmolekulare Energiewanderung und Fluoreszenz. Ann. Phys. 1948; 6:55–75.
70. Borgia A, Zheng W, Buholzer K, Borgia MB, Schuler A, Hofmann H, Soranno A, Nettels D, Gast K, Grishaev A, Best RB, Schuler B. Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. Submitted. 2016
71. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc. Natl. Acad. Sci. U. S. A. 2010; 107:8183–8188. [PubMed: 20404210]

72. Hoefling M, Lima N, Haenni D, Seidel CAM, Schuler B, Grubmuller H. Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. PLoS One. 2011; 6(5):e19791. [PubMed: 21629703]
73. Hoefling M, Grubmüller H. In silico FRET from simulated dye dynamics. Comput. Phys. Comm. 2013; 184:841–852.
74. Graen T, Hoefling M, Grubmüller H. AMBER-DYES: characterization of charge fluctuations and force field parameterization of fluorescent dyes for molecular dynamics simulations. J. Chem. Theor. Comput. 2014; 10:5505–5512.
75. Nettels D, Gopich IV, Hoffmann A, Schuler B. Ultrafast dynamics of protein collapse from single molecule photon statistics. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:2655–2660. [PubMed: 17301233]
76. Nettels D, Hoffmann A, Schuler B. Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. J. Phys. Chem. B. 2008; 112:6137–6146. [PubMed: 18410159]
77. Borgia A, Wensley BG, Soranno A, Nettels D, Borgia MB, Hoffmann A, Pfeil SH, Lipman EA, Clarke J, Schuler B. Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. Nature Comm. 2012; 3:1195.
78. Nettels D, Muller-Spath S, Küster F, Hofmann H, Haenni D, Rüegger S, Reymond L, Hoffmann A, Kubelka J, Heinz B, Gast K, Best RB, Schuler B. Single molecule spectroscopy of the temperature-induced collapse of unfolded proteins. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:20740–20745. [PubMed: 19933333]
79. Piana S, Klepeis JL, Shaw DE. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr. Opin. Struct. Biol. 2014; 24:98–105. [PubMed: 24463371]
80. Nerenberg PS, Jo B, Tripathy A, Head-Gordon T. Optimizing solute-water van der Waals Interactions to Reproduce Solvation Free Energies. J. Phys. Chem. B. 2012; 116:4524–4534. [PubMed: 22443635]
81. 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]
82. Wang Y, Trewhella J, Goldenberg DP. Small-angle X-ray scattering of reduced ribonuclease A: effects of solution conditions and comparisons with a computational model of unfolded proteins. J. Mol. Biol. 2008; 377:1576–1592. [PubMed: 18329044]
83. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco aKW. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:12491–12496. [PubMed: 15314214]
84. Chahine J, H N, Leite VBP, Socci ND, Onuchic JN. Specific and nonspecific collapse in protein folding funnels. Phys. Rev. Lett. 2002; 88:168101. [PubMed: 11955268]
85. Camacho CJ, Thirumalai D. Kinetics and thermodynamics of folding in model proteins. Proc. Natl. Acad. Sci. U. S. A. 1993; 90:6369–6372. [PubMed: 8327519]
86. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc. Natl. Acad. Sci. U. S. A. 2000; 97:8868–8873. [PubMed: 10908673]
87. De Sancho D, Best RB. Modulation of an IDP binding mechanism and rates by helix propensity and non-native interactions: association of HIF1a with CBP. Molecular Biosystems. 2012; 8:256–267. [PubMed: 21892446]
88. Huang, J-r, Gabel, F., Jensen, MR., Grzesiek, S., Blackledge, M. Sequence-specific mapping of the interaction between urea and unfolded ubiquitin from ensemble analysis of NMR and small angle scattering data. J. Am. Chem. Soc. 2012; 134:4429–4436. [PubMed: 22309138]
89. Timasheff SN. Water as a ligand: preferential binding and exclusion of denaturants in protein folding. Biochemistry. 1992; 31:9857–9864. [PubMed: 1390769]
90. Baynes BM, Trout BL. Proteins in mixed solvents: a molecular-level perspective. J. Phys. Chem. B. 2003; 107:14058–14067.

91. Ma L, Pegram L, Record MT, Cui Q. Preferential interactions between small solutes and the protein backbone: a computational analysis. Biochemistry. 2010; 49:1954–1962. [PubMed: 20121154]
92. Canchi DR, García AE. Backbone and side-chain contributions in protein denaturation by urea. Biophys. J. 2011; 100:1526–1533. [PubMed: 21402035]
93. Voronoi GF. Nouvelles applications des paramètres continus a là théorie de formes quadratiques. J. Reine Angew. Math. 1908; 134:198–287.
94. Naidoo KJ, Kuttel M. Water structure about the dimer and hexamer repeat units of amylose from molecular dynamics computer simulations. J. Comput. Chem. 2001; 22:445–456.
95. Auton M, Bolen DW. Predicting the energetics of osmolyte-induced protein folding/unfolding. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:15065–15068. [PubMed: 16214887]
96. König G, Boresch S. Hydration free energies of amino acids: why side chain analog data are not enough. J. Phys. Chem. B. 2009; 113:8967–8974. [PubMed: 19507836]
97. Ziv G, Haran G. Protein folding, protein collapse and Tanford's transfer model: lessons from single-molecule FRET. J. Am. Chem. Soc. 2009; 131:2942–2947. [PubMed: 19239269]
98. Bolen WD, Rose GD. Structure and energetics of the hydrogen-bonded backbone in protein folding. Annu. Rev. Biochem. 2008; 77:339–362. [PubMed: 18518824]
99. Stokes RH. Thermodynamics of aqueous urea solutions. Austr. J. Chem. 1967; 20:2087–2100.
100. Nozaki Y, Tanford C. The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. J. Biol. Chem. 1970; 245:1648–1652. [PubMed: 5438355]
101. Lim WK, Rosgen J, Englander SW. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. P Natl Acad Sci USA. 2009; 106(8):2595–2600.
102. Perrin CL, Lollo CP, Johnston ER. NMR site-to-site rate constants and the mechanisms of acid-catalyzed proton exchange in secondary amides. J. Am. Chem. Soc. 1984; 106:2749–2753.
103. Canchi DR, García AE. Cosolvent effects on protein stability. Annu. Rev. Phys. Chem. 2013; 64:273–293. [PubMed: 23298246]
104. Kjaergaard M, Nørholm A-B, 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]
105. Bernadó P, Blackledge M. A self-consistent description of the conformational behaviour of chemically denatured proteins from NMR and small angle scatter. Biophys. J. 2009; 97(10): 2839–2845. [PubMed: 19917239]
106. Wuttke R, Hofmann H, Nettels D, Borgia MB, Mittal J, Best RB, Schuler B. Temperature-dependent solvation modulates the dimensions of disordered proteins. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:5213–5218. [PubMed: 24706910]
107. Zerze GH, Best RB, Mittal J. Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. J Phys Chem B. 2015; 119(46):14622–14630. [PubMed: 26498157]
108. Lakowicz, JR. Principles of Fluorescence Spectroscopy. Springer; New York: 2006.
109. Zerze GH, Best RB, Mittal J. Modest influence of FRET chromophores on the properties of unfolded proteins. Biophys. J. 2014; 107:1654–1660. [PubMed: 25296318]
110. Best RB, Mittal J. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. J Phys Chem B. 2010; 114(46): 14916–14923. [PubMed: 21038907]
111. Lindner B, Smith JC. Sassena - X-ray and neutron scattering calculated from molecular dynamics trajectories using massively parallel computers. Comput. Phys. Comm. 2012; 183:1491–1501.

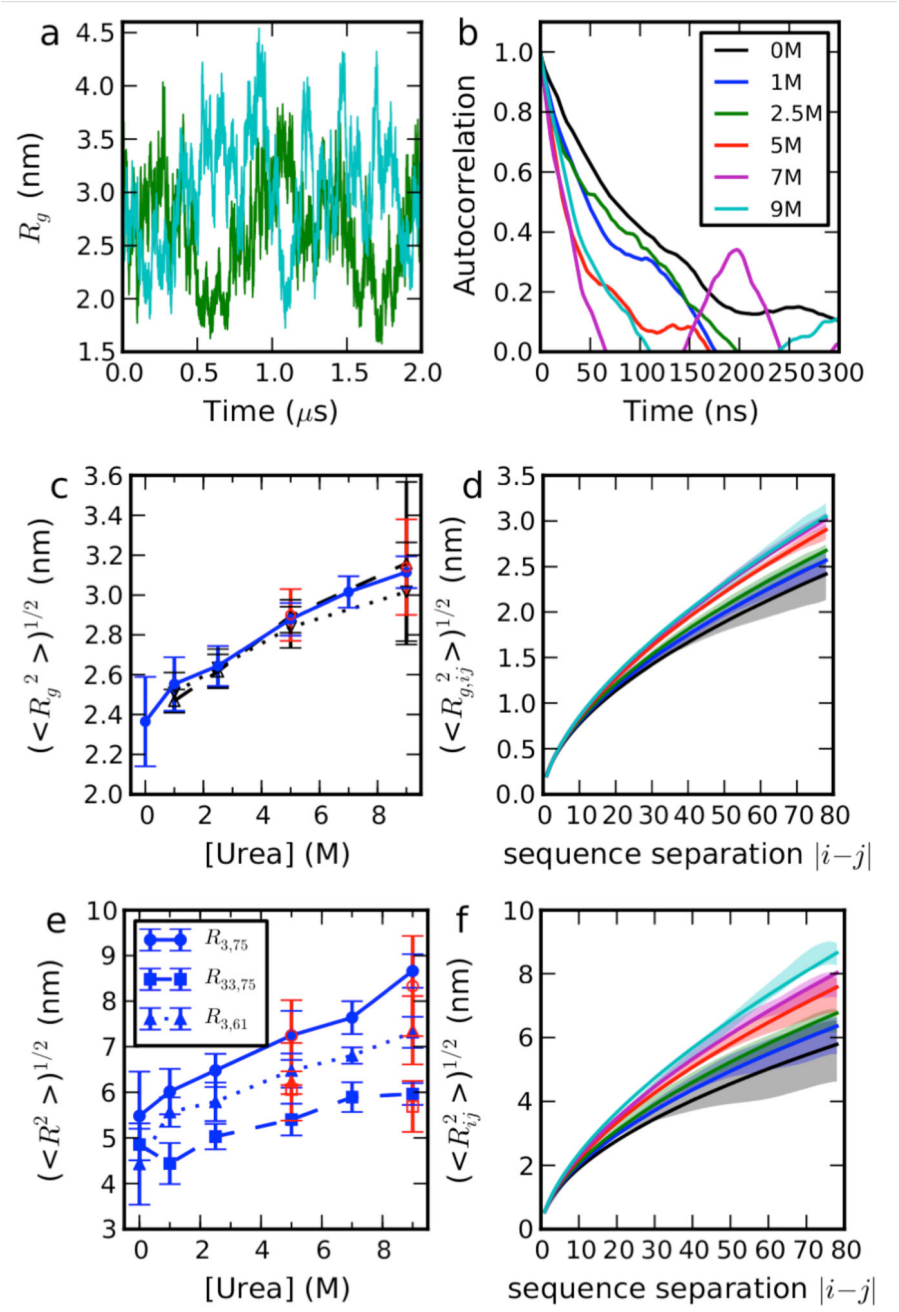


Figure 1.

Equilibrium properties of ACTR in Urea. (a) Fluctuations in R_g over time at selected denaturant concentrations, concentration as labeled in (b). (b) Time correlation functions for R_g. (c) Dependence of mean R_g from simulations (blue solid line, filled circles), Guinier fit of SAXS intensity with $q < 0.04$ Å^{-1} from explicit solvent calculation (black dash line, up triangles) and from implicit solvent calculation (black dotted line, down triangles). Red symbols show the data from large box simulations (see Table S1). (d) the dependence of the R_g included between residues i and j on the sequence separation $|i - j|$. (e) Dependence of root mean square intramolecular distances on denaturant concentration. Red symbols show

data from large box simulations (see Table S1). (f) Dependence of root-mean-square distance between *i* and *j* on $|i - j|$. Scaling exponents were calculated by fitting a power law to the dependence on sequence separation $|i - j|$ of either the root mean square distance between *i* and *j* in (f) or the R_g of the chain included between residues *i* and *j*. Solid lines show fits to $<R_{\mathrm{ij}}^2>^{1/2} = \sqrt{2l_p b}N^{\nu}$ and $<R_{g,\mathrm{ij}}^2>^{1/2} \approx \sqrt{\frac{2l_p b}{(2\nu+1)(2\nu+2)}}N^{\nu}$, in which l_p=0.4 nm and b = 0.38 nm[35] (the expression for R_g is an approximation which assumes Flory scaling to hold for all *i,j* pairs, although it is only strictly valid for sufficiently large $|i - j|$). Simulation error bars are obtained from block averages, as described in the SI text, and refer to the standard error of the mean.

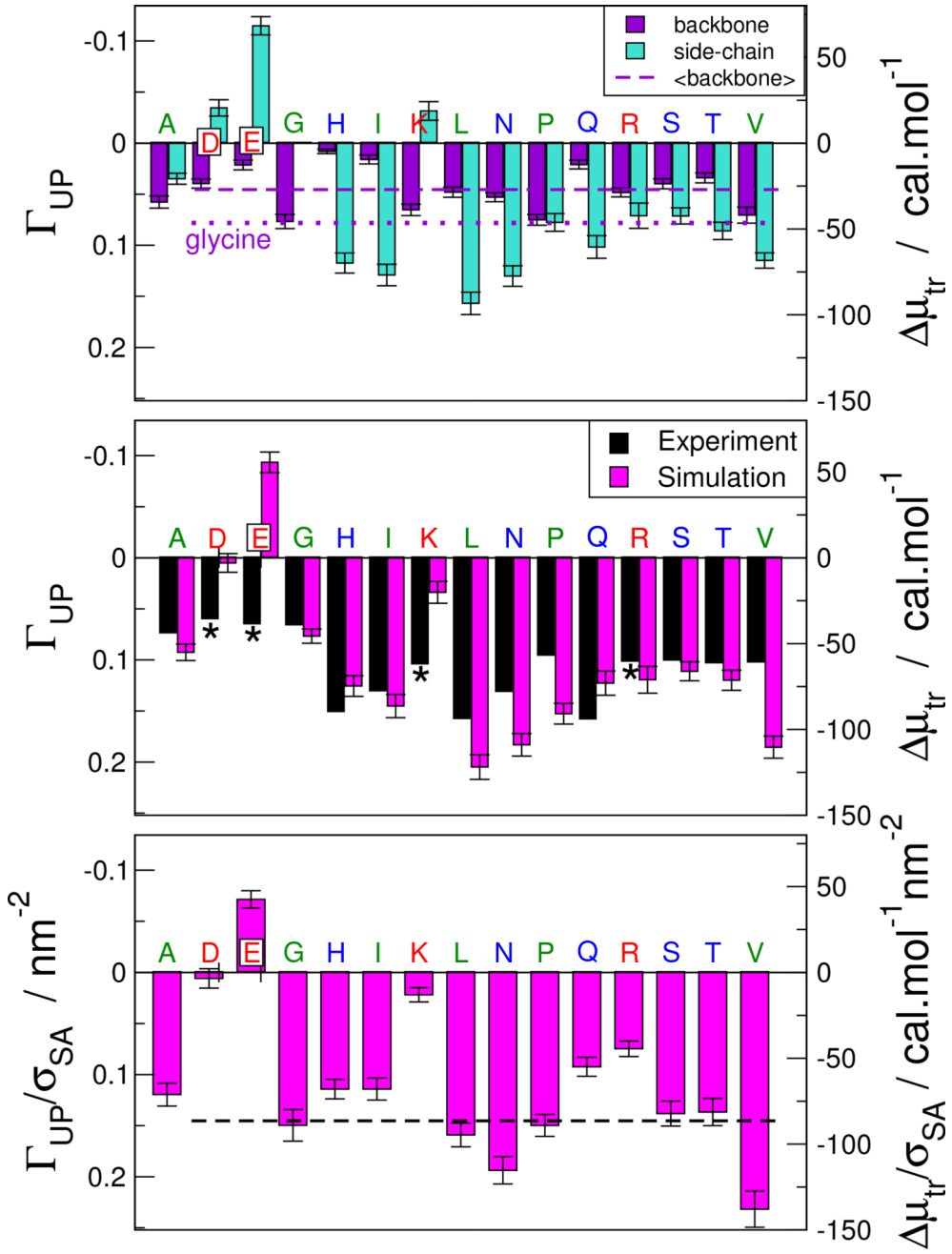


Figure 2.

Preferential interaction coefficients Γ_{UP} (left axis) for association of urea with the protein surface at 1 M urea. (Top) Decomposition of Γ_{UP} for each type of residue into backbone and side-chain contributions. Broken line is the average backbone contribution across all residues. Dotted line is Γ_{UP} for glycine. (Centre) Comparison of Γ_{UP} for whole residues with experimental water to urea transfer free energies.[95] Transfer free energies (right axis) were approximated from Γ_{UP} using $\Delta\mu_{tr} \approx -RT\Gamma_{UP}$, where R and T are the molar gas constant and absolute temperature, respectively.[90] Transfer free energies for charged residues (*) are not directly comparable, because they necessarily include counterions.

(Bottom) Preferential interaction coefficients Γ_{UP} and transfer free energies $\Delta\mu_{tr}$ normalized by the average surface area of each residue, σ_{SA}. Broken line is the average Γ_{UP}/σ_{SA} over all charge neutral residues. Residue labels are colored according to residue type: hydrophobic: green, polar: blue, charged: red. Simulation error bars refer to the standard error of the mean, computed from block averages.

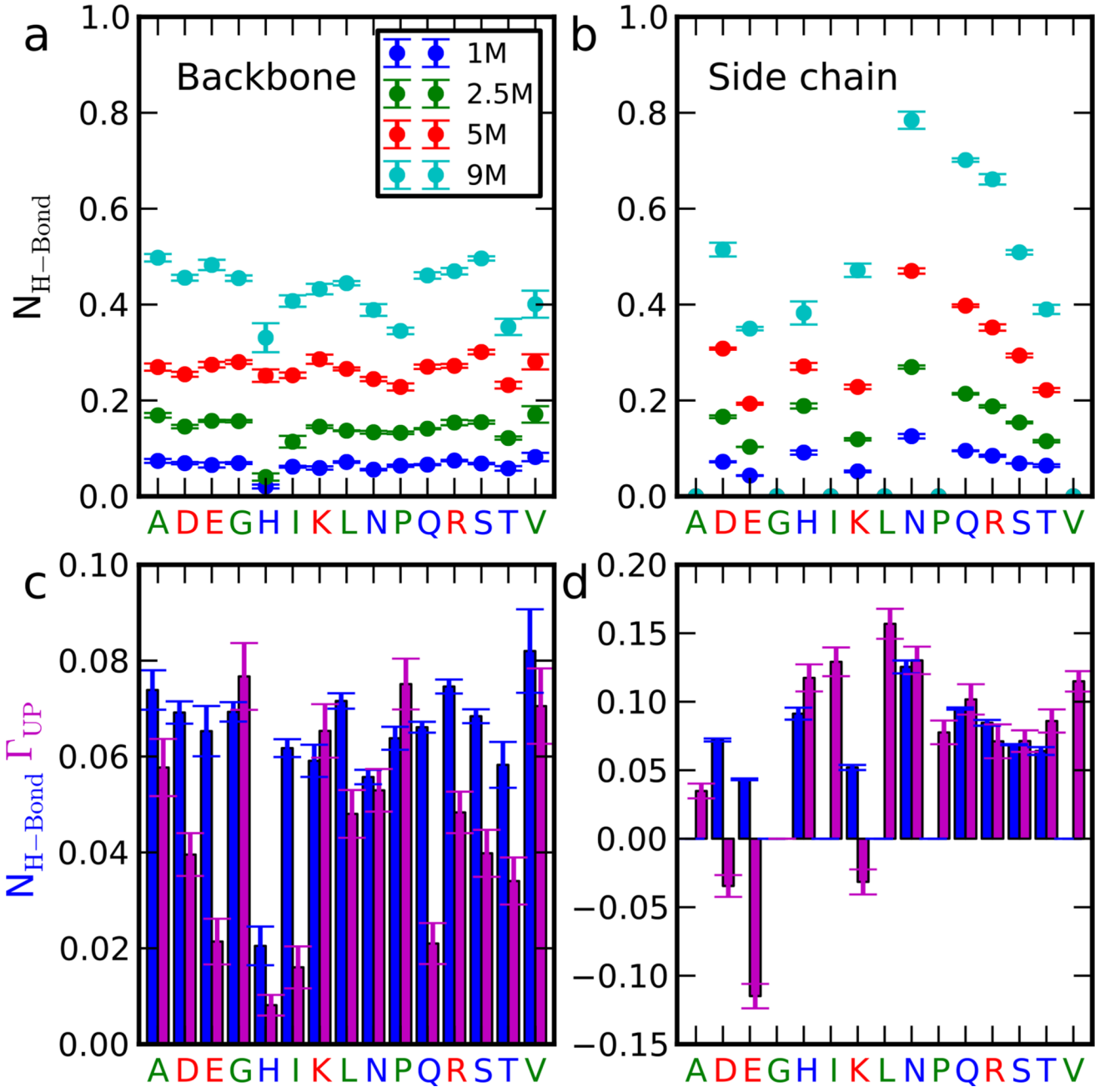


Figure 3.

Hydrogen bonding between urea and protein. In (a) and (b) are shown the number of hydrogen bonds per urea molecule to the backbone and side-chains respectively, averaged by residue type, at different urea concentrations. In (c) and (d) we compare the average number of hydrogen bonds with the preferential interaction coefficients, a measure of the number of excess urea molecules in the vicinity of each group, relative to bulk. Error bars refer to the standard error of the mean, computed from block averages.

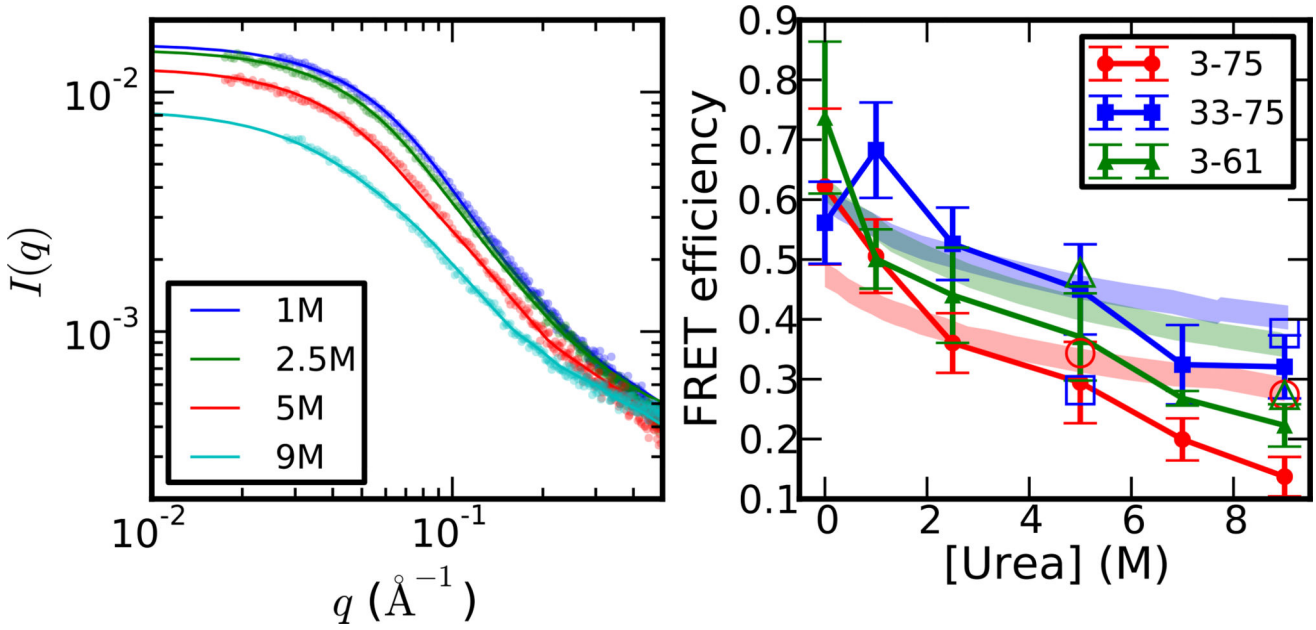


Figure 4.

Comparison with experimental observables. Left: SAXS; Right: FRET. Experimental data (from Ref[70]) and uncertainties are represented by shaded areas and simulation data by solid lines. Solid symbols show the results of simulations with a common box size (12 nm rhombic dodecahedron), and open symbols the results using larger box sizes (15 nm and 17 nm at 5 M and 7 M, respectively). Large box simulations were run for 0.6 μs versus 2 μs for the small box simulations. For calculating FRET efficiencies, the distance between the C_α atoms of the labeled residues has been rescaled by a factor $[\frac{N+9}{N}]^{\nu}$ as detailed in the Methods section. For FRET, simulation error bars give the standard error of the mean and the shaded regions account for the systematic error in experiment.

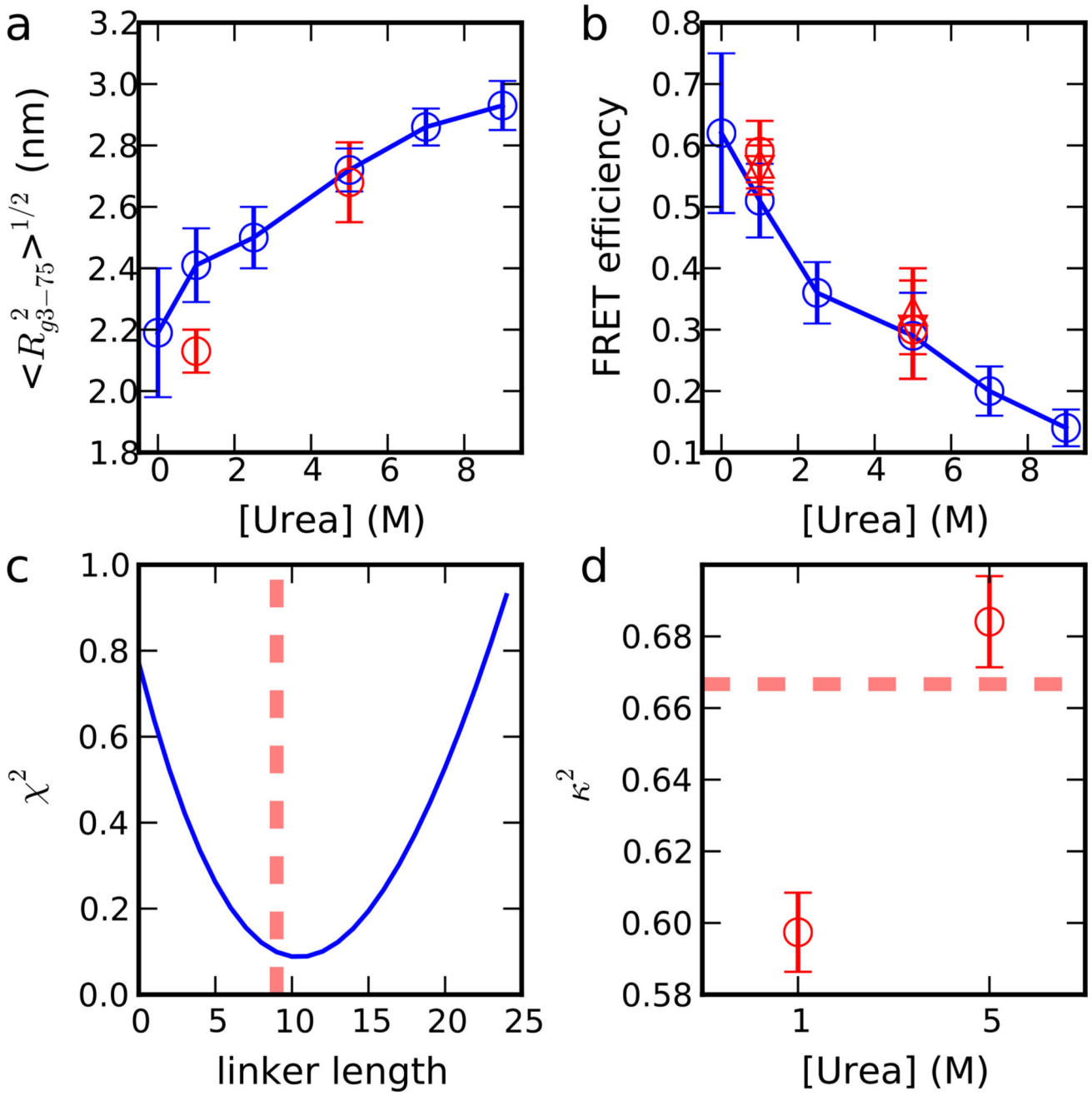


Figure 5.

Explicit and implicit treatment of dyes in FRET calculation. (a) R_g of residue 3–75 fragment for simulation with dyes (red) and without dyes (blue). (b) FRET efficiency for simulation with dyes (red) and without dyes (blue). For simulations with dyes, FRET efficiency is estimated from three different ways described in the Methods section, including C_α distance with a correction of linker length of 9 residues (circle), C_1 distance without the correction (lower triangle), and integration of the intensity decay directly. (c) Variation of χ^2 between the two ways of calculating FRET efficiency described in (b) as a function of the linker length, and the linker length obtained from the experiment (dash line). (d) κ^2 of dyes in the

simulation and the value expected for complete rotational averaging (dash line). Error bars are the standard error of the mean, computed from block averages.

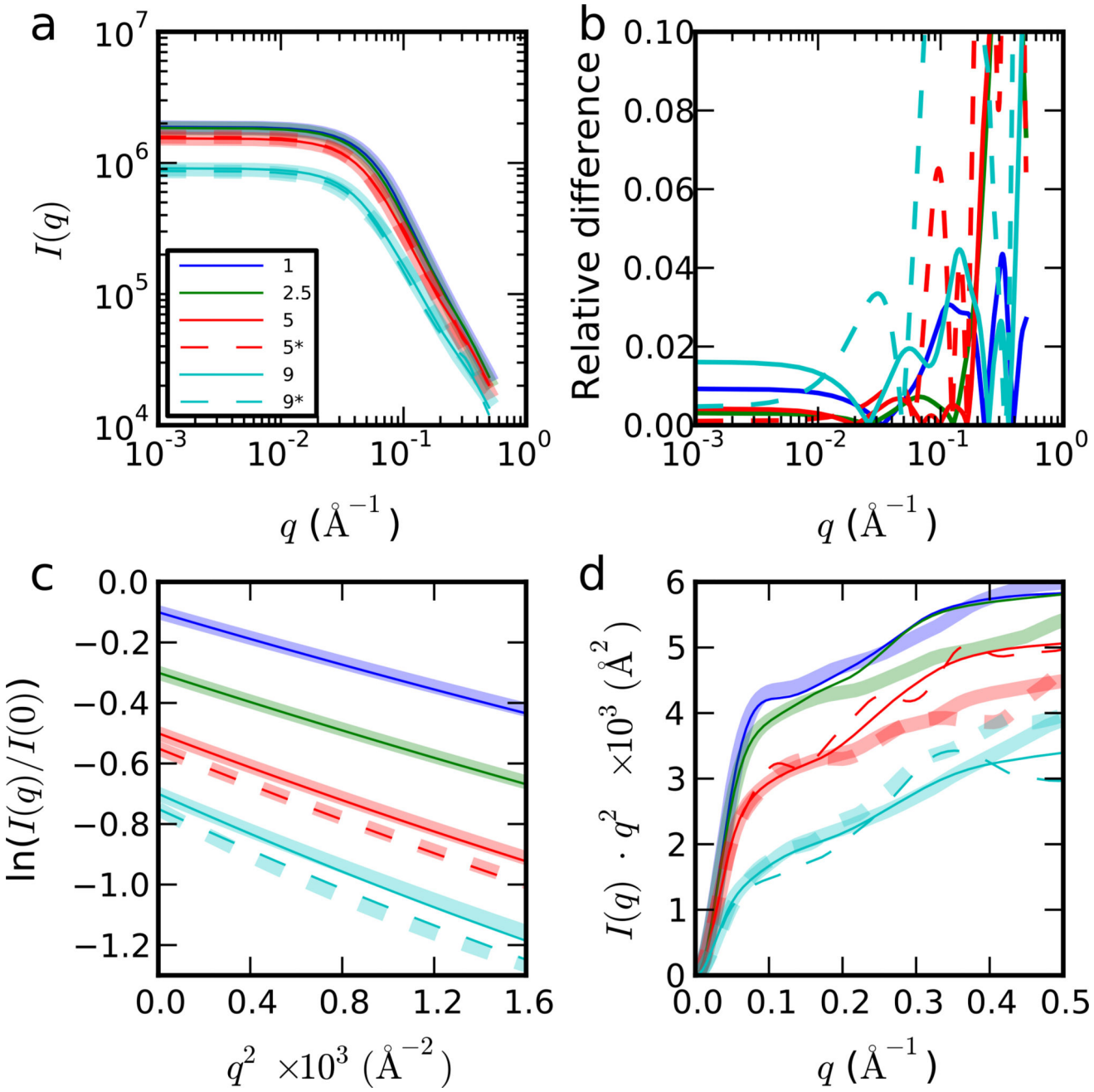


Figure 6.

Explicit (always shown by the thick curve in (a), (c) and (d)) and implicit (thin line) treatment of solvent in SAXS calculation. (a) Log-log plot of scattering intensity with legend showing the urea concentration. Dashed lines (* in legend) indicate simulations with a larger simulation box at 5 and 9 M urea. (b) Relative difference between the scattering intensity of explicit and implicit treatment of solvent in SAXS calculation, $|I_{\text{implicit}}(q) - I_{\text{explicit}}(q)|/I_{\text{explicit}}(q)$. (c) Guinier plot. (d) Kratky plot. We obtain essentially identical results either from all-atom or CRYSOL calculation at low scattering angles (i.e. solvent structuring does not influence the measured R_g). For all the comparisons, the scattering intensity from

implicit treatment of solvent is scaled by a factor a to minimize $\Sigma_q (aI_{\text{implicit}}(q) - I_{\text{explicit}}(q))^2$, to correct for differences of scattering intensities between simulation and experiment.

Table 1

Power-law fitting exponent relating scaling of internal distances R_{ij} or included radii of gyration R_{gij} to sequence separation |i−j|. Errors in brackets are calculated by bootstrapping.

Urea (M)	R	R_g
0	0.54 (0.07)	0.55 (0.06)
1	0.56 (0.05)	0.57 (0.05)
2.5	0.58 (0.04)	0.58 (0.04)
5	0.60 (0.05)	0.61 (0.03)
7	0.62 (0.04)	0.62 (0.03)
9	0.63 (0.04)	0.62 (0.03)

Table 2

Pairwise reduced χ^2 between the simulation and experimental observables. $\chi^2_{\mathrm{FRET}}=<(E_{\mathrm{sim}}-E_{\mathrm{exp}})^2/(\sigma^2_{\mathrm{sim}}+\sigma^2_{\mathrm{exp}})>$ where and E_{sim} and E_{exp} are the simulation and experimental FRET efficiencies, and σ^2_{sim} and σ^2_{sim} the corresponding variances and the average is over denaturant concentrations. $\chi^2_{\mathrm{SAXS}}=(\mathrm{I}_{\mathrm{sim}}-\mathrm{I}_{\mathrm{exp}})^T\sum^{-1}_{\mathrm{exp}}(\mathrm{I}_{\mathrm{sim}}-\mathrm{I}_{\mathrm{exp}})$, in which Σ_{exp} is the covariance matrix from 30 independent experimental measurements for each urea concentration, and $\sum^{-1}_{\mathrm{exp}}$ its pseudo-inverse. Rows and columns correspond to simulation and experimental urea concentrations, respectively, i.e. the lowest number in each column (in boldface) indicates the urea concentration (row) of the simulation which agrees best with that experiment (column). Ideally, the lowest numbers should be on the diagonal, or close to it, which occurs in most cases. For SAXS, we show the comparison for $q < 0.04$ Å^{-1}, the full range comparison (which is similar) can be found in Table S3. The scattering intensity from simulation is scaled by a factor α to correct for differences of magnitude of scattering intensities between simulation and experiment. (*L*) indicates the large box simulations at 5 and 9 M urea concentration.

		Experiment					
	FRET χ^2	*0 M*	*1.0 M*	*2.5 M*	*5.0 M*	*7.0 M*	*9.0 M*
Simulation	*0.0 M*	**0.92**	1.44	2.81	5.02	6.51	7.95
	1.0 M	1.89	1.67	3.25	6.67	9.18	11.7
	2.5 M	3.75	**1.16**	**0.27**	1.07	2.16	3.45
	5.0 M	7.26	3.65	1.29	**0.28**	**0.26**	0.52
	5.0 M (L)	37.9	24.4	13.6	6.19	3.55	2.08
	7.0 M	95.5	63.2	36.8	17.8	10.6	6.06
	9.0 M	66.3	46.2	29.7	17.2	12.1	8.43
	9.0 M (L)	8.38	5.16	2.69	1.08	0.56	**0.30**
	SAXS χ^2	*1.0 M*	*2.5 M*	*5.0 M*	*9.0 M*		
	1.0 M	1.59	3.99	5.98	1.90		
	2.5 M	**1.53**	**3.26**	4.79	1.54		
	5.0 M	1.96	4.25	3.94	1.26		
	5.0 M (L)	2.43	5.73	**3.77**	1.17		
	9.0 M	2.71	6.67	3.78	**1.16**		
	9.0 M (L)	3.35	9.97	4.27	1.19		