

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

Wenwei Zheng,<sup>\*,†</sup> Alessandro Borgia,<sup>‡</sup> Karin Buholzer,<sup>‡</sup> Alexander Grishaev,<sup>§</sup> Benjamin Schuler,<sup>‡</sup> and Robert B. Best<sup>\*,†</sup>

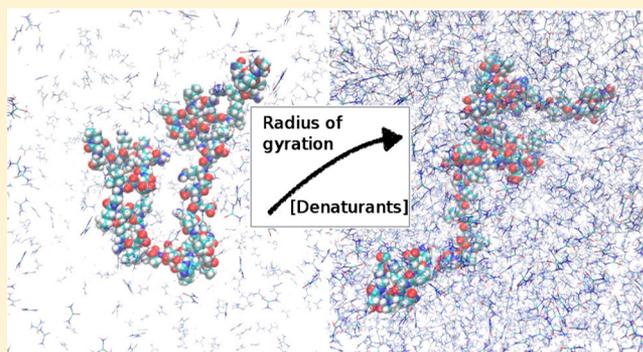
<sup>†</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, United States

<sup>‡</sup>Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

<sup>§</sup>National Institute of Standards and Technology and the Institute for Bioscience and Biotechnology Research, Rockville, Maryland 20850, United States

## S Supporting Information

**ABSTRACT:** Chemical denaturants are the most commonly used agents 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. Studies using Förster resonance energy transfer (FRET) and other methods found an increase in radius of gyration with denaturant concentration, but most small-angle X-ray scattering (SAXS) studies found 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 favorable 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 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 preferential solvent partitioning into the interior of the chain do not interfere with the determination of radius of gyration from the SAXS experiments.



## INTRODUCTION

Addition of chemical denaturants represents one of the most straightforward and widely used methods of perturbing the stability of proteins because they are simple to use and rarely involve protein aggregation artifacts that 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.<sup>1–3</sup> In the most accepted model, the denaturant binds weakly to the protein, favoring unfolding due to the greater surface available for binding in the unfolded state.<sup>1,4–10</sup> 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,<sup>11,12</sup> providing a sensitive measure of the denaturant action. Indeed, studies using a large number of experimental techniques have reached this conclusion,<sup>13</sup> including ensemble and single-molecule Förster resonance energy transfer (FRET),<sup>14–22</sup> dynamic light-scattering

Received: May 26, 2016

Published: September 1, 2016

(DLS),<sup>10,23</sup> nuclear magnetic resonance (NMR),<sup>24–26</sup> and small-angle X-ray scattering (SAXS).<sup>25,27</sup>

However, a significant number of studies, using SAXS on unfolded proteins,<sup>28–32</sup> have reached a qualitatively different conclusion, i.e. the radius of gyration ( $R_g$ ) of the unfolded state of a two-state protein does not vary with denaturant concentration. For intrinsically disordered proteins<sup>33</sup> that are easier to study experimentally because there is no folded state which needs to be separated, the results also differ according to the method used: FRET experiments suggest an expansion with denaturant concentration for a number of IDPs,<sup>34,35</sup> whereas 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.<sup>36</sup> The reasons for the differences between the conclusions drawn from these different experiments are unclear but important to resolve. First, 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. Second, if proteins do not expand with increasing denaturant concentration, then 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.<sup>37,38</sup>

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 extent of collapse in the absence of chromophores.<sup>31,32</sup> SAXS requires highly homogeneous samples and relatively high protein concentration, but should provide a robust readout of the protein dimensions 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 because 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, toward the center of the coil rather than at the periphery or vice versa. This type of effect could arise 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 measurably alter  $R_g$  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<sup>39,40</sup> or

a continuum solvent model to interpret SAXS experiments,<sup>41</sup> in order to solve the inverse problem of reconstructing molecular properties from a small number of observables. 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 (because they are themselves based on empirically derived force fields).<sup>42</sup> 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, then they can 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.<sup>43</sup> Molecular simulations have already been used in a large number of studies to determine the effect of chemical denaturants on protein stability,<sup>5,44,45</sup> the unfolded state,<sup>44,46,47</sup> and the mechanism by which they denature proteins.<sup>4,5,8,46,48–53,54</sup> 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  $\sim 0.5$  kcal/mol too favorable per residue.<sup>6</sup> Other studies have also found too strong an association of urea with peptide models using a variety of force fields and different methods.<sup>5,55</sup> 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 parametrized 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.<sup>56</sup> 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<sup>35</sup> in which the chromophores were explicitly represented, as well as  $m$ -values for denaturation of the Trp cage miniprotein.<sup>44</sup> Here, we apply this force field to study the denaturation of the intrinsically disordered protein ACTR (activator for thyroid hormone and retinoid receptors)<sup>57,58</sup> 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 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 artifacts 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<sup>59</sup> simulation code at a constant temperature of 298 K (maintained by a Langevin thermostat with a friction coefficient of 0.1 ps<sup>-1</sup>) and pressure of 1 bar (with a Parrinello–Rahman barostat).<sup>60</sup> The time step was 2 fs; electrostatic energies and forces were computed with particle-mesh Ewald<sup>61</sup> using a 0.12 nm grid spacing and real-space cutoff 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<sup>56</sup> was used for the protein together with the TIP4P/2005 water model<sup>62</sup> and KBFFs model for urea and GdmCl,<sup>44</sup> i.e., the Kirkwood–Buff force field (KBFF) model<sup>63,64</sup> including scaled denaturant–protein interactions in order to balance protein–denaturant interactions. The dye force field was described in a previous work.<sup>65</sup> 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<sup>66</sup> and is briefly described here. The SAXS intensity  $I(q)$ , as a function of the momentum transfer  $q$ , is calculated by

$$I(q) = \sum_{ij} f_i(q) f_j(q) [\Delta I_{ij}(q) + \nu I_{ij}(q)] \quad (1)$$

in which  $\Delta I_{ij}$  represents the partial intensity difference between protein with solution (foreground) and pure solution (background), and  $f_i(q)$  is the form factor of species  $i$ . The term  $\nu I_{ij}(q)$  adds back the bulk solvent contribution oversubtracted in  $I_{ij}$ . This correction is not used in the current work because it has been shown in the original literature<sup>66</sup> to have limited influence on the SAXS intensity for the range of  $q$  we are interested in ( $q < 0.5 \text{ \AA}^{-1}$ ). 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.  $\Delta I_{ij}$  can then be calculated from all-atom MD data by

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

in which  $\Delta N_i$  is the difference of the average number of particles of species  $i$  between the foreground and background simulations,  $\Delta H_{ij}$  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 CRY SOL<sup>41</sup> and FOXS.<sup>67</sup> Here, we briefly describe CRY SOL and refer the readers to the original literature for more details. The scattering intensity is calculated by three terms in the following

expression:  $I(q) = \langle |A_a(\mathbf{q}) - \rho_w A_c(\mathbf{q}) + (\rho_b - \rho_w) A_b(\mathbf{q})|^2 \rangle_\Omega$ . The first scattering amplitude is that of the protein *in vacuo*, the second is from the volume excluded to the solvent, and the third is from the surface hydration layer with a higher density than the bulk solution.  $\rho_b$  and  $\rho_w$  represent the scattering density of the surface layer and pure solvent, respectively, and  $\Omega$  is 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. The hydration layer is empirically estimated, so just protein coordinates are required to calculate the SAXS intensity using CRY SOL. 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 donor fluorescence lifetime<sup>68</sup> 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<sup>68,69</sup>

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

where the averaging is over all frames of the trajectory, and  $R_0$  is the spectroscopically determined Förster radius<sup>70</sup> for the donor and acceptor dyes used here, AlexaFluor 488 and AlexaFluor 594.<sup>14</sup>  $R_0$  is defined from the refractive index,  $n$ , and its value in the absence of denaturant,  $R_0(0) = 5.4 \text{ nm}$ , as<sup>70</sup>

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

In this expression,  $n(0)$  is the refractive index at 0 M and  $n([\text{urea}]) = 1.3361 + 0.00841[\text{urea}]$ , in which [urea] is the urea concentration in M (this curve is determined for a solution of 50 mM sodium phosphate, 140 mM  $\beta$ -mercaptoethanol, and 0.01% Tween 20).<sup>71</sup> It is assumed that changes in the donor quantum yield and spectral overlap integral do not significantly change with denaturant concentration. The chromophores were not present in most of the simulations, so in these cases  $R$  was calculated between the  $C_\alpha$  of the residues labeled in the experiment. The distance is then rescaled by a factor of

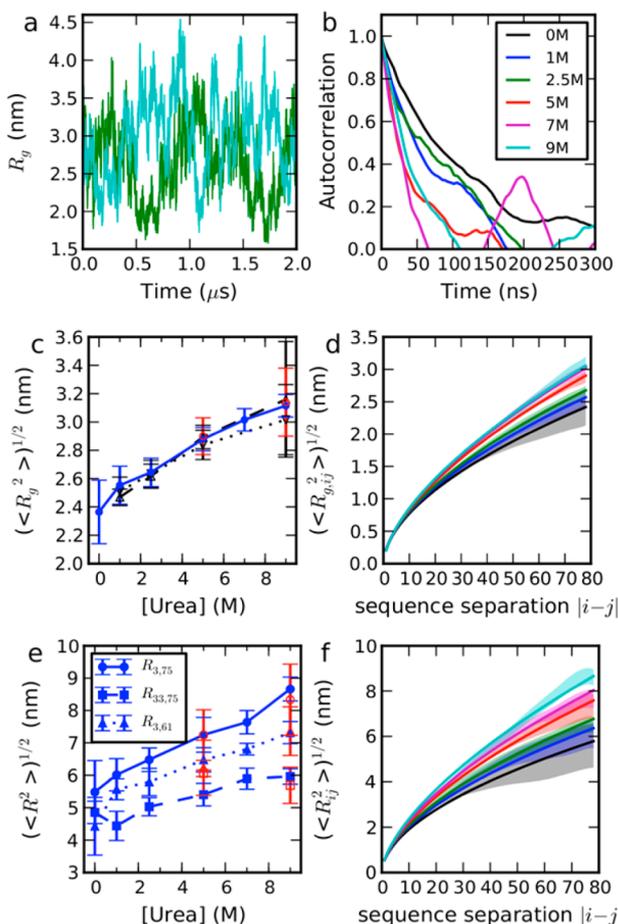
$$\gamma = \left[ \frac{N + N_{\text{linker}}}{N} \right]^\nu \quad (5)$$

in which  $N$  is the number of protein residues between the FRET dyes in the experiment,  $\nu$  is the Flory scaling exponent determined from the scaling of internal chain distances with sequence separation in the simulation (Figure 1),<sup>72</sup> and  $N_{\text{linker}}$  is a free parameter representing the length of both linkers between the dyes and the protein in terms of an equivalent number of protein residues. It has previously been estimated empirically to be  $\sim 9$ .<sup>17,35</sup>

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  $C_\alpha$  are used. The second is to use the distances between the  $C_1$  atoms of each chromophore as described previously<sup>65</sup> 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.<sup>65,73–75</sup> In this case, the transfer rate  $k_{\text{ET}}(x)$  for configuration  $x$  in the simulation trajectory is given by

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

where  $k_D$  is the donor fluorescence decay rate in the absence of an acceptor and the orientational factor  $\kappa$  is given by



**Figure 1.** Equilibrium properties of ACTR in urea. (a) Fluctuations in  $R_g$  over time at selected denaturant concentrations (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  $q < 0.04 \text{ \AA}^{-1}$  from explicit solvent calculation (black dashed 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) Dependence of  $R_g$  for the segment included between residues  $i$  and  $j$  on the sequence separation  $|i - j|$ . (e) Dependence of root-mean-square intramolecular distances ( $R$ ) on denaturant concentration. Red symbols show data from large box simulations (see Table S1). (f) Dependence of root-mean-square distance between residues  $i$  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  $\langle R_{ij}^2 \rangle^{1/2} = \sqrt{2l_p b} N^\nu$  and  $\langle R_{g,ij}^2 \rangle^{1/2} \approx \sqrt{\frac{2l_p b}{(2\nu + 1)(2\nu + 2)}} N^\nu$ , in which  $l_p = 0.4 \text{ nm}$  and  $b = 0.38 \text{ nm}^3$ <sup>35</sup> (the expression for  $R_g$  is an approximation which assumes Flory scaling to hold for all  $ij$  pairs, although it is only strictly valid for sufficiently large  $|i - j|$ ). Simulation error bars are obtained from block averages, as described in the Supporting Information text and refer to the standard error of the mean.

$$\kappa = \hat{\mu}_D \cdot \hat{\mu}_A - 3(\hat{R} \cdot \hat{\mu}_A)(\hat{R} \cdot \hat{\mu}_D) \quad (7)$$

where  $\hat{\mu}_D$  and  $\hat{\mu}_A$  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.<sup>65</sup> 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:

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

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

$$\langle E \rangle = 1 - k_D \int_0^{t_{\max}} I(t) dt \quad (9)$$

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

## RESULTS AND DISCUSSION

**ACTR Expands in Denaturant Solution.** To sample the configurations of the intrinsically disordered protein ACTR, we ran multiple unbiased,  $2 \mu\text{s}$  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  $0.05\text{--}0.1 \mu\text{s}$ .<sup>17,76–78</sup> We use force field models for protein, urea, and water which we have recently parametrized to reproduce the balance of interactions between the protein, water, and denaturant components of the system.<sup>44,56</sup> 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,<sup>79,80</sup> with several suggested corrections proposed.<sup>56,81,82</sup> This would confound any attempt at quantitative comparison with experiment.<sup>56</sup> 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 functions 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.<sup>77,78</sup> 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 1e). 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 1f).<sup>72</sup> 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 1d). 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  $\theta$ -solvent conditions in water to close to the excluded volume limit in denaturant.<sup>35</sup> The finding of near

**Table 1. Power-Law Fitting Exponent Relating Scaling of Internal Distances  $R_{ij}$  or Included Radii of Gyration  $R_{g,ij}$  to Sequence Separation<sup>a</sup>  $|i - j|$**

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)

<sup>a</sup>Errors in brackets are calculated by bootstrapping.

$\theta$ -solvent conditions in water is also consistent with the fractal dimension from SAXS experiments on reduced RNase A in water,<sup>83</sup> 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.<sup>84</sup> In addition, our results are consistent, qualitatively, with the expectations of polymer theory for the changes which occur when the solvent quality is improved.<sup>12</sup> 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  $\theta$ -state.<sup>35</sup> Thus, in analogy with protein folding,<sup>85,86</sup> 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.<sup>87,88</sup>

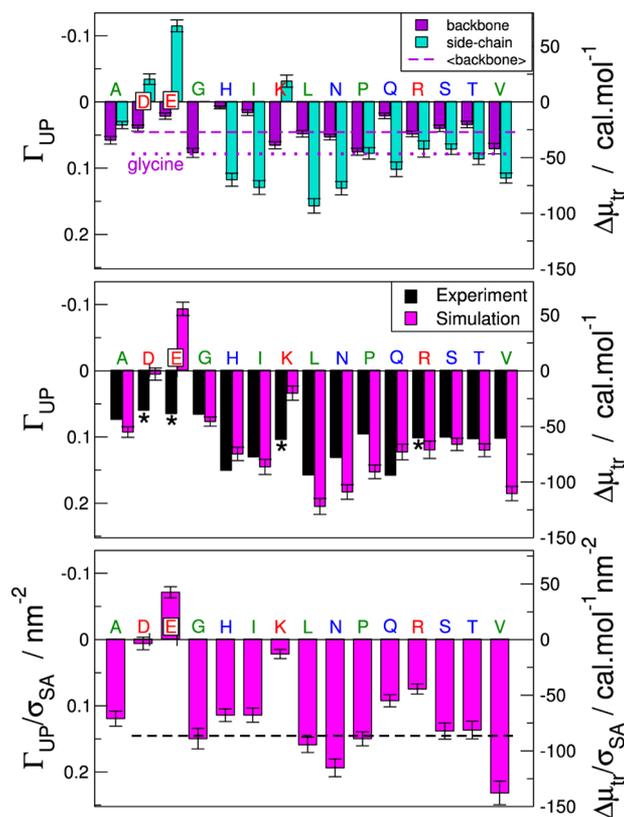
**Denaturing Mechanism of Urea.** The clear expansion of the chain in urea implies an improvement in solvent quality with increasing denaturant concentration (an alternative explanation for increased  $R_g$  might be an increase in chain stiffness, but that would not explain the increase of scaling exponent). 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;<sup>89</sup> 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  $\Gamma_{UP}$  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.<sup>90,91</sup> That is,  $\Gamma_{UP}$  measures how much urea must be added to keep the bulk urea chemical potential  $\mu_U$  constant when a protein is added to the solution and is expected to be positive if urea interacts favorably with the protein and vice versa. In simulations, the coefficient can be estimated very simply from the heuristic relation:<sup>90–93</sup>

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

In this equation,  $n_U^P$  and  $n_W^P$  are the numbers 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.,  $\Gamma_{UP}$  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 cutoff 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  $\Gamma_{UP}$  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.<sup>94,95</sup> 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.



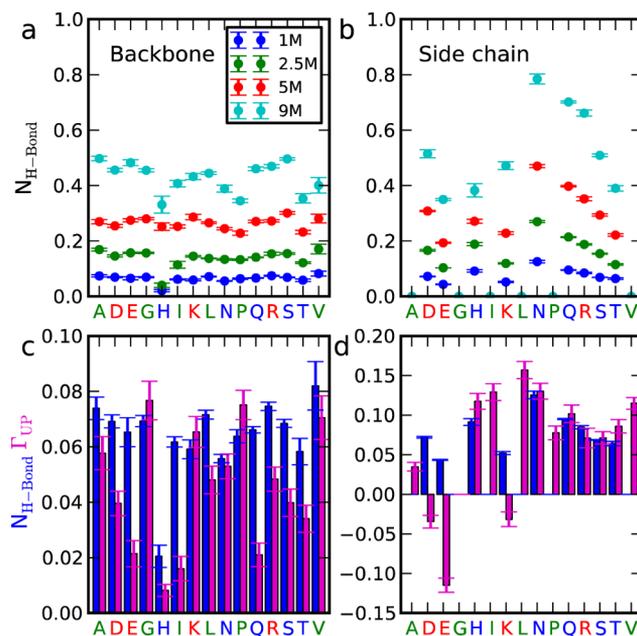
**Figure 2.** Preferential interaction coefficients  $\Gamma_{UP}$  (left axis) for association of urea with the protein surface at 1 M urea. (Top) Decomposition of  $\Gamma_{UP}$  for each type of residue into backbone and side-chain contributions. Broken line is the average backbone contribution across all residues. Dotted line is  $\Gamma_{UP}$  for glycine. (Center) Comparison of  $\Gamma_{UP}$  for whole residues with experimental water to urea transfer free energies.<sup>96</sup> Transfer free energies (right axis) were approximated from  $\Gamma_{UP}$  using  $\Delta\mu_{tr} \approx -RT\Gamma_{UP}$ , where  $R$  and  $T$  are the molar gas constant and absolute temperature, respectively.<sup>91</sup> Transfer free energies for charged residues (\*) are not directly comparable because they necessarily include counterions. (Bottom) Preferential interaction coefficients  $\Gamma_{UP}$  and transfer free energies  $\Delta\mu_{tr}$  normalized by the average surface area of each residue,  $\sigma_{SA}$ . Broken line is the average  $\Gamma_{UP}/\sigma_{SA}$  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.

We see that urea interacts favorably with the backbone of all residue types, although there is some residue to residue variation. While we have presented the average  $\Gamma_{UP}$  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,<sup>1,9,96</sup> namely, the decomposition of protein folding  $m$ -values as a sum over independent contributions from different functional groups in the polypeptide chain.<sup>9</sup> However, 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 Figure 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,<sup>97</sup> 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.<sup>8</sup> 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.<sup>7,93</sup> 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)<sup>98</sup> would be needed to evaluate the relative contribution of the backbone to folding  $m$ -values.<sup>99</sup>

Experimental preferential interaction coefficients are not available for all residue types, so we have compared our results with per-residue transfer free energies from water to 1 M urea.<sup>96</sup> 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.<sup>91</sup> A concentration of 1 M is sufficiently low for the first assumption to be valid, and urea solutions are known to be very close to ideal<sup>100</sup> (also reflected in the properties of the KBFF force field).<sup>63</sup> 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 that this is also a reasonable first approximation. 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 (the Pearson correlation coefficient of the transfer free energies between the simulation and experiment is 0.63 with a  $p$ -value of 0.01). 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  $\text{Na}^+$  or  $\text{Cl}^-$  salts, so these 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, with 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 that of bulk (given by the preferential interaction



**Figure 3.** Hydrogen bonding between urea and protein. (a) and (b) Number of hydrogen bonds per urea molecule to the backbone and side-chains, respectively, averaged by residue type, at different urea concentrations. (c) and (d) Comparison of 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.

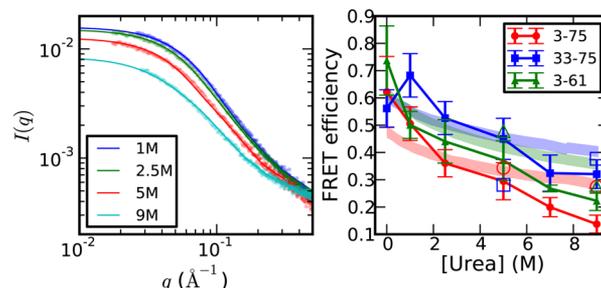
coefficient). This strongly suggests that hydrogen bonding is the main mode of specific interaction with the backbone, with many of the side-chains also making hydrogen-bonded interactions with urea. In addition, however, it is clear that the hydrophobic side-chains interact favorably without forming 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. There are no residues with aromatic side chains in ACTR, yet these usually have the most favorable water–urea transfer free energies in experiment.<sup>96</sup> We have therefore calculated the preferential interaction coefficients of the unfolded Trp cage mini protein using published simulations with the same force field in 3 M urea<sup>44</sup> (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 for sequences containing also aromatic residues.

*Denaturing Mechanism of Guanidinium Chloride.* 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<sup>+</sup> ions have a significant preferential interaction with the backbone, both Gdm<sup>+</sup> and Cl<sup>-</sup> 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.<sup>101</sup> 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*,<sup>98</sup> although the native state must also contribute.<sup>99</sup> 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<sup>+</sup> and urea interact by hydrogen bonding with the backbone. However, the type of hydrogen bonds formed is different, with Gdm<sup>+</sup> hydrogen bonding exclusively to the CO group of the amide bond (as may be expected from its lack of hydrogen bond acceptors) and urea bonding 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.<sup>102</sup> Base-catalyzed exchange was found to be blocked by urea and unaffected by GdmCl, which is expected because 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<sup>+</sup> does not hydrogen bond to the CO group either,<sup>102</sup> 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.<sup>103</sup> 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,<sup>1</sup> the model favored by most recent studies,<sup>4–8,49,50,52,104</sup> 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 NMR spectroscopy.<sup>57,105</sup> 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.<sup>106</sup> 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 not only ACTR to collapse (due to strengthened hydrophobic effect)<sup>107</sup> but also an apparent reduction of helix content<sup>105</sup> as has also been observed in all-atom simulations of unfolded proteins.<sup>108</sup> Thus, there is not a simple connection between 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<sup>71</sup> with that calculated from the all-atom simulations. In Figure 4, we show the mean FRET efficiency



**Figure 4.** Comparison of simulations of ACTR in urea with experimental observables. Left: SAXS; right: FRET. Experimental data (from ref 71) and uncertainties are represented by shaded areas and simulation data are represented by solid lines. Solid symbols show the results of simulations with a common box size (12 nm rhombic dodecahedron), and open symbols represent the results of using larger box sizes (15 and 17 nm at 5 and 7 M, respectively). Large box simulations were run for 0.6  $\mu$ s versus 2  $\mu$ s for the small box simulations. For calculating FRET efficiencies, the distance between the C $_{\alpha}$  atoms of the labeled residues has been rescaled by a factor  $\left[\frac{N+9}{N}\right]^{\nu}$  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.

computed from the simulations using the Förster equation (eq 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 the simulation FRET efficiency from experiment. This is probably due to the limited box size affecting the end-to-end distance of more expanded configurations and the 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  $\mu$ s versus 2  $\mu$ s for the small box simulations. Even with the deviation at high urea concentration, the Pearson correlation coefficient between the FRET efficiencies from simulation and experiment is 0.91 with a *p*-value of the order of 10<sup>-7</sup>, suggesting the agreement between experiment and simulation is overall quite good. In Table 2, an all-versus-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 because 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 $_{\alpha}$  of the labeled residues (eq 5). In addition, we assume that the efficiency is determined only by the donor–acceptor distance and that the FRET orientational factor  $\kappa^2 = 2/3$ .<sup>109</sup> We will revisit and justify both of these assumptions in the next section.

**Table 2. Pairwise Reduced  $\chi^2$  between the Simulation and Experimental Observables<sup>a</sup>**

Simulation	Experiment					
FRET $\chi^2$	0 M	1.0 M	2.5 M	5.0 M	7.0 M	9.0 M
0.0 M	<b>0.92</b>	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	<b>1.16</b>	<b>0.27</b>	1.07	2.16	3.45
5.0 M	7.26	3.65	1.29	<b>0.28</b>	<b>0.26</b>	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	<b>0.30</b>
SAXS $\chi^2$	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	<b>1.53</b>	<b>3.26</b>	4.79	1.54		
5.0 M	1.96	4.25	3.94	1.26		
5.0 M (L)	2.43	5.73	<b>3.77</b>	1.17		
9.0 M	2.71	6.67	3.78	<b>1.16</b>		
9.0 M (L)	3.35	9.97	4.27	1.19		

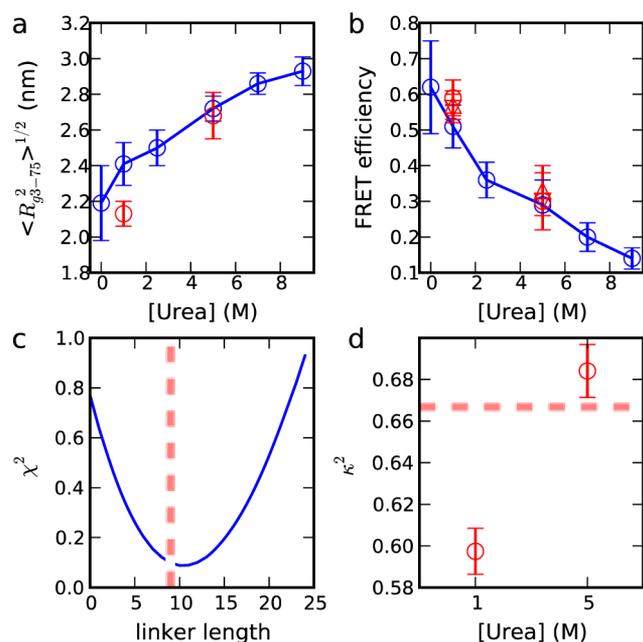
<sup>a</sup> $\chi_{\text{FRET}}^2 = \langle (E_{\text{sim}} - E_{\text{exp}})^2 / (\sigma_{\text{sim}}^2 + \sigma_{\text{exp}}^2) \rangle$  where  $E_{\text{sim}}$  and  $E_{\text{exp}}$  are the simulation and experimental FRET efficiencies,  $\sigma_{\text{sim}}^2$  and  $\sigma_{\text{exp}}^2$  the corresponding squared standard errors, and the average is over denaturant concentrations.  $\chi_{\text{SAXS}}^2 = (I_{\text{sim}} - I_{\text{exp}})^T \Sigma_{\text{exp}}^{-1} (I_{\text{sim}} - I_{\text{exp}})$ , in which  $\Sigma_{\text{exp}}$  is the covariance matrix from 30 independent experimental measurements for each urea concentration, and  $\Sigma_{\text{exp}}^{-1}$  its pseudoinverse. 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  $q < 0.04 \text{ \AA}^{-1}$ , the full range comparison (which is similar) can be found in Table S3. The scattering intensity from simulation is scaled by a factor  $\alpha$  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.

We have also computed SAXS scattering profiles  $I(q)$  using the all-atom coordinates via an established procedure.<sup>66</sup> To do this, we compute within a spherical volume around the protein center of mass the atom–atom pair distance distribution functions (PDDFs), 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 that does not completely contain the vast majority (i.e., ~99%) of the disordered protein configurations distorts the results, for example, by 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 considerations 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 artifact of confinement due to the smaller system size 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), with 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  $\chi^2$  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.<sup>31,32</sup> 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 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.<sup>65</sup> An initial comparison of the radius of gyration shows that at both 1 and 5 M urea  $R_g$  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 5a). It is clear, however, that  $R_g$  increases with denaturant concentration both for the labeled and unlabeled systems. In a previous study,<sup>110</sup> 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)<sup>111</sup> 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  $C_\alpha$  atoms are included in Figure 5b (detailed 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,  $\kappa^2$ , needs to be considered, i.e.,  $\langle \kappa^2 \rangle = 2/3$ . In the simulations with explicit chromophores, we have a complete, unbiased trajectory of the chromophore positions, so we can directly calculate 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



**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  $C_\alpha$  distance with a correction of linker length of 9 residues (circle),  $C_1$  distance between the dyes without the correction (down triangle), and integration of the intensity decay directly (up triangle). (c) Variation of  $\chi^2$  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 (dashed line). (d)  $\kappa^2$  of dyes in the simulation and the value expected for complete rotational averaging (dashed line). Error bars are the standard error of the mean, computed from block averages.

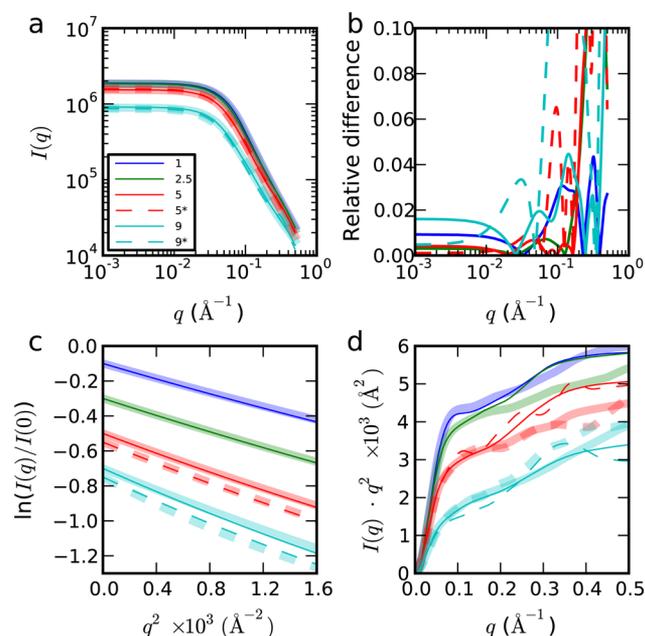
donor fluorescence decay is shown in Figure S12, and the transfer efficiencies are shown in Figure 5b. The consistency of the different calculations provides strong support both for the simple distance-based FRET estimate as well as for the assumption of  $\langle \kappa^2 \rangle = 2/3$ . Indeed, the equilibrium average  $\kappa^2$  computed from the simulations is very close to the expected value of  $2/3$  for an isotropic distribution of chromophore orientations (Figure 5d), as seen in an earlier study.<sup>110</sup> 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  $\sim 1$  ns, with a similar correlation time for  $\kappa^2$  itself (Figure S12), compared with donor lifetimes of  $\sim 2$  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  $\kappa^2$ . 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  $\kappa^2$  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 5b) 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  $R_{\text{obs}}$  by assuming that the linkers effectively add a certain number of extra residues to the length of the chain ( $N$ ) so that the distance between protein residues is  $R = (N/(N + N_{\text{linker}}))^\nu R_{\text{obs}}$ , where the number of extra residues  $N_{\text{linker}}$  has been chosen to be around 9 from the literature<sup>35</sup> and  $\nu$  is the polymer scaling exponent (from Table 1). In the simulations with attached chromophores we can measure both distances, so we determine  $N_{\text{linker}}$  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  $N_{\text{linker}}$ -dependent correction. The  $\chi^2$  between these two estimates is shown in Figure 5c, yielding a minimum at  $N_{\text{linker}} \approx 10$  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.<sup>66,112</sup> 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, e.g., if 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 CRY SOL<sup>41</sup> and FOXS.<sup>67</sup>

Note that CRY SOL 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  $\sim 10\%$  of the bulk density, respectively, while FOXS is also optimized for water. For noncompact 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 CRY SOL 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  $q < 0.3 \text{ \AA}^{-1}$  and excellent for  $q < 0.04 \text{ \AA}^{-1}$ , 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 and Figure S13) and of leaving out the hydration shell altogether (Figure S14). Although the latter of course leads to larger deviations for  $q > 0.08 \text{ \AA}^{-1}$ , 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,<sup>67</sup> in which the surface solvent is modeled by adjusting the atomic form factor of the solvent-



**Figure 6.** Explicit (always shown by the thick curve in panels 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 CRY SOL 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  $\alpha$  to minimize  $\sum_q (\alpha I_{\text{implicit}}(q) - I_{\text{explicit}}(q))^2$ , to correct for differences of scattering intensities between simulation and experiment.

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 and 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 1c). 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 toward the center 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, while 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 results of simulations with GdmCl suggest that it operates via a similar mechanism, although with preferential interaction coefficients about double those of urea, consistent with protein folding  $m$ -values. The major difference from urea is that GdmCl only hydrogen bonds to the CO, and not the NH, moiety of the amide group, and that the Gdm<sup>+</sup> ion has a higher affinity for acidic side-chains.

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 artifacts 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. There is no fundamental inconsistency between the two experiments, so in the absence of the above artifacts, the experimental discrepancy most likely relates to the challenging inverse problem of determining properties of IDPs from limited experimental data.<sup>71</sup>

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 all-atom, explicit solvent, force fields used are reaching the point of being a useful tool for the investigation of intrinsically disordered proteins.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b05443.

Methods, figures, and tables (PDF)

## ■ AUTHOR INFORMATION

## Corresponding Authors

\*wenwei.zheng@nih.gov

\*robertbe@helix.nih.gov

## Notes

The authors declare no competing financial interest.

## ■ 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, and A.G. was supported by intramural funding from the National Institute of Standards and Technology. The work at the University of Zurich was supported by the Swiss National Science Foundation. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>).

## ■ REFERENCES

- (1) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1963**, *238*, 4074–4081.
- (2) Frank, H. S.; Franks, F. *J. Chem. Phys.* **1968**, *48*, 4746–4757.
- (3) Schellman, J. A. *Biopolymers* **1978**, *17* (5), 1305–1322.
- (4) Camilloni, C.; Guerini Rocco, A.; Eberini, I.; Gianazza, E.; Broglia, R. A.; Tiana, G. *Biophys. J.* **2008**, *94*, 4654–4661.
- (5) Canchi, D. R.; Paschek, D.; García, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338–2344.
- (6) Horinek, D.; Netz, R. R. *J. Phys. Chem. A* **2011**, *115*, 6125–6136.
- (7) Moeser, B.; Horinek, D. *J. Phys. Chem. B* **2014**, *118*, 107–114.
- (8) Stumpe, M. C.; Grubmüller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126–16131.
- (9) Guinn, E. J.; Pegram, L. M.; Capp, M. W.; Pollock, M. N.; Record, M. T. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16932–16937.
- (10) Holthauzen, L. M. F.; Rösgen, J.; Bolen, D. W. *Biochemistry* **2010**, *49*, 1310–1318.
- (11) De Gennes, P.-G. *Scaling Concepts in Polymer Physics*. Cornell University Press: Ithaca and London, 1979.
- (12) Flory, P. J. *Principles of Polymer Chemistry*. Cornell University Press: Ithaca and London, 1953.
- (13) Haran, G. *Curr. Opin. Struct. Biol.* **2012**, *22*, 14–20.
- (14) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743–747.
- (15) Sherman, E.; Haran, G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11539–11543.
- (16) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1528–1533.
- (17) Hoffmann, A.; Kane, A.; Nettels, D.; Hertzog, D. E.; Baumgärtel, P.; Lengsfeld, J.; Reichardt, G.; Horsley, D. A.; Seckler, R.; Bakajin, O.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 105–110.
- (18) Deniz, A. A.; Laurence, T. A.; Beligere, G. S.; Dahan, M.; Martin, A. B.; Chemla, D. S.; Dawson, P. E.; Schultz, P. G.; Weiss, S. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (10), 5179–5184.
- (19) Kuzmenkina, E. V.; Heyes, C. D.; Nienhaus, G. U. *J. Mol. Biol.* **2006**, *357*, 313–324.
- (20) Huang, F.; Sato, S.; Sharpe, T. D.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 123–127.
- (21) Mukhopadhyay, S.; Krishnan, R.; Lemke, E. A.; Lindquist, S.; Deniz, A. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2649–2654.
- (22) Möglich, A.; Joder, K.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12394–12399.
- (23) Nöppert, A.; Gast, K.; Müller-Frohne, M.; Zirwer, D.; Damaschun, G. *FEBS Lett.* **1996**, *380*, 179–182.
- (24) Mok, Y.-K.; Kay, C. M.; Kay, L. E.; Forman-Kay, J. *J. Mol. Biol.* **1999**, *289*, 619–638.
- (25) Choy, W.-Y.; Mulder, F. A. A.; Crowhurst, K. A.; Muhandiram, D. R.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. *J. Mol. Biol.* **2002**, *316*, 101–112.
- (26) Lindorff-Larsen, K.; Kristjansdottir, S.; Teilum, K.; Fieber, W.; Dobson, C. M.; Poulsen, F. M.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 3291–3299.
- (27) Konuma, T.; Kimura, T.; Matsumoto, S.; Goto, Y.; Fujisawa, T.; Fersht, A. R.; Takahashi, S. *J. Mol. Biol.* **2011**, *405*, 1284–1294.
- (28) Plaxco, K. W.; Millett, I. S.; Segel, D. J.; Doniach, S.; Baker, D. *Nat. Struct. Biol.* **1999**, *6*, 554–556.
- (29) Jacob, J.; Krantz, B.; Dothager, R. S.; Thiyagarajan, P.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *338*, 369–382.
- (30) Jacob, J.; Dothager, R. S.; Thiyagarajan, P.; Sosnick, T. R. *J. Mol. Biol.* **2007**, *367*, 609–615.
- (31) Yoo, T. Y.; Meisburger, S. P.; Hinshaw, J.; Pollack, L.; Haran, G.; Sosnick, T. R.; Plaxco, K. *J. Mol. Biol.* **2012**, *418*, 226–236.
- (32) Watkins, H. M.; Simon, A. J.; Sosnick, T. R.; Lipman, E. A.; Hjelm, R. P.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 6631–6636.
- (33) Das, R. K.; Ruff, K. M.; Pappu, R. V. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112.
- (34) Müller-Spätth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Ruegger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14609–14614.
- (35) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16155–16160.
- (36) Johansen, D.; Trehwella, J.; Goldenberg, D. P. *Protein Sci.* **2011**, *20*, 1955–1970.
- (37) Takahashi, S.; Kamagata, K.; Oikawa, H. *Curr. Opin. Struct. Biol.* **2016**, *36*, 1–9.
- (38) Gianni, S.; Jemth, P. *Biophys. Chem.* **2016**, *212*, 17–21.
- (39) Song, J.; Gomes, G.-N.; Gradinaru, C. C.; Chan, H.-S. *J. Phys. Chem. B* **2015**, *119*, 15191–15202.
- (40) O'Brien, E.; Morrison, G.; Brooks, B. R.; Thirumalai, D. *J. Chem. Phys.* **2009**, *130*, 124903.
- (41) Svergun, D.; Barberato, C.; Koch, M. H. J. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
- (42) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134* (8), 3787–3791.
- (43) Peterscu, A.-J.; Receveur, V.; Calmettes, P.; Durand, D.; Smith, J. C. *Protein Sci.* **1998**, *7*, 1396–1403.
- (44) Zheng, W.; Borgia, A.; Borgia, M. B.; Schuler, B.; Best, R. B. *J. Chem. Theory Comput.* **2015**, *11*, 5543–5553.
- (45) Liu, Z.; Reddy, G.; Thirumalai, D. *J. Phys. Chem. B* **2016**, DOI: 10.1021/acs.jpcc.6b00327.
- (46) Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 5142–5147.
- (47) Holehouse, A. S.; Garai, K.; Lyle, N.; Vitalis, A.; Pappu, R. V. *J. Am. Chem. Soc.* **2015**, *137*, 2984–2995.
- (48) Wallqvist, A.; Covell, D. G.; Thirumalai, D. *J. Am. Chem. Soc.* **1998**, *120*, 427–428.
- (49) Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16928–16933.
- (50) O'Brien, E. P.; Dima, R.; Brooks, B. R.; Thirumalai, D. *J. Am. Chem. Soc.* **2007**, *129*, 7346–7353.
- (51) Yang, Z.; Xiu, P.; Shi, B.; Hua, L.; Zhou, R. *J. Phys. Chem. B* **2012**, *116*, 8856–8862.
- (52) Tirado-Rives, J.; Orozco, M.; Jorgensen, W. L. *Biochemistry* **1997**, *36*, 7313–7329.
- (53) Cafilisch, A.; Karplus, M. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 1746–1750.
- (54) Mountain, R. D.; Thirumalai, D. *J. Am. Chem. Soc.* **2003**, *125*, 1950–1957.
- (55) Shukla, D.; Shinde, C.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 12546–12554.
- (56) Best, R. B.; Zheng, W.; Mittal, J. J. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.
- (57) Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E. *Nature* **2002**, *415*, 549–553.

- (58) Soranno, A.; Koenig, I.; Borgia, M. B.; Hofmann, H.; Zosel, F.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4874–4879.
- (59) Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447.
- (60) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.
- (61) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (62) Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.
- (63) Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003**, *107*, 3891–3898.
- (64) Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2004**, *121* (5), 2180–2186.
- (65) Best, R. B.; Hofmann, H.; Nettels, D.; Schuler, B. *Biophys. J.* **2015**, *108*, 2721–2731.
- (66) Köfinger, J.; Hummer, G. *Phys. Rev. E* **2013**, *87*, 052712.
- (67) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. *Biophys. J.* **2013**, *105* (4), 962–974.
- (68) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2754–2759.
- (69) Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späh, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17800–17806.
- (70) Förster, T. *Ann. Phys.* **1948**, *437*, 55–75.
- (71) Borgia, A.; Zheng, W.; Buholzer, K.; Borgia, M. B.; Schuler, A.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; Best, R. B.; Schuler, B., Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *Submitted* 2016.
- (72) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8183–8188.
- (73) Hoefling, M.; Lima, N.; Haenni, D.; Seidel, C. A. M.; Schuler, B.; Grubmüller, H. *PLoS One* **2011**, *6* (5), e19791.
- (74) Hoefling, M.; Grubmüller, H. *Comput. Phys. Commun.* **2013**, *184*, 841–852.
- (75) Graen, T.; Hoefling, M.; Grubmüller, H. *J. Chem. Theory Comput.* **2014**, *10*, 5505–5512.
- (76) Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2655–2660.
- (77) Nettels, D.; Hoffmann, A.; Schuler, B. *J. Phys. Chem. B* **2008**, *112*, 6137–6146.
- (78) Borgia, A.; Wensley, B. G.; Soranno, A.; Nettels, D.; Borgia, M. B.; Hoffmann, A.; Pfeil, S. H.; Lipman, E. A.; Clarke, J.; Schuler, B. *Nat. Commun.* **2012**, *3*, 1195.
- (79) Nettels, D.; Müller-Späh, S.; Küster, F.; Hofmann, H.; Haenni, D.; Ruegger, S.; Reymond, L.; Hoffmann, A.; Kubelka, J.; Heinz, B.; Gast, K.; Best, R. B.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 20740–20745.
- (80) Piana, S.; Klepeis, J. L.; Shaw, D. E. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98–105.
- (81) Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T. *J. Phys. Chem. B* **2012**, *116*, 4524–4534.
- (82) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.
- (83) Wang, Y.; Trewhella, J.; Goldenberg, D. P. *J. Mol. Biol.* **2008**, *377*, 1576–1592.
- (84) Kohn, J. E.; Millett, I. S.; Jacob, J.; Zagrovic, B.; Dillon, T. M.; Cingel, N.; Dothager, R. S.; Seifert, S.; Thiyagarajan, P.; Sosnick, T. R.; Hasani, M. Z.; Pande, V. S.; Ruczinski, I.; Doniach, S.; Plaxco, K. W. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12491–12496.
- (85) Chahine, J.; Nymeyer, H.; Leite, V. B. P.; Socci, N. D.; Onuchic, J. N. *Phys. Rev. Lett.* **2002**, *88*, 168101.
- (86) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6369–6372.
- (87) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8868–8873.
- (88) De Sancho, D.; Best, R. B. *Mol. Biosyst.* **2012**, *8*, 256–267.
- (89) Huang, J.-r.; Gabel, F.; Jensen, M. R.; Grzesiek, S.; Blackledge, M. *J. Am. Chem. Soc.* **2012**, *134*, 4429–4436.
- (90) Timasheff, S. N. *Biochemistry* **1992**, *31*, 9857–9864.
- (91) Baynes, B. M.; Trout, B. L. *J. Phys. Chem. B* **2003**, *107*, 14058–14067.
- (92) Ma, L.; Pegram, L.; Record, M. T.; Cui, Q. *Biochemistry* **2010**, *49*, 1954–1962.
- (93) Canchi, D. R.; García, A. E. *Biophys. J.* **2011**, *100*, 1526–1533.
- (94) Voronoi, G. F. *J. Reine Angew. Math.* **1908**, *134*, 198–287.
- (95) Naidoo, K. J.; Kuttel, M. *J. Comput. Chem.* **2001**, *22*, 445–456.
- (96) Auton, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15065–15068.
- (97) König, G.; Boresch, S. *J. Phys. Chem. B* **2009**, *113*, 8967–8974.
- (98) Ziv, G.; Haran, G. *J. Am. Chem. Soc.* **2009**, *131*, 2942–2947.
- (99) Bolen, W. D.; Rose, G. D. *Annu. Rev. Biochem.* **2008**, *77*, 339–362.
- (100) Stokes, R. H. *Aust. J. Chem.* **1967**, *20*, 2087–2100.
- (101) Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1970**, *245*, 1648–1652.
- (102) Lim, W. K.; Rosgen, J.; Englander, S. W. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (8), 2595–2600.
- (103) Perrin, C. L.; Lollo, C. P.; Johnston, E. R. *J. Am. Chem. Soc.* **1984**, *106*, 2749–2753.
- (104) Canchi, D. R.; García, A. E. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273–293.
- (105) Kjaergaard, M.; Nørholm, A.-B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B. *Protein Sci.* **2010**, *19*, 1555–1564.
- (106) Bernadó, P.; Blackledge, M. *Biophys. J.* **2009**, *97* (10), 2839–2845.
- (107) Wuttke, R.; Hofmann, H.; Nettels, D.; Borgia, M. B.; Mittal, J.; Best, R. B.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 5213–5218.
- (108) Zerze, G. H.; Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2015**, *119* (46), 14622–14630.
- (109) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. Springer: New York, 2006.
- (110) Zerze, G. H.; Best, R. B.; Mittal, J. *Biophys. J.* **2014**, *107*, 1654–1660.
- (111) Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2010**, *114* (46), 14916–14923.
- (112) Lindner, B.; Smith, J. C. *Comput. Phys. Commun.* **2012**, *183*, 1491–1501.

# **Supporting Information for: “Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment”**

Wenwei Zheng<sup>†</sup>, Alessandro Borgia<sup>‡</sup>, Karin Buholzer<sup>‡</sup>, Alexander Grishaev<sup>§</sup>,  
Benjamin Schuler<sup>‡</sup>, Robert B. Best<sup>†</sup>

<sup>†</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520, USA.

<sup>‡</sup>Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

<sup>§</sup>National Institute of Standards and Technology and the Institute for Bioscience and Biotechnology Research, Rockville, Maryland 20850, USA.

## SI Text

### Details of molecular dynamics simulation.

**Table S1**, below, summarizes the system size, run length and composition of all the trajectories. There are three sets of simulations.

The first set is with a box size of  $\sim 12$  nm (all boxes are rhombic dodecahedrons, and the box size is the shortest distance between periodic images). In each denaturant concentration, for both urea and GdmCl, and in pure water solution,  $2 \mu\text{s}$  trajectories were generated. The last  $1.8 \mu\text{s}$  of each trajectory was used for further analysis. In addition, frames in which the minimum distance of the protein to its periodic image was less than  $3 \text{ \AA}$  were excluded.

The second set uses larger box sizes of  $\sim 15$  nm and  $\sim 17$  nm for urea concentrations of 5 M and 9 M, respectively. The initial configurations for these simulations were generated from the last frame of the 12 nm-box simulation, by adding extra denaturant and water molecules. Initial 200 ps runs were used for equilibration followed by 600 ns trajectories. Since the minimum distance to its periodic image for each frame is always larger than 0.8 nm and 2.8 nm for urea concentrations of 5 M and 9 M, respectively, all frames from the 600 ns simulations are used for further analysis.

The third set includes simulations at 1 M and 5 M urea with explicit dyes attached with a box size of  $\sim 12$  nm.  $2 \mu\text{s}$  trajectories were generated and the first 200 ns as equilibration. Frames in which the minimum distance of the protein to its periodic image was less than  $3 \text{ \AA}$  were excluded.

For the first two sets of simulations, the sequence of the protein used in SAXS experiments (with 79 residues) is used in simulations, whereas for the third set of simulations with explicit dyes, the sequence of the protein used in FRET experiments with 75 residues is used (**Table S2**).

### Details of calculating SAXS profiles from all-atom simulations.

For simulations with a box size of  $\sim 12$  nm, 18000 evenly spaced frames were selected from the  $1.8 \mu\text{s}$  equilibrated portion of the trajectories as described above. For simulations with box sizes of  $\sim 15$  nm and  $\sim 17$  nm simulation, 15000 frames were selected from the 600 ns trajectories. For each denaturant concentration, a background simulation of pure solvent was run with the same number of denaturant and ion molecules. Additional water molecules were added to keep the box size the same. 1000 evenly spaced frames were selected from each of the 100 ns background simulations to calculate the radial distribution functions of the bulk solvent.

As described in the Methods section of the main text, to calculate the SAXS intensity from all-atom data sets, the difference distribution functions of interparticle pair distances between the foreground and background simulations are calculated. Before obtaining the

difference of the distance distribution functions, a solvent electron density matching between the foreground bulk solvent region and the background solvent is performed. There are two parameters required to define the bulk solvent region in the foreground simulation, namely the inner and outer radii, defining two spheres. The volume between these is defined as bulk solvent. The outer radius also determines the maximum distance from the center of mass of protein atoms which are taken into account in the SAXS intensity calculation. Since we study unfolded proteins, for which there are always some configurations which are much more extended than the average, there can be several problems.

First, some atoms of the proteins may lie outside the sphere defined by the outer radius, causing underestimation of  $R_g$ . We investigated that by increasing the size of the sphere to construct the distance histogram. For each of the denaturant concentrations we performed the calculation, the  $R_g$  appears to converge to a stable value with increasing sphere size, at which point the contribution from protein atoms lying outside this sphere must be limited (Figure S8).

Second, some atoms of the proteins may intrude into the region of the foreground simulation defined as “bulk”, distorting the electron density of the bulk solvent. We addressed this problem when matching the solvent electron density between foreground and background simulation by only selecting frames in which all atoms of the proteins are within the inner sphere for this calculation. In this way, the scaling of the solvent electron density of the foreground is not affected by the expanded structure of unfolded proteins. However, in the next step of calculating distance histograms, all frames are used. All configurations sampled were weighted equally, i.e. there was no energetic bias toward experimental data employed.

Third, if too small a region of bulk solvent is defined in the foreground simulation, there will not be enough statistics to match the electron density of the background solvent. We investigated that by varying the inner radius as shown in Figure S8: when increasing the inner radius and consequently decreasing the region of the bulk solvent, the  $R_g$  from SAXS intensity is stable at the beginning and then start fluctuating rapidly because of the limited statistics. We always pick an inner radius before the  $R_g$  becomes unstable (Figure S8).

To compare with the experiment in the Guinier region ( $q < 0.04 \text{ \AA}^{-1}$ ), the simulated intensity is scaled by a factor to minimize the residues to the experimental intensities (this has no effect on the  $R_g$  extracted from Guinier fit). To compare with the experimental SAXS intensity in the full range of  $q$ , the simulated intensity is both scaled and shifted (i.e.  $I'(q) = aI(q) + b$ ) to take into account small differences in background.

## Error analysis

To estimate the errors of  $R_g$ , distance and FRET efficiency we use block averages with 5 blocks, i.e. for the simulations with the 12nm-box for which we have 18000 frames in the analysis as described above, the block size was 3600 frames. Note here we are not using the block transformation described in the paper by Flyvbjerg and Petersen<sup>1</sup>, but averages over a number of contiguous data blocks of equal size, where we vary the block size. To justify the choice of the block size, we have shown the errors of  $R_g$  and FRET efficiency when varying the block size in **Figure S17**. We obtain a similar shape of error curve as in the original literature when varying the number of block transformations, in which the errors are initially underestimated due to correlations between adjacent blocks, increase to a stable value before becoming noisy when only a few blocks are used. We always pick our block size to be at the beginning of the plateau in the shape of the error curve. The same block average method and block size is used to estimate the errors of the other quantities calculated from the trajectories.

One exception is the errors of explicit SAXS intensity calculation, in which a block average of 3 blocks is used to estimate the errors. The reason of doing that is because we need sufficient frames to match the bulk solvent between the foreground and background as described in the previous section.

## SI Tables

**Table S1.** System sizes and simulation lengths used for ACTR in urea and GdmCl. Trajectories for all systems were obtained initially in a 12 nm box. Additional runs in 15 and 17 nm boxes were performed for SAXS calculations in 5 M and 9 M urea, respectively. (*L*) indicates large box simulations referenced in the main text.

System	Solvent	Box Size (nm)	N <sub>denaturant</sub>	Length ( $\mu$ s)
Unlabeled ACTR	Water	12.28	0	2
Unlabeled ACTR	1.04 M Urea	11.96	761	2
	2.50 M Urea	11.96	1824	2
	5.00 M Urea	11.96	3642	2
	7.00 M Urea	11.97	5096	2
	9.01 M Urea	11.96	6560	2
	5.01 M Urea ( <i>L</i> )	14.90	7049	0.6
	9.02 M Urea ( <i>L</i> )	16.90	18507	0.6
Unlabeled ACTR	0.32 M GdmCl	11.96	233	2
	1.04 M GdmCl	11.96	758	2
	2.50 M GdmCl	11.96	1821	2
	4.99 M GdmCl	11.95	3635	2
Dye-labeled ACTR	1.00 M Urea	11.96	761	2
	5.00 M Urea	11.96	3642	2

**Table S2.** Sequences of two peptides simulated. In the simulations without dyes, the “BARE” sequence is used. The residues considered to attach the dyes when calculating FRET efficiency are shown in blue. In the simulations with dyes (using “DYE” sequence), positions of cysteines labeled with AlexaFlour 488 and AlexaFlour 594 are indicated by **C** and **C** respectively. Note that residue numbers are offset by 2 from those in Ref. <sup>2</sup>.

BARE:	1-GP	<b>S</b> GTQNRPL	LRNSLDDL	VG	PPSNLEGQSD	ER	<b>A</b> LLDQLHT-40
	41-LLSNTDATGL	EEIDRALGIP	<b>E</b> LVNQGQALE	PKQD	<b>S</b> GGPR		-79
DYES:	1-GP	<b>C</b> GTQNRPL	LRNSLDDL	VG	PPSNLEGQSD	ER	<b>A</b> LLDQLHT-40
	41-LLSNTDATGL	EEIDRALGIP	<b>E</b> LVNQGQALE	PKQD	<b>C</b>		-75

**Table S3.** Comparison between the errors of SAXS intensity in simulation and experiment. The ratio between the trace of the covariance matrix of 30 independent SAXS measurements and the sum of the squared errors from simulation (equal to the trace of the simulation covariance matrix) for each combination of denaturant concentration is shown. The ratio is always much larger than 1, showing that the errors of the simulation are negligible when calculating  $\chi^2$  between simulation and experimental SAXS intensity in different denaturant concentrations shown in Table 2 of the main text and Table S4. (*L*) indicates the large box simulations.

		$q < 0.04 \text{ \AA}^{-1}$			
		Experiment			
Simulation		<i>1.0 M</i>	<i>2.5 M</i>	<i>5.0 M</i>	<i>9.0 M</i>
	<i>1.0 M</i>	54.2	38.9	227.2	213.4
	<i>2.5 M</i>	80.1	73.3	497.1	343.8
	<i>5.0 M</i>	20.1	12.2	63.8	74.1
	<i>5.0 M (L)</i>	45.7	34.6	207.1	183.8
	<i>9.0 M</i>	153.3	139.2	937.1	656.4
	<i>9.0 M (L)</i>	5.4	3.7	19.4	21.2
		$q < 0.5 \text{ \AA}^{-1}$			
		Experiment			
Simulation		<i>1.0 M</i>	<i>2.5 M</i>	<i>5.0 M</i>	<i>9.0 M</i>
	<i>1.0 M</i>	68.1	37.0	201.4	248.4
	<i>2.5 M</i>	37.2	25.3	159.2	147.6
	<i>5.0 M</i>	30.2	13.9	67.3	104.6
	<i>5.0 M (L)</i>	37.3	23.0	134.2	143.1
	<i>9.0 M</i>	64.2	44.4	280.4	239.1
	<i>9.0 M (L)</i>	4.7	2.7	14.7	17.7

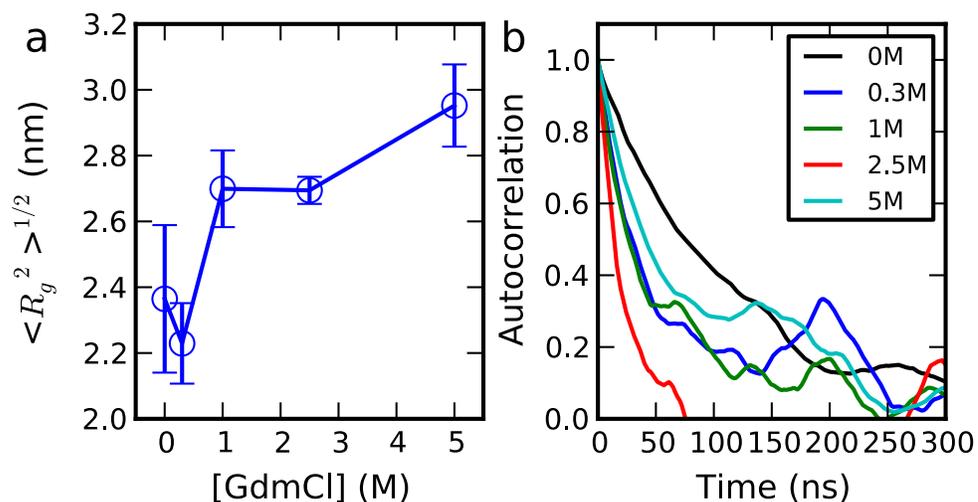
**Table S4.**  $\chi^2$  between the simulation and experimental observables for full range of SAXS data in urea, i.e.  $q < 0.5 \text{ \AA}^{-1}$ . Rows and columns correspond to simulated and experimental urea concentrations, respectively. (*L*) indicates the large box simulations.

		Experiment			
		<i>1.0 M</i>	<i>2.5 M</i>	<i>5.0 M</i>	<i>9.0 M</i>
Simulation	<i>1.0 M</i>	<b>0.48</b>	1.05	1.41	1.58
	<i>2.5 M</i>	0.49	<b>0.41</b>	0.85	0.88
	<i>5.0 M</i>	1.76	1.42	<b>0.74</b>	0.47
	<i>5.0 M (L)</i>	2.67	2.51	0.88	0.39
	<i>9.0 M</i>	3.18	3.24	0.98	0.38
	<i>9.0 M (L)</i>	1.56	2.69	0.84	<b>0.36</b>

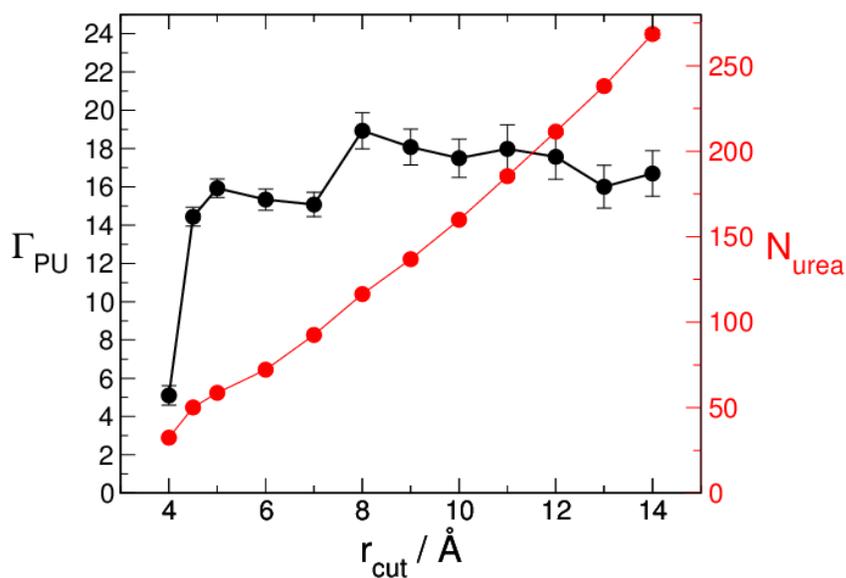
**Table S5.** Radii of gyration (in nm) from Guinier fitting scattering intensity using a continuum solvent model compared to that calculated using explicit solvent coordinates. The fitting range is  $q < 0.04 \text{ \AA}^{-1}$ . Numbers in brackets show the relative deviation of  $R_g$  from the one extracted from explicit solvent scattering intensity,  $|R_{g,\text{implicit}} - R_{g,\text{explicit}}|/R_{g,\text{explicit}}$ . (*L*) indicates the simulation with large box (**Table S1**).

	Explicit	Crysol			FOXS
		Hydration, urea contrast	Hydration, water contrast	No hydration, urea contrast	
[Urea] (M)	Figure 2, 6	Figure 2, 6	Figure S13	Figure S14	Figure S15
1	2.47	2.51 (2%)	2.51 (2%)	2.47 (0%)	2.47 (0%)
2.5	2.62	2.63 (0.4%)	2.64 (0.8%)	2.62 (0%)	2.59 (1%)
5	2.80	2.83 (1%)	2.83 (1%)	2.80 (0%)	2.79 (0.4%)
5 ( <i>L</i> )	2.89	2.84 (2%)	2.84 (2%)	2.90 (0.3%)	2.89 (0%)
9	2.95	3.03 (3%)	3.03 (3%)	2.95 (0%)	3.00 (2%)
9 ( <i>L</i> )	3.16	3.02 (4%)	3.01 (5%)	3.18 (0.6%)	3.03 (4%)

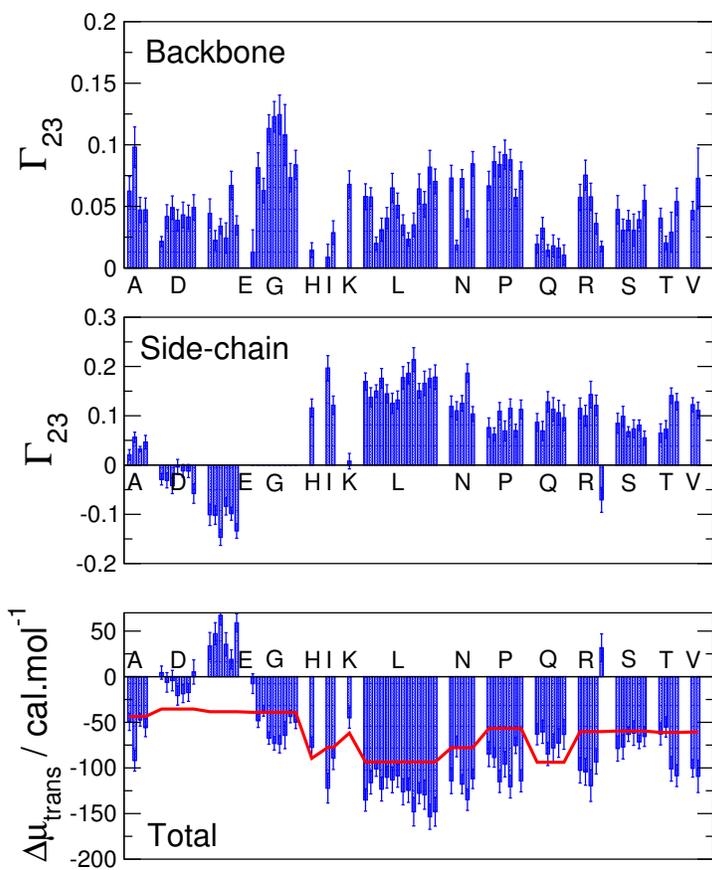
## SI Figures



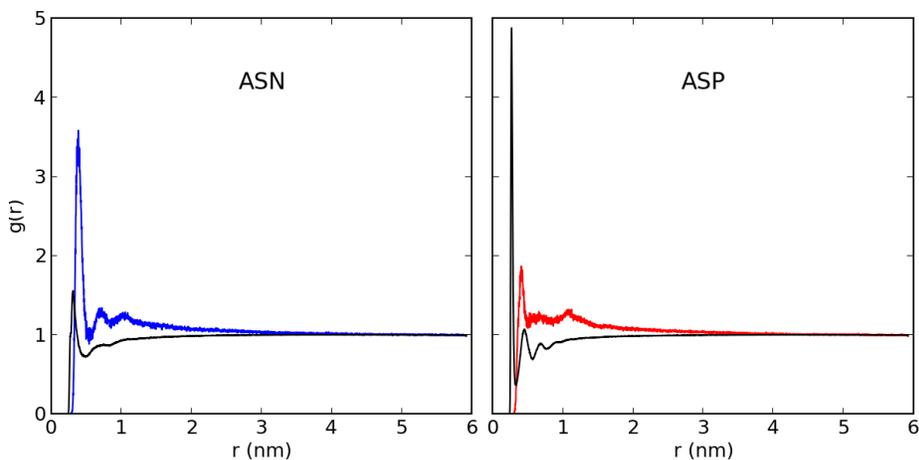
**Figure S1.** a) Dependence of  $R_g$  of ACTR on concentration of GdmCl. b) Time correlation function for  $R_g$ .



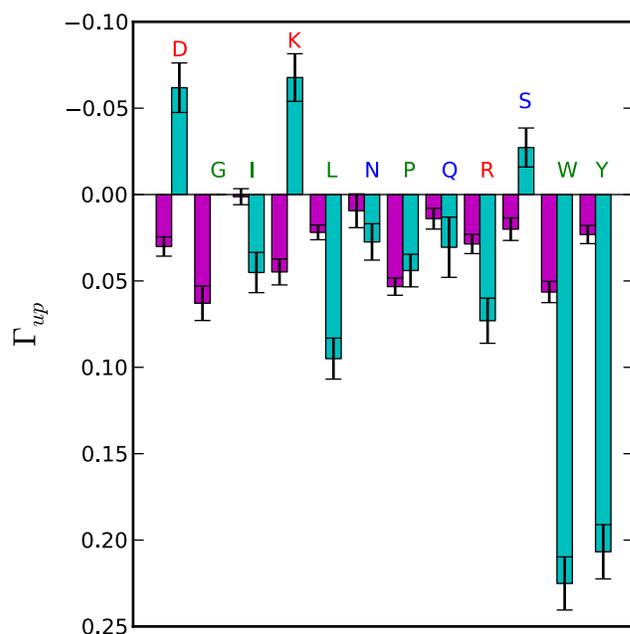
**Figure S2.** Dependence of total preferential interaction coefficient (i.e. for whole chain) on the cutoff distance between protein heavy atoms and water oxygens used to define the region near the protein.<sup>3</sup>



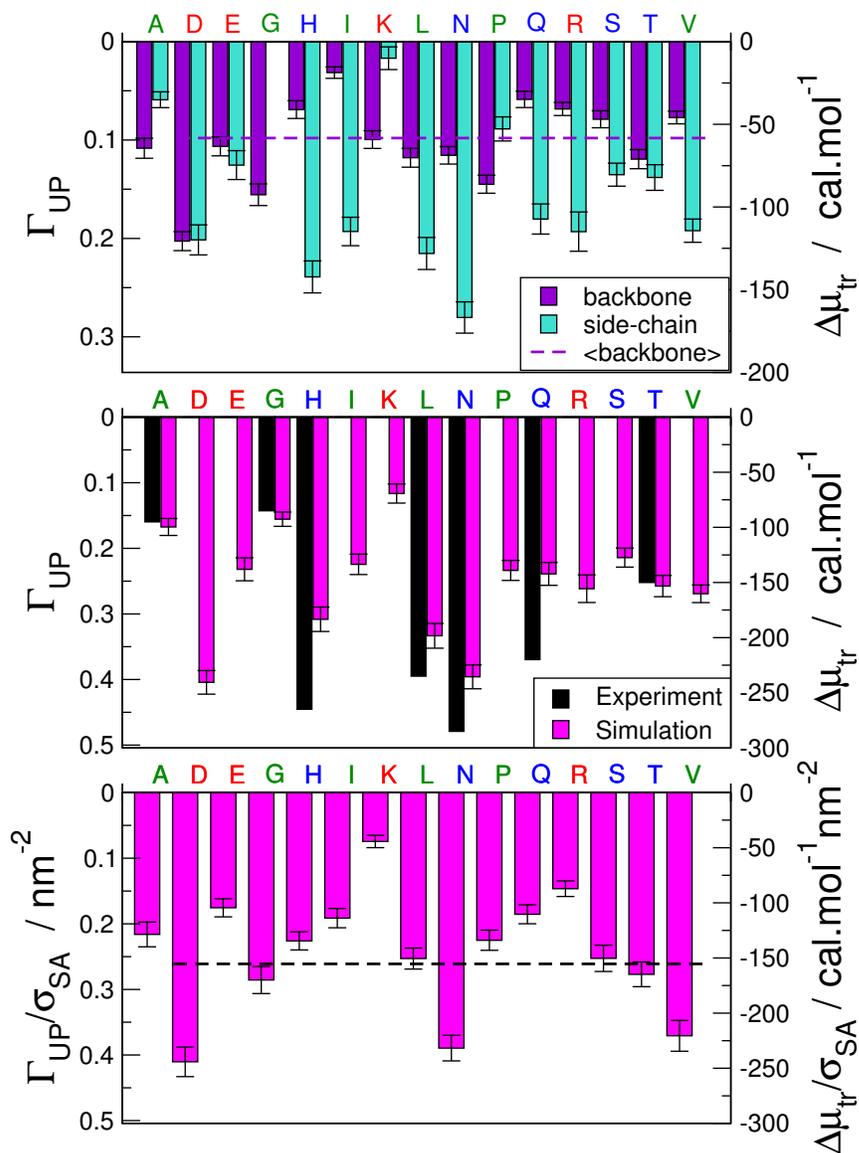
**Figure S3.** Preferential interaction coefficients (histograms) for urea for each residue of ACTR, grouped by residue type. Results are fairly insensitive to sequence context. Red curve in lower panel is experimental free energy transfer data from Auton and Bolen.<sup>4</sup>



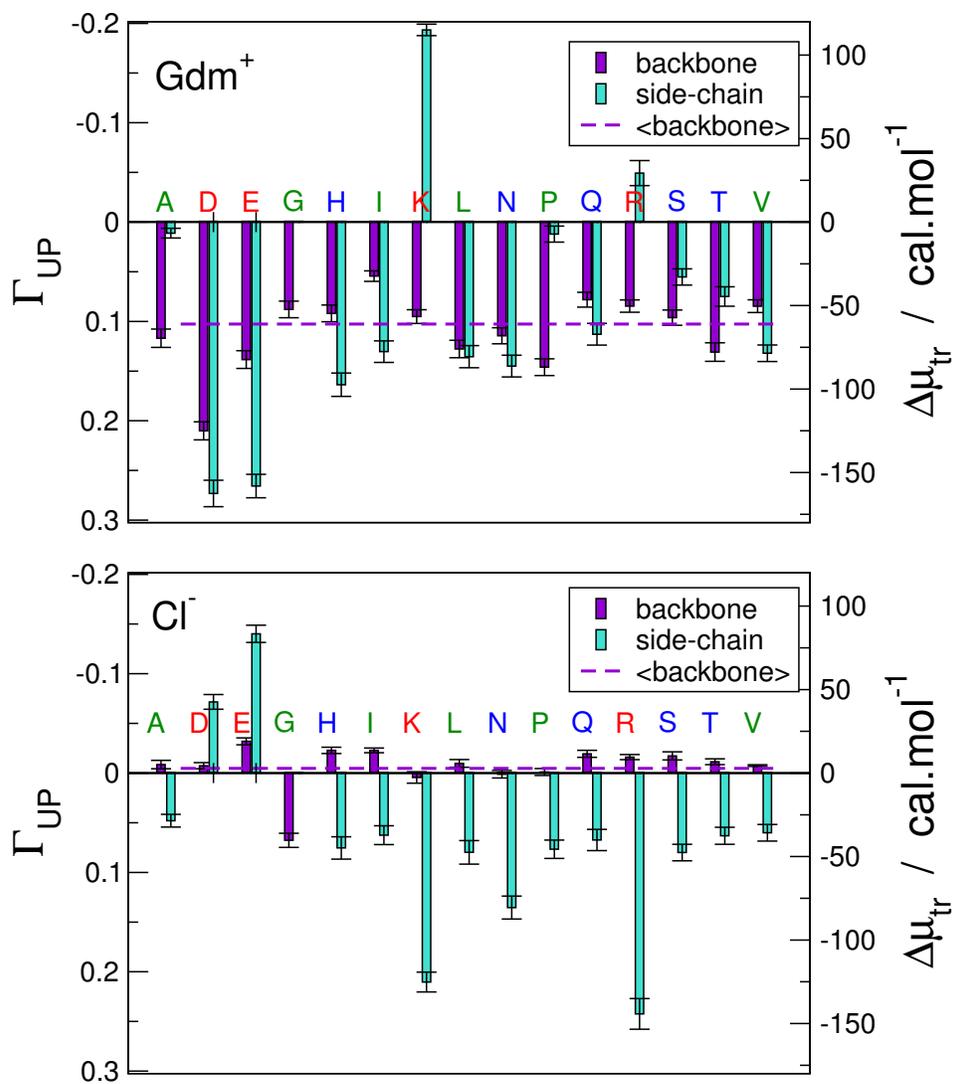
**Figure S4.** Pair distribution functions for water and urea around Asn and Asp side-chains. Black solid lines are RDF of center of mass of water to the two oxygen atoms in Asp side-chains or oxygen and nitrogen atoms in Asn side-chains, and coloured lines are RDF of center of mass of urea to these atoms in Asp or Asn side-chains.



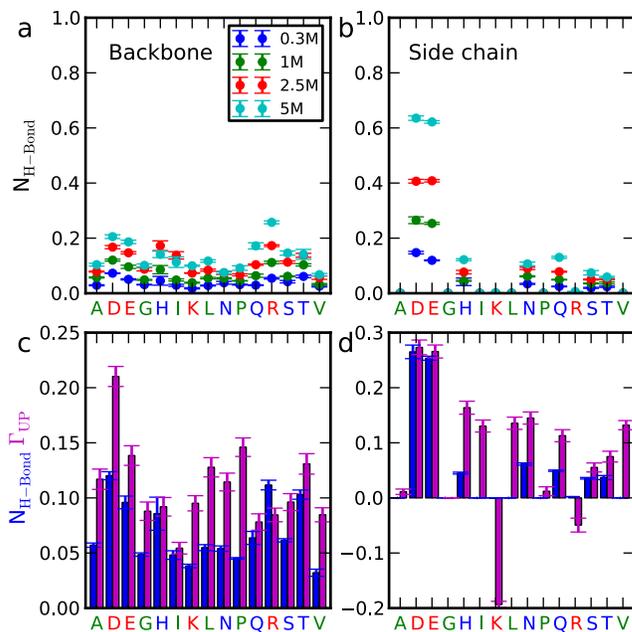
**Figure S5.** Preferential interaction coefficient calculated from the Trp cage simulation data in 3M urea concentration from a previous publication<sup>5</sup> (backbone: purple; side-chains: cyan). Interaction interaction coefficients have been multiplied by a factor 1/3 to place them on a similar scale to those in the main text, computed from simulations in 1 M urea, including only unfolded conformations at 300 K (unfolded state is defined when the fraction of native contacts is smaller than 0.63 as described in the original literature<sup>5</sup>). Here we show that the two amino acids Trp and Tyr which are not present in the ACTR protein sequence have much larger preferential coefficients than the other residues, consistent with the much more favorable transfer free energy of these two amino acids in experiment<sup>4</sup>. Unfortunately, we cannot directly compute the transfer free energy because the linear relation between preferential interaction coefficient and transfer free energy may not be valid at this denaturant concentration.



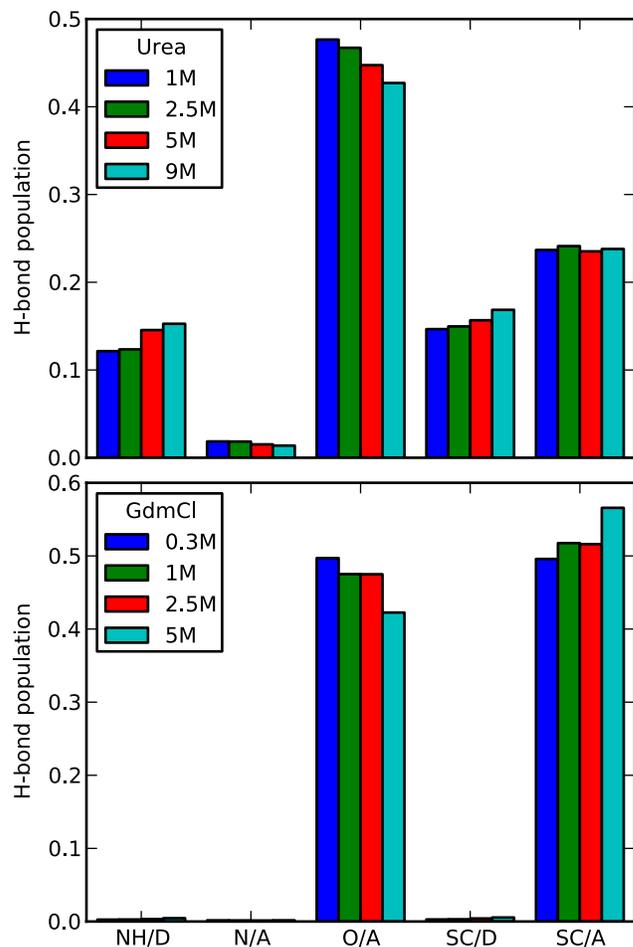
**Figure S6.** Preferential interaction coefficients and transfer free energies to guanidinium chloride as shown for urea in main text. The preferential interaction coefficient is reported as the sum of those for the  $Gdm^+$  and  $Cl^-$  ions (see **Figure S7** for a breakdown into  $Gdm^+$  and  $Cl^-$  contributions). Experimental data taken from Nozaki and Tanford 1970<sup>6</sup>. Missing bars for experiment are due to lack of data for those residues.



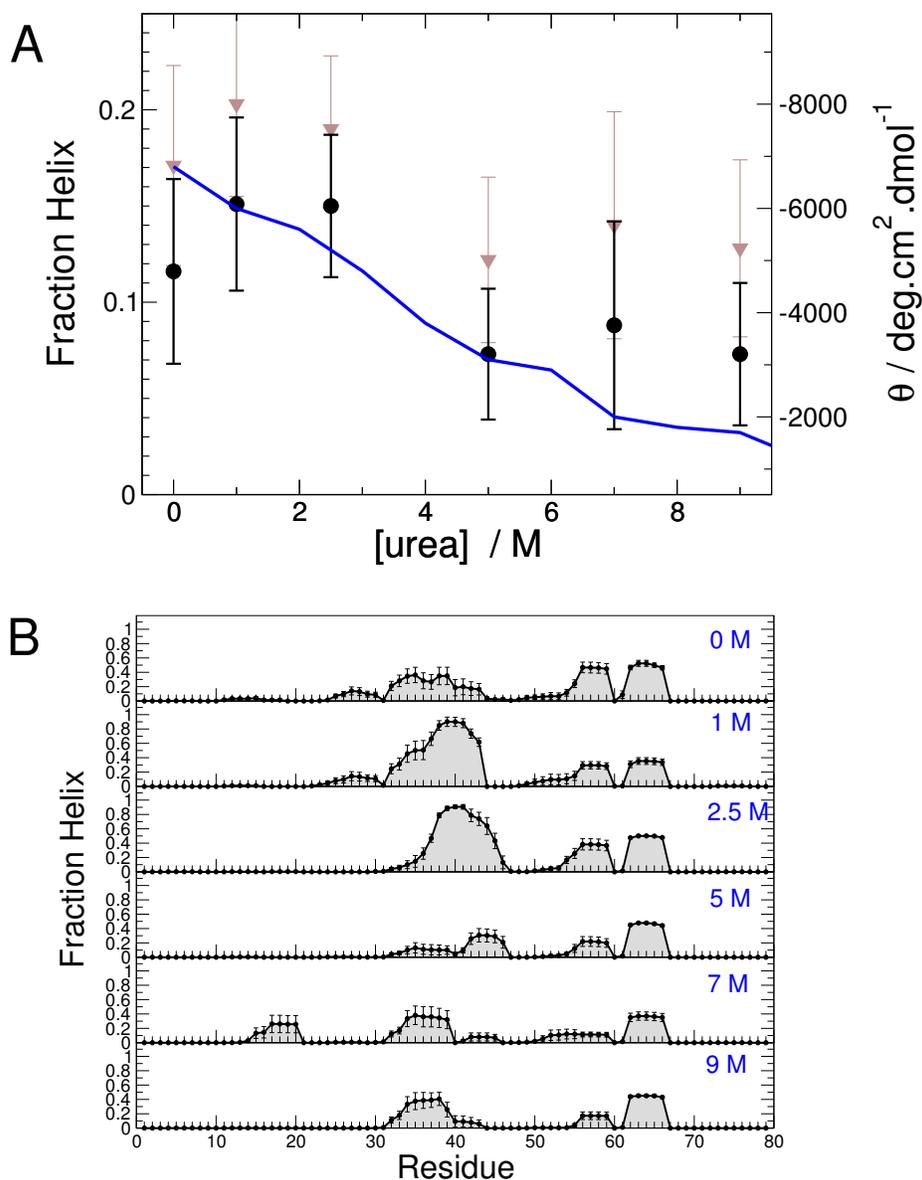
**Figure S7.** Transfer free energies for GdmCl broken down into contributions from Gdm<sup>+</sup> and Cl<sup>-</sup> ions. Broken lines indicate an average for the backbone contribution for all uncharged residues.



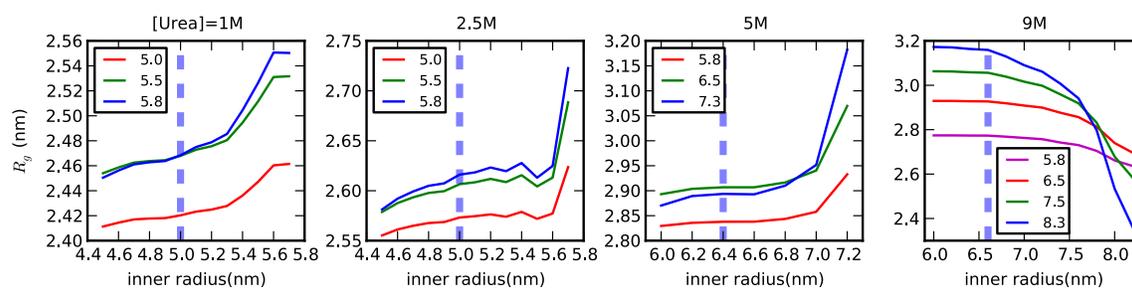
**Figure S8.** Hydrogen bonding between guanidinium and protein. In (a) and (b) are shown the number of hydrogen bonds per guanidinium molecule to the backbone and side-chains respectively, averaged by residue type, at different GdmCl concentrations. In (c) and (d) we compare the average number of hydrogen bonds with the preferential interaction coefficients, a measure of the number of excess guanidinium molecules in the vicinity of each group, relative to bulk.



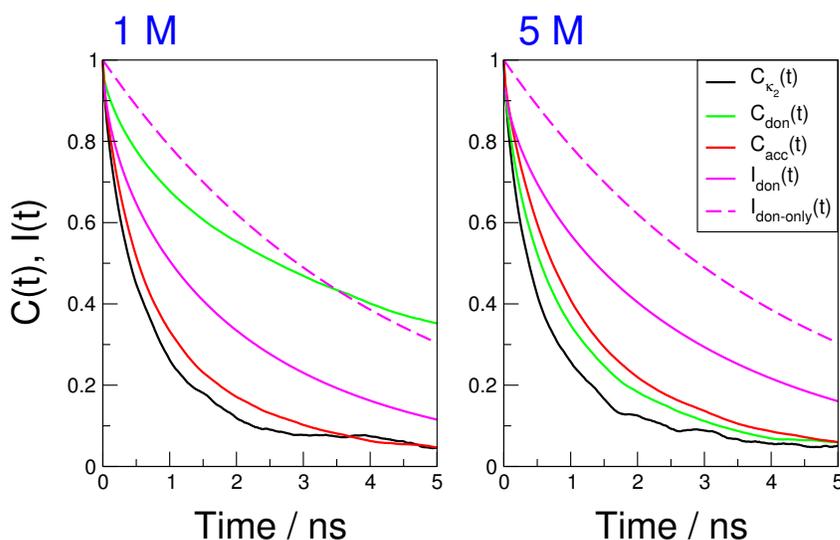
**Figure S9.** Population of different types of hydrogen bond. NH/D indicates hydrogen bond formed with backbone NH group as the donor; N/A backbone nitrogen atom as the acceptor; O/A backbone oxygen atom of the carboxyl group as the acceptor; SC/D side chain as the donor; and SC/A side chain as the acceptor. Since there is no carboxyl group in the guanidinium ion compared to the urea molecule, it has a much lower chance to form a hydrogen bond with both the backbone and side-chain of the protein.



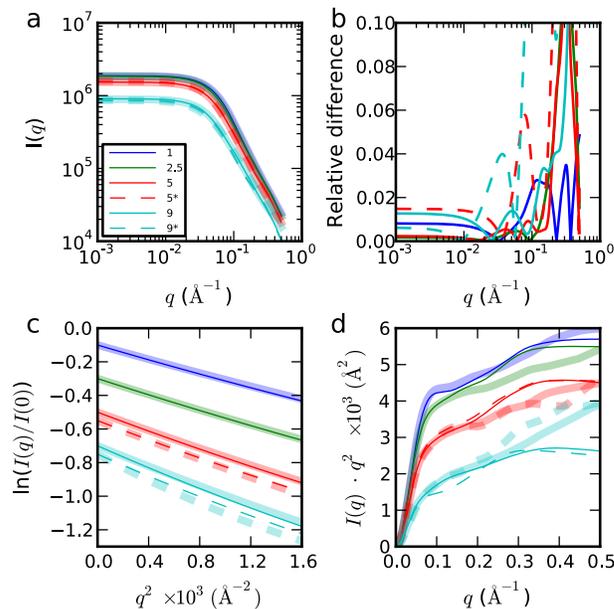
**Figure S10.** Dependence of helix fraction on urea concentration. (A) Sequence-averaged fraction helix. Black filled circles were computed from simulation using DSSP<sup>7</sup> (left axis), counting the fraction of residues assigned as  $\alpha$ -helix, while blue line shows the experimental 222 nm CD mean residue ellipticity data  $\theta$ .<sup>8</sup> (right axis). The left and right scales were aligned assuming ideal coil and helical values of molar ellipticity of  $\theta_c = -510$  and  $\theta_h = -37425$  deg.cm<sup>2</sup>.dmol<sup>-1</sup> respectively, based on literature data<sup>9</sup>. Some caution is needed in interpreting the CD data, because non-helical contributions have been suggested to be significant for this protein<sup>10</sup>. The brown symbols were determined by including  $3_{10}$  helix as well as  $\alpha$ -helix by DSSP in the helix fraction. (B) Per-residue fraction helix (considering only  $\alpha$ -helix) in each simulation, from DSSP. Differences at the residue level should not be over-interpreted given that the length of the simulations is comparable to typical relation times associated with helix formation/breaking (microseconds).<sup>11</sup>



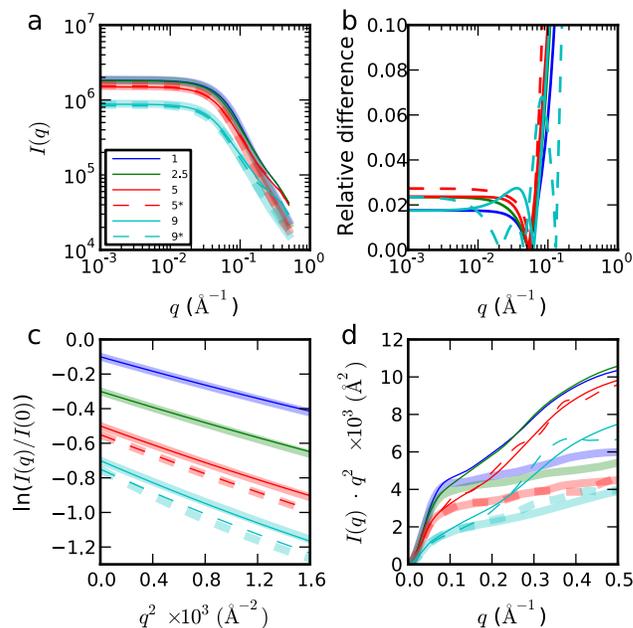
**Figure S11.** Dependence of  $R_g$  on the inner and outer radius defining the background region via Guinier fit within  $q < 0.04 \text{ \AA}^{-1}$ . Legend shows the outer radius tested and blue dashed lines the selected inner radii. The outer radius is always selected to be  $2 \text{ \AA}$  smaller than the box size due to the boundary effect.



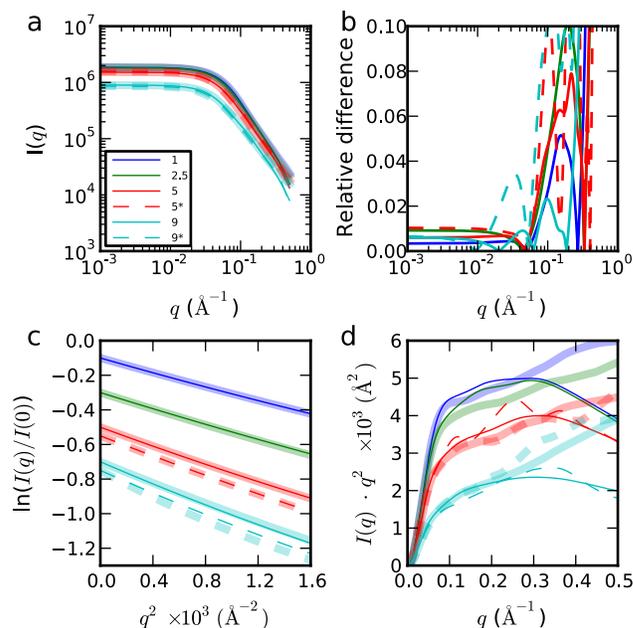
**Figure S12.** Decay of donor fluorescence intensity  $I(t)$  and autocorrelation functions  $C(t)$  of donor, acceptor and  $\kappa^2$  from simulations of ACTR labeled with Alexa 488 and Alexa 594 in 1 M and 5 M urea. Donor and acceptor correlation functions are  $P_2$  autocorrelations of a vector  $\mu$  defined with respect to the largest conjugated ring system of each chromophore, i.e.  $C(t) = \langle P_2[\mu(0) \cdot \mu(t)] \rangle$ , where  $P_2$  is the second order Legendre polynomial. The intrinsic decay of the donor fluorescence (taken as 4.2 ns) is shown as a broken line.



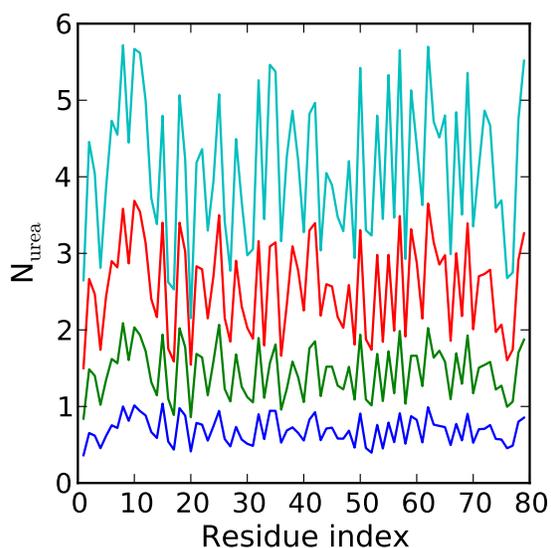
**Figure S13.** Comparison between Crysol *with fixed solvent density* (water density) and explicit solvent calculation of SAXS intensity in different urea concentration. Color schemes of figures are the same as Figure 6 in the main text. We obtain essentially identical results either from all-atom or Crysol calculation with fixed solvent density at low scattering angles, showing that the background contrast does not influence  $R_g$  (Table S5).



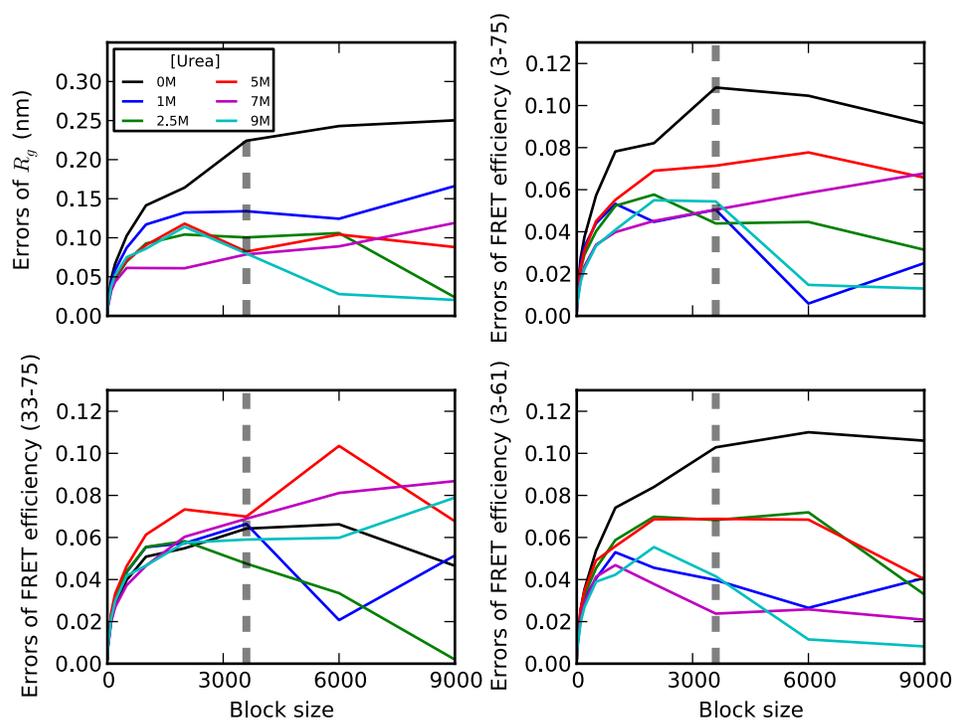
**Figure S14.** Comparison between Crysol *without hydration shell* and explicit solvent calculation of SAXS intensity in different urea concentration. Color schemes of figures are the same as Figure 6 in the main text. We obtain large deviations starting from  $q < 0.1 \text{ \AA}^{-1}$ , showing hydration shell structuring influences low scattering angles. However the errors are within 3% in the Guinier region (Table S5).



**Figure S15.** Comparison between FOXS and explicit solvent calculation of SAXS intensity. Default parameters of FOXS are used, which means no background adjustment is introduced. Color schemes of figures are the same as Figure 6 in the main text. We obtain essentially identical results either from all-atom or FOXS calculation (**Table S5**). Color schemes of figures are the same as Figure 6 in the main text.



**Figure S16:** Distribution of urea molecules along the protein chain. The urea molecule is defined as being in contact with protein residues if the all-atom minimum distance between them is smaller than 4.5 Å. Color schemes of figures are the same as Figure 6 in the main text.



**Figure S17:** Errors of  $R_g$  and FRET efficiency as a function of block size using the block average method<sup>1</sup>. The black dash lines show our choice of block size, which is usually at the beginning of the plateau for the errors.

## References

1. Flyvbjerg, H.; Petersen, H. G., Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91*, 461-466.
2. Borgia, A.; Zheng, W.; Buholzer, K.; Borgia, M. B.; Schuler, A.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; Best, R. B.; Schuler, B., Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *Submitted* **2016**.
3. Baynes, B. M.; Trout, B. L., Proteins in mixed solvents: a molecular-level perspective. *J. Phys. Chem. B* **2003**, *107*, 14058-14067.
4. Auton, M.; Bolen, D. W., Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15065-15068.
5. Zheng, W.; Borgia, A.; Borgia, M. B.; Schuler, B.; Best, R. B., Empirical optimization of interactions between proteins and chemical denaturants in molecular simulations. *J. Chem. Theor. Comput.* **2015**, *11*, 5543-5553.
6. Nozaki, Y.; Tanford, C., The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **1970**, *245*, 1648-1652.
7. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577-2637.
8. Demarest, S. J.; Martinez-Yamout, M.; Chung, J.; Chen, H.; Xu, W.; Dyson, H. J.; Evans, R. M.; Wright, P. E., Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators. *Nature* **2002**, *415*, 549-553.
9. Scholtz, J. M.; Marqusee, S.; Baldwin, R. L.; York, E. J.; Stewart, J. M.; Bolen, D. W., Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 2854-2858.
10. Kjaergaard, M.; Nørholm, A.-B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B., Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein Sci.* **2010**, *19*, 1555-1564.
11. Thompson, P. A.; Muñoz, V.; Jas, G. S.; Henry, E. R.; Eaton, W. A.; Hofrichter, J., The helix-coil kinetics of a heteropeptide. *J. Phys. Chem. B* **2000**, *104* (2), 378-389.