Understanding the Structural Ensembles of a Highly Extended Disordered Protein†

Gary W. Daughdrill[a], **Stepan Kashtanov**[b], **Amber Stancik**[c], **Shannon E. Hill**[d], **Gregory Helms**[e], **Martin Muschol**[d], **Véronique Receveur-Bréchot**[f], and **F. Marty Ytreberg**[b]

Véronique Receveur-Bréchot: veronique.brechot@ibsm.cnrs-mrs.fr; F. Marty Ytreberg: ytreberg@uidaho.edu

[a]Department of Cell Biology, Microbiology, and Molecular Biology and the Center for Drug Discovery and Innovation, University of South Florida, 3720 Spectrum Blvd., Suite 321, Tampa, FL 33612

[b]Department of Physics, University of Idaho, Engineering and Physics Building, Rm. 333, Moscow, ID 83844-0903

[c]Department of Biological Sciences, University of Idaho, Life Science Building, Rm. 252, Moscow, ID 83844-3051

[d]Department of Physics, University of South Florida, 4202 East Fowler Ave, Physics Building, Rm. 114, Tampa, FL 33620

[e]Department of Chemistry, Washington State University, Fulmer Hall, B3, Pullman, WA 99164-4630

[f]IMR – CNRS, UPR3243, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

Abstract

Developing a comprehensive description of the equilibrium structural ensembles for intrinsically disordered proteins (IDPs) is essential to understanding their function. The p53 transactivation domain (p53TAD) is an IDP that interacts with multiple protein partners and contains numerous phosphorylation sites. Multiple techniques were used to investigate the equilibrium structural ensemble of p53TAD in its native and chemically unfolded states. The results from these experiments show that the native state of p53TAD has dimensions similar to a classical random coil while the chemically unfolded state is more extended. To investigate the molecular properties responsible for this behavior, a novel algorithm that generates diverse and unbiased structural ensembles of IDPs was developed. This algorithm was used to generate a large pool of plausible p53TAD structures that were reweighted to identify a subset of structures with the best fit to small angle X-ray scattering data. High weight structures in the native state ensemble show features that are localized to protein binding sites and regions with high proline content. The features localized to the protein binding sites are mostly eliminated in the chemically unfolded ensemble; while, the regions with high proline content remain relatively unaffected. Data from NMR experiments support these results, showing that residues from the protein binding sites experience larger environmental changes upon unfolding by urea than regions with high proline content. This behavior is consistent with the urea-induced exposure of nonpolar and aromatic side-chains in the protein binding sites that are partially excluded from solvent in the native state ensemble.

Introduction

It is well established that intrinsically disordered proteins (IDPs) have fewer stabilizing intramolecular interactions and are more dynamic than ordered proteins.[1–13] This is because they do not contain a high enough fraction of nonpolar residues to form a hydrophobic core.[4,5,14,15] However, many IDPs do exhibit some degree of collapse relative to a classical random coil and in general IDPs have a broader range of compactness as a function of polymer length when compared to chemically unfolded proteins.[4,16] It is necessary to understand the forces responsible for this behavior before a comprehensive picture of the ensemble structure of IDPs can be developed.

A random polymer model has been successfully used to describe the hydrodynamic dimensions of chemically unfolded proteins that are ordered in their native state.[17–21] A simple power law relationship can be used to predict the radius of gyration (R_g) for such proteins based solely on polymer length.[17,18] There is a growing consensus that this model cannot be accurately applied to IDPs because of the obvious compositional and physicochemical differences between the chemically unfolded states of ordered proteins and the native states of IDPs. For instance, IDPs generally have a higher net charge and proline content than most ordered proteins and it is likely that these differences will play a prominent role in defining their hydrodynamic dimensions. This was recently investigated by Forman-Kay and colleagues where they showed that accounting for proline content and overall net charge improved the prediction of the hydrodynamic dimensions for IDPs.[16] When these attributes are taken into account many IDPs are more compact than expected for a classical random coil.

If IDPs are not random coils then what are they? Several groups have used a variety of techniques to investigate the structural ensembles of IDPs but a general approach to characterize and classify these structural ensembles has not emerged.[3,10,22–32] Only a handful of IDP structural ensembles have been determined so it is too early to tell if they are providing a robust and realistic representation of the equilibrium ensemble. However, there are some notable features that the experimentally determined structural ensembles of some IDPs share. For instance, short segments of transient helical secondary structure are commonly observed, and these short helical segments often correspond to protein binding sites.[9,33–35] Transient long-range contacts have also been observed for a few IDPs.[24,31,32,36,37]

Most investigations of IDP structural ensembles rely heavily on nuclear magnetic resonance (NMR) spectroscopy because of its ability to provide atomic level information on IDP structure and dynamics. Small angle x-ray scattering (SAXS) is also an important tool because it provides a comprehensive, low-resolution picture of the equilibrium ensemble. SAXS is also the premiere solution state tool for investigating protein compactness. Defining the degree of compactness for IDPs is an important step that can be taken to understand their equilibrium structural ensembles. In particular, understanding how the relative compactness of an IDP is affected by chemical denaturants will help identify the forces that contribute to their marginal stability.

The transactivation domain of the tumor suppressor protein, p53, was selected as a disordered model system for this study. Residues 1–73 of p53 form a transactivation domain (p53TAD) that is responsible for regulating its transcriptional activity and cellular stability.[38–40] This domain contains binding sites for the ubiquitin ligase, MDM2, and the 70 kDa subunit of replication protein A, RPA70. This domain also contains thirteen prolines and has a net negative charge of −14e. When bound to MDM2, p53 becomes ubiquinated and targeted for proteosome-mediated degradation. When bound to RPA70, p53 may be

stabilized and available to amplify the cellular response to DNA damage.[41] Several studies have shown that human p53TAD is intrinsically disordered but there is no consensus description of its equilibrium structural ensemble.[31,32,41–45] One model of the p53TAD structural ensemble was based on distance restraints from paramagnetic relaxation enhancement (PRE) experiments and another was based on a combination of SAXS data and angular restraints from residual dipolar coupling (RDC) experiments[31,32,45]. The model based on the PRE restraints was probably more compact than the true equilibrium ensemble but it did reliably identify an interesting asymmetry in the distribution of negative charge that helped rationalize the recruitment of basic factors to the site of transcription. The model based on SAXS and RDC restraints was less compact than the PRE model but the backbone dihedral angles of the structures in that ensemble were biased toward values observed in the coil and loop regions of ordered proteins.

In this study, SAXS, size exclusion chromatography (SEC), and dynamic light scattering (DLS) were used to show that the radius of gyration and the hydrodynamic radius of p53TAD significantly increase when the protein is chemically unfolded with urea. This increase in the hydrodynamic dimensions of chemically unfolded p53TAD means that the native state of p53TAD has an inherent degree of compactness that is disrupted in the presence of urea. This hypotheses is supported by results from a new protocol we developed to generate structural ensembles of IDPs. Structures with the largest contribution to fitting the SAXS data show persistent features in the native state of p53TAD that are eliminated in the unfolded state.

Materials and Methods

Preparation of p53TAD samples

Samples of p53TAD, residues 1–73 of human p53 with the R72 polymorphism (molecular mass 8205 Da), were prepared as previously described.[41] In summary, a pET28a clone containing the cDNA for p53TAD was overexpressed in *Escherichia coli* BL21(DE3) cells for six hours. The cells were pelleted, resuspended in a buffer containing protease inhibitors, and lysed using a French press. The clarified lysate was loaded onto a column containing Ni-NTA Superflow resin and eluted with a buffer containing 300 mM imidazole. Fractions containing p53TAD were combined and dialyzed into a buffer suitable for SEC. The Histidine tag was cleaved using thrombin and the solution was loaded onto a HiLoad 16/60 Superdex 75 prep grade column (Pharmacia, 17 1068-01). The p53TAD protein was eluted at a flow rate of 1.0 ml/min. The fractions containing p53TAD were pooled, dialyzed into an appropriate buffer for SAXS, SEC, or DLS, and concentrated to 2–10 mg/ml.

SAXS measurements of p53TAD

SAXS experiments were performed on ID02 beamline at ESRF synchrotron, Grenoble, France, as previously described.[46] The sample-to-detector distance was set at 1.5 m, and the wavelength (λ) at 1.0 Å, giving access to scattering vectors q ranging from 0.018 to 0.39 Å^{-1}, where $q = 4\pi\sin\theta/\lambda$, and θ is the scattering angle. Ten successive frames of 0.5 seconds were collected for each sample. Between each frame the protein solution was circulated through an evacuated quartz capillary. Measurements on the buffer solution at the same urea concentration as the protein were performed before and after each protein sample.

The individual frames were averaged if no radiation damage or bubble formation was observed. The buffer signal was then subtracted from the protein scattering curves after proper normalization and correction from detector response. All the experiments were conducted at 20°C. SAXS data was collected on p53TAD samples at 4 mg/ml and at 10 mg/ml, in the presence of 0, 2, 4, 6, and 8 M urea. The initial buffer solution was 50 mM

NaH_2PO_4 (pH 6.5), 50 mM NaCl, 1 mM EDTA, 0.02% NaN_3. Data at 4 mg/ml protein and 4 M urea had to be discarded because of the presence of a bubble in the capillary during data acquisition.

The R_g was inferred for p53TAD, at all protein and urea concentrations, from the Guinier law, in the appropriate q-range (qR_g 1) : $I(q) = I(0) \exp(-q^2 R_g^2/3)$, where $I(q)$ is the scattering intensity and $I(0)$ is the forward scattering intensity[47]. The 4 mg/ml data was used for subsequent analysis because it was less affected by interparticle repulsion. The distance distribution function $P(r)$ was calculated from the scattered intensity by indirect Fourier-transform using the program GNOM.[48]

The SAXS data for p53TAD was also modeled as a Kratky-Porod chain, using the following relationship[49]:

$$I(q)/I(0) = 2/x^2 \, (x - 1 + e^{-x}) + b/L \, [4/15 + 7/15x - (11/15 + 7/15x) \, e^{-x}] \quad (1)$$

where $x = q^2 Lb/6$, L is the contour length, and b is the Kuhn statistical length, which is twice the persistence length. This equation is valid for $L/b > 10$ and in the q-range $q < 3/b$. The persistence length is a measure of the rigidity of the polypeptide and is defined for polymers as the length over which the polymer naturally stays straight. A higher persistence length reveals a higher rigidity of the polymer. The persistence length therefore accounts for the rigidity of the polypeptide, and for local interactions. The contour length is the length of the fully extended chain. The contour length of a fully unfolded protein cannot exceed the product $n \cdot l_0 \cdot f$, where n is the number of residues, l_0 is the distance between two consecutive residues ($l_0 = 3.8$ Å), and f is a geometrical factor equal to 0.95 for a polypeptide chain. When the SAXS data for p53TAD was fit using the Krakty-Porod model, L values larger than the upper bound were observed. Since this is not physically possible, the value of L was constrained to not exceed the upper bound. It is unclear why L values larger than the upper bound were observed but the perfect linearity of the Guinier plots and the presence of repulsive interactions based on a systematic decrease in R_g as the protein concentration is increased argues strongly against aggregation. The large values of L might be due to eqn (1) that is mathematically more sensitive to the product $L*b$ included in the adjustable parameter $x = q^2 Lb/6$, than to each parameter L and b individually, leading to a possibility of a correct $L*b$ product with too large a value of L.

Size Exclusion Chromatography

Hydrodynamic radii for p53TAD were determined using SEC. A prepacked 16/60 Superdex 75 column (GE Healthcare, Piscataway New Jersey) with a V_o of 39.81 ml was attached to an AKTA FPLC system (GE Healthcare, Piscataway New Jersey) and equilibrated with five column volumes of 50 mM TrisHCl (pH 7.5) and 300 mM KCl. Molecular weight standards were purchased from Sigma (MW-GF-70) and included Aprotinin, Bovine Serum Albumin, Carbonic Anhydrase, and Cytochrome C. V_e was determined for each of the standards at a flow rate of 0.14 ml/min. A standard curve was generated by plotting log(MW) for the standards versus V_e/V_o. Least squares fitting of the data to a linear function gave a standard equation with a correlation coefficient of 0.997. V_e was also determined for p53TAD at 0, 2, 4, 6, and 8 M urea. Three replicate experiments were run at each concentration of urea. The elution volume, measured at the point of maximum absorbance in the elution profile, never deviated by more than ±0.4 ml at a given urea concentration, which corresponds to a change in the Stokes radius (R_h) of less than 0.4 Å. The V_e values were averaged and used to calculate an apparent molecular weight based on the standard curve and this was used to calculate the R_h based on an empirical relationship between the two quantities.[50] The elution volumes of blue dextran and acetone were measured at all urea concentrations to determine whether the permeation properties of the Superdex 75 changed in the presence of urea.[50]

The elution volumes of blue dextran and acetone did not change significantly when the concentration of the urea was changed (data not shown).

Static and Dynamic Light Scattering Measurements

Both static (SLS) and dynamic (DLS) light scattering measurements were carried out using a Zetasizer Nano S (Malvern Instruments, Worcestershire, UK) with a 4 mW He-Ne laser at λ = 633 nm. The unit collects back-scattered light at an angle of θ = 173°. Sample temperature during measurements, controlled by the built-in Peltier element, was 25 ± 0.1 °C. Stock solutions of human p53TAD were sequentially filtered through 220 nm and 20 nm pore size syringe filters. The resulting samples had no large aggregates and the polydispersity obtained from DLS was less than 0.1. Actual p53TAD concentrations where obtained from uv absorption measurements at λ = 280 nm (a_{280}nm = 1.364 ml mg^{-1} cm^{-1}). Scattering intensities and autocorrelation functions were determined from the average of eight to ten correlation functions, with a typical acquisition time of 180 s per correlation function. A viscosity value of η = 0.9026 mPa·s was used for the 50 mM NaH_2PO_4/50 mM NaCl buffer. This value accounts for the contributions of both salts to the water viscosity.[51]

Static Light Scattering Analysis

SLS measures the excess scattering due to the dissolved protein. This value is obtained by comparison against a standard of known scattering cross section (in our case, toluene), and is expressed as Rayleigh ratio R_θ.

$$R_\theta = [(I_{tot} - I_{sol})/I_{tol}]\,[n/n_{tol}]^2\,R_{\theta,tol} \quad (2)$$

where I_{tot}, I_{sol}, and I_{tol} are the measured scattering intensity of the protein solution, the salt/buffer background, and of the toluene standard, respectively. $R_{\theta,tol}$ is the Rayleigh ratio for toluene at λ = 633 nm and the ratio n/n_{tol} accounts for the difference in scattering volume imaged onto the detector due to the refractive index differences between the aqueous solvent and toluene. For our set-up, the manufacturer quotes a Rayleigh ratio of $R_{tol} = 13.52 \times 10^{-6}$ cm^{-1}. The normalized Rayleigh ratio R_θ can be related to the properties of the protein in solution via

$$KC_p/R_\theta = 1/M + 2\,B_{22}\,C_p \quad (3)$$

where KC_p/R_θ is the Debye Ratio, M is the molecular weight of the protein, C_p is the protein concentration in mg/ml, and B_{22} is the virial coefficient, which accounts for the net effect of all intermolecular protein interactions on the thermal fluctuations. The ellipsis indicates that non-linear terms can become important at higher protein concentrations. The instrument constant K in eqn (3) is given by

$$K = (2\pi^2\,n_0{}^2/N_A\,\lambda_0{}^4)\,(dn_0/dC_p)^2 \quad (4)$$

where n_0 is the solvent's refractive index, N_A is Avogadro's number, λ_0 is the wavelength of incident light, and $(dn_0/dC_p)_\lambda$ is the refractive index increment of the solvent due to p53TAD. We used the value of $(dn_0/dC_p)_\lambda$ = 0.185 which is typical for many proteins.[52] The molecular weight of p53TAD obtained from eqn (2) was 8.1 ± 0.2 kDa.

Dynamic Light Scattering Analysis

The autocorrelation function of scattered light measured in DLS yields the decay rates, G, of local concentration fluctuations for macromolecules in solution.[53–55] For data analysis, the experimentally measured (and normalized) autocorrelation function of intensity fluctuations $g_2(\tau)$ is first converted into the field autocorrelation function $g_1(\tau)$ via the Siegert relation.[54]

$$g_1(\tau)=\sqrt{[g_2(\tau)-1]} \quad (5)$$

For the essentially homogeneous distributions of monomeric protein molecules we are concerned with, the field autocorrelation function $g_1(\tau)$ will decay with a single rate

$$\Gamma=D_m q^2, \quad (6)$$

where D_m is the mutual diffusion coefficient and q is the magnitude of the scattering wave vector given by

$$q=(4\pi n_0/\lambda_0)\sin(\theta/2) \quad (7)$$

Here, n_0 is the solution's refractive index, λ_0 is the wavelength of the incident laser in air, and θ is the in-plane angle at which the scattered light is detected. Since all measurements are performed at elevated protein concentrations, both direct (e.g. electrostatic, dipole-dipole, van der Waals, hydrophobic interactions) and solvent-mediated hydrodynamic interactions among the protein molecules will alter the decay rates compared to purely thermally driven concentration fluctuations.[56–58] At moderate protein concentrations, the mutual diffusion coefficient D_m is related to the single particle diffusivity D_0 via

$$D_m=D_0[1+k_D\varphi] \quad (8)$$

where k_D is the sum of direct and hydrodynamic protein interactions, φ is the protein volume fraction, and D_0 is the single-particle diffusivity of the protein given by the Stokes-Einstein relation

$$D_0=k_B T/(6\pi\eta R_H) \quad (9)$$

Here, k_B is the Boltzmann constant, T is the absolute temperature, $\eta = \eta(Cs,T)$ is the (salt- and temperature-dependent) solution viscosity, and R_h is the hydrodynamic radius.

NMR of unfolded p53TAD

^{1}H-^{15}N HSQC experiments were performed on uniformly ^{15}N-labeled p53TAD samples in 0-6 M urea at 1 M increments. The p53TAD concentration for all samples was 0.3 mM and the buffer was 50 mM NaH_2PO_4 (pH 6.5), 50 mM NaCl, 1 mM EDTA, 0.02% NaN_3 plus the different concentrations of urea. Backbone resonances assignments for p53TAD in 0M urea were previously reported.[41] Assignments of amide ^{1}H and ^{15}N resonances in 1-6 M urea were determined by following the movement of the peaks at the different concentrations of urea. Amide ^{1}H and ^{15}N chemical shifts were measured for each assigned residue at each urea concentration and the average chemical shift differences were calculated. These residue-specific chemical shift differences were plotted against urea concentration. The plots were best fit using the formula for a straight line and the correlation coefficients for all of the fits were greater than 0.99. The slopes for each of the residue

specific fits were then plotted to show how the magnitude of the chemical shifts change as a function of urea.

Generating structural ensembles

All atom structures of p53TAD were determined using our new approach termed broad ensemble generator with re-weighting (BEGR). The BEGR method is inspired by the previous work of Ytreberg and Zuckerman and was developed to determine structural ensembles for IDPs.[59] For the current study the following three steps were utilized: (i) Generate a large pool of plausible p53TAD structures. (ii) Calculate the simulated spectrum for each structure in the pool (in this case SAXS spectra were simulated but other types of spectroscopic data can be used). (iii) Determine the probability (i.e., weights) for each structure in the pool such that the average simulated spectrum best fits the experimental spectrum.

For the first BEGR step a pool of 100,000 p53TAD structures was generated. The BEGR software uses a random build-up scheme to generate structurally diverse conformations with no bias toward structures contained in the protein data bank (PDB).[60] This is in contrast to other methods such as RANCH and ROSETTA.[27,61] The RANCH algorithm uses randomly selected amino acid conformations from a library derived from a database of coil conformations found in high-resolution X-ray structures while ROSETTA generates structures using an empirically optimized procedure to select protein fragments from the PDB. For the BEGR approach, single amino acids with optimized side-chain geometries are used as building blocks. Starting from the N-terminus, the first two amino acids of p53TAD were joined together with randomly generated Φ-Ψ angles. A test is then performed for a steric clash between the two amino acids. If a steric clash is detected then a new pair of Φ-Ψ angles is randomly generated. This process is repeated until this clash is eliminated. Using this procedure, subsequent amino acids were added until the C-terminal amino acid was reached. Generating 100,000 structures for p53TAD required less than a day of simulation on a modern desktop computer.

For the second BEGR step, the simulated SAXS spectrum for each of the 100,000 structures in the pool was generated using CRYSOL.[62] These simulated spectra were then normalized to the experimental spectrum by multiplying the intensity values for each simulated spectrum by the mean experimental intensity averaged over the full scattering angle and then dividing this value by the average simulated intensity.

For the third BEGR step, a weight is assigned to each structure in the pool such that the weighted average simulation spectrum best fits the experimental SAXS spectrum. In order to assign these weights, the fit between the 100,000 simulated spectra and the experimental spectrum was optimized. Mathematically this corresponds to minimizing the following quantity:

$$\chi^2 = \frac{1}{k-1} \sum_{j=1}^{k} \left(\frac{I_j^{sim} - I_j^{\exp}}{\sigma_j^{\exp}} \right), \text{ where } I_j^{sim} = \sum_{i=1}^{N} w_i \cdot I_{i,j}^{sim}, \text{ and } \sum_{I}^{N} w_i = 1.0 \quad (10)$$

N is the number of structures in the pool (100,000 for this study), w_i is the weight of structure i, k is the number of data points from the SAXS curve (420 for this study), the experimental data points are $I_j^{\exp}$ with error $\sigma_j^{\exp}$ and the simulated SAXS data points are I_j^{sim}. To minimize χ^2 a Metropolis Monte Carlo simulated annealing approach was implemented.[63,64] Simulations were performed for 50 million Monte Carlo steps and the pseudo-temperature was lowered such that the temperature change decreased exponentially

during the simulation. Trial moves consisted of randomly changing a single w_i value by $\pm 0.01/N$.

The numerical analysis performed for BEGR is severely underdetermined since the number of degrees of freedom is 100,000 (number of weights to be calculated) and the number of constraints is 421 (number of experimental data points plus one since the weights must sum to 1.0). However, there are several reasons that the weights obtained from BEGR are expected to be reasonable. First, solutions can be found via optimization, such as the Monte Carlo approach used here, since most of the weights are zero and the weights must be positive[65]. Second, the Monte Carlo approach will not necessarily produce a unique solution but it will produce a valid solution. This is indicated by the excellent fit to the SAXS data obtained using BEGR and the consistency of these results with NMR experiments. Recent work by Stultz and collaborators confirms the notion that many different sets of weights can fit the experimental data equally well but this does not detract from the validity of the solutions determined using BEGR, especially when these solutions can be used to rationalize multiple forms of independent experimental measurements[66,67]. Third, repeating BEGR using five independent pools of structures gave consistent results (data not shown).

Analyzing the structural ensembles

To test for the presence of local compactness or stiffness in the reweighted ensembles, the following analysis was performed: 1) Count the number of alpha carbons within 15 Å of a given alpha carbon for a single structure from a reweighted ensemble. This results in a value for the number of neighbors within 15 Å of each alpha carbon. These values are referred to as X1 for alpha carbon 1, X2 for alpha carbon 2, and so on. 2) Repeat step 1 for each structure in the reweighted ensemble. 3) Compute the X1–X2 correlation. This is the correlation between number of neighbors for alpha carbon 1 and number of neighbors for alpha carbon 2 in all of the structures in the reweighted ensemble. This tests whether the number of alpha carbons surrounding alpha carbons 1 and 2 change in a correlated fashion between the different structures in the reweighted ensembles. (4) Repeat step 3 for all possible pairs of alpha carbons, X1–X3, X1–X4,…, X2–X4, X2–X5,…, X(n–1)-X(n), where n is the number of residues in the polypeptide (n=73 in this study). By definition, the X2–X1 correlation is the same as the X1–X2 correlation and thus does not need to be explicitly calculated. After step 4, there are $n \times n$ values representing the correlations between the numbers of neighbors for each pair of alpha carbons in the reweighted ensemble. A high correlation value means that, within the reweighted ensemble, the numbers of neighbors for those alpha carbons tend to change together. If one of these alpha carbons has a large number of neighbors in a given structure, then the other alpha carbon is likely to also have a large number of neighbors in that same structure. If one alpha carbon has a small number of neighbors in a given structure, then the other the other alpha carbon is likely to also have a small number of neighbors. Conversely, a correlation that is zero means that, within the BEGR ensemble, the number of neighbors change independently of each other.

Results and Discussion

Radius of gyration of p53TAD

SAXS data was collected on p53TAD and processed as described in the materials and methods. The Guinier plots were perfectly linear and no trace of aggregation was detected (data not shown). Analysis of SAXS data collected over a range of concentrations (2–10 mg/ml) showed the R_g values obtained at high [p53TAD] were systematically lower than the R_g values at low [p53TAD], indicating a repulsive interaction. This is no surprise given the estimated net charge on p53TAD at pH 6.5 is 14e. Figure 1a shows a plot of R_g values for the 4 mg/ml and 10 mg/ml p53TAD scattering data collected at 0, 2, 4, 6, and 8 M urea. The

scattering data collected at 4 M urea and 4 mg/ml p53TAD was discarded because the intensity of the buffer sample was so much greater than the protein plus buffer sample that it could not be reliably subtracted. This was probably due to a small bubble entrained in the capillary during data collection. The low contrast between the solvent and the solute and the resulting low signal-to-noise ratio prevented the use of data collected on 2 mg/ml p53TAD at 6 and 8 M urea. Therefore, the data collected at 4 mg/ml was used because the repulsive interactions were still weak and the signal-to-noise ratio was good.

Figure 1b shows the distance distribution for all internuclear distances based on the Fourier transform of the entire scattering curve for a 4 mg/ml sample of p53TAD in 0 and 8 M urea. The value for the maximum diameter, D_{max}, extracted from this distribution is 100 Å in 0 M urea and 110 Å in 8 M urea. Figure 1c shows the corresponding fit to the scattering curve in reciprocal space, determined using GNOM, for the 4 mg/ml sample of p53TAD in 0 and 8 M urea. The scattering curves for both data sets are shown in black and the fits are shown in blue and red, for 0 and 8 M urea, respectively. This fitting procedure gave the same R_g values as the Guinier approximation. In 0 M urea a value of 28.0 ± 0.3 Å was calculated for R_g and in 8 M urea, R_g increases to 31.6 ± 0.1 Å. According to the power-law relationship between polymer length and the ensemble average R_g, initially proposed by Flory and experimentally validated by Doniach and Plaxco, the R_g value for a random polymer that is the same length as p53TAD should be 25.8 ± 3.0 Å.[17,18] According to our measurements, the R_g value for native p53TAD is at the upper bound of the predicted random coil value and increases upon the addition of urea.

Krakty-Porod analysis of the SAXS data

To gain some additional insight into how the p53TAD ensemble changes in response to increasing urea concentrations, Kratky-Porod analysis was performed on the 4 mg/ml SAXS data. In 0 M urea, p53TAD has a contour length of $L = 262\pm10$ Å and a statistical length of $b = 22\pm1$ Å. In 0 M urea, the contour length and the statistical length are consistent with the values predicted for a random coil.[68,69] The values predicted for the statistical length of a random coil are between 19 and 25 Å. In 8 M urea, the contour length is $L = 261\pm15$ Å and the statistical length is $b = 29\pm3$ Å. These values are typical of the values expected for a random polymer with an excluded volume effect, which is induced by the high concentration of urea.[70]

Hydrodynamic radius of p53TAD by SEC and DLS

To determine the hydrodynamic radius of p53TAD, SEC was performed in different urea concentrations. Figure 2a shows the SEC elution profiles for p53TAD in 0M and 8M urea. A hydrodynamic radius (R_h) of 23.8 Å was previously reported for p53TAD in the absence of denaturant.[32] A nearly identical value of 23.6 Å was determined using the V_e measured in 0 M urea. R_h values were also determined in 2, 4, 6, and 8M urea and these values are shown in Fig. 2b. The error bars shown in Figure 2b are based on the maximum variation in V_e, which was observed during the 0 M runs. There is a clear trend toward increasing R_h values as the concentration of urea is increased with a maximum value of 29.5 Å in 8 M urea.

The hydrodynamic radius of p53TAD was also measured using dynamic light scattering. Figure 2c and 2d show the results of static and dynamic light scattering measurements of p53TAD in 0 M urea, respectively. The positive slopes of either curve indicate that the prevailing direct and hydrodynamic interactions among the p53TAD monomers are repulsive, most likely due to incomplete shielding of intermolecular protein charge repulsion. The linear extrapolation of the mutual diffusivity to vanishing protein concentration results in a single-particle diffusivity of $D_0 = 101.1\pm0.3$ $\mu m^2/s$. Use of the Stokes-Einstein relation yields a corresponding hydrodynamic radius for p53TAD of 23.9 $\pm$

0.1 Å, which is in excellent agreement with the SEC results. The resolution of the diffusivity data matches previous results from the Muschol group.[51]

Evaluation of theoretical models for hydrodynamic radius

An empirical relationship between R_h and MW was previously developed for IDPs that behave like random coils.[4] Using this relationship, the R_h of p53TAD is predicted to be 23.9 Å while a chemically unfolded globular protein the same size as p53TAD is predicted to have an R_h of 24.5 Å. By comparing the measured R_h values in Fig. 2 with this prediction, one would conclude that p53TAD behaves like a random coil in 0 M urea and is more extended than a random coil in 8 M urea. This conclusion is consistent with the R_g values calculated from the SAXS data collected in 0 and 8 M urea.

A recent study by Forman-Kay and colleagues suggests that proline content and net charge have a significant effect on the inherent compactness of an IDP.[16] They developed an empirical relationship between polymer length and R_h for IDPs that accounts for proline content and net charge. Using this relationship, p53TAD is predicted to have an R_h of 26.6 Å. This predicted value for R_h is greater than the value measured for p53TAD in 0 M urea and smaller than the value measured in 8 M urea. The correction factors for proline content and net charge were not tested against R_g values. However, we applied the same correction factor used for R_h to the R_g value predicted using the power law relationship proposed by Flory. In this case, the predicted R_g value of 30.2 Å, is also larger than the R_g value measured in 0 M urea and is close to the R_g value measured in the presence of 8 M urea.

Comparing the predictions for R_h developed by either Uversky or Forman-Kay to the measured values indicates that the native state ensemble of p53TAD has an inherent degree of compactness that is eliminated in the chemically unfolded state. This is consistent with our previous results on the ensemble structure of p53TAD that showed a tendency for partially collapsed structures and our previous analysis of hydrophobicity and net charge, which showed that p53TAD is near the border between disordered and ordered proteins.[31,32] These comparisons also suggest that there are structures in the chemically unfolded ensemble of p53TAD that are more extended than expected for a random coil.

The empirical relationship developed by Forman-Kay and colleagues is a first attempt to use the unique sequence properties of IDPs in the prediction of R_h. However, this relationship is based on the average properties of the sequence, and does not take into account the local features. It is therefore not surprising that the study by Forman-Kay does not show any correlation between R_h and global hydrophobicity. In the next sections, we demonstrate that the degree of collapse observed for p53TAD is, in part, due to the presence of local hydrophobic clusters that are generating local residual structures. These local hydrophobic clusters are observed at the MDM2 and RPA70 binding sites of p53TAD and are involved in the coupled folding and binding of these two important protein partners.

Localized structural changes upon unfolding

NMR resonances from ^{1}H-^{15}N HSQC experiments of p53TAD in 0, 2, 4, and 6 M urea were collected and analyzed to identify the local regions of p53TAD most affected by unfolding. Amide ^{1}H and ^{15}N chemical shifts for individual residues were averaged and these averaged chemical shifts were plotted as a function of urea concentration. Figure 3a shows an overlay of a selected region from the ^{1}H-^{15}N HSQC spectra at 0, 2, 4, and 6 M urea. The inset of Figure 3a shows a plot for one of the most affected residues (L26) and one of the least affected residues (D48). The average chemical shift changes were fit using a formula for a straight line and the fits are shown for L26 and D48. The R^2 values for all the fits were >0.99 and the average slope was 0.02±0.008. The slopes for the linear fits are plotted as

black bars in Figure 3b and can be used to identify the regions with the largest environmental changes. For comparison, the red line is a hydrophobicity plot for the p53TAD sequence. Some of the largest slope values correspond to nonpolar residues in the MDM2 binding site (L25 and L26) and aromatic residues in the RPA70 binding site (W54 and F55). Previous structural studies of p53TAD have shown that these two segments have some transient secondary structure and form transient long-range contacts.[32,41,44]

Structural ensembles of p53TAD

The details of our new BEGR algorithm are presented in the materials and methods. Briefly, a pool of 100,000 plausible p53TAD structures was generated using a random buildup scheme that fully samples the allowed regions of Ramachandran space and does not introduce any bias related to secondary structure preference or the structures in the PDB. These p53TAD structures are reweighted to minimize the difference between the simulated and experimental SAXS data. After performing the reweighting step, most of the structures have zero weights. In fact, less than 1% of the structures from the initial pool of 100,000 structures contribute significantly to the weighted average spectrum. For 0 M urea, 227 structures had non-zero weights, and for 8 M urea, 189 structures had non-zero weights. The term "BEGR ensemble" will be used hereafter to refer to the subset of structures with weights greater than zero.

Figure 4a shows the agreement between the simulated SAXS spectra from the BEGR ensembles and the experimental SAXS spectra for p53TAD. The blue curve shows the fit to the 0 M SAXS data and the red curve shows the fit to the 8 M SAXS data. Both the 0 M and 8 M BEGR ensembles were generated using the same initial pool of 100,000 structures. These results demonstrate that broad sampling of Ramachandran space combined with reweighting is an effective way to generate an ensemble of structures that provide an excellent fit to the experimental SAXS data.

Figure 4b shows R_g distributions for the BEGR ensembles used to fit the SAXS data collected in 0 M (blue curve) and 8 M urea (red curve). The shift in the distribution of R_g values involves a reduction in the number of structures with R_g values less than 25 Å and an increase in the number of structures with R_g values greater than 30 Å. BEGR ensembles from five independently generated pools of 100,000 structures produced very similar R_g distributions for the 0 M and 8 M datasets (data not shown).

Several groups have developed protocols for generating structural ensembles of disordered and chemically denatured proteins.[10,22,23,26–28,66,67,71–75] The ensemble optimization method (EOM) developed by Blackledge, Svergun, and colleagues was chosen for comparison with BEGR because the software is freely available and has been optimized for use with SAXS data.[27] Ten thousand random structures were generated using the RANCH algorithm (distributed as part of the EOM software package). This ensemble size was chosen to be consistent with the author's suggested use of RANCH. Then, the simulated SAXS spectra were calculated for the 10,000 structures using CRYSOL. Finally, the best-fit simulated SAXS curve was generated using the GAJOE algorithm (also distributed as part of the EOM software package). The dashed lines in Fig. 4b show the R_g distributions for the EOM ensembles. The R_g distributions for the BEGR and EOM ensembles are similar with one important distinction: the 0 and 8M BEGR ensembles both contain more extended structures than the 0 and 8 M EOM ensembles. This is interesting because the R_g distributions from the initial pools of structures used by BEGR and EOM are nearly identical, which means that EOM has more extended structures available but does not select these structures. Since the EOM code is not open source it is difficult to make an in depth conclusion about the differences with the BEGR method. However, a comparison of the average R_g from the EOM and BEGR ensembles with the R_g inferred from the Guinier law

does provide some indication that BEGR is doing a better job of selecting structures from the initial pool. The average R_g for the 0 M and 8 M EOM ensembles are 27.5 and 31.0 Å, respectively, and the average R_g for the 0 M and 8 M BEGR ensembles are 29.3 and 32.7 Å, respectively. The R_g inferred from the Guinier law for the 0 M SAXS data is 28.0 Å and for the 8 M SAXS data is 31.6 Å. It has been demonstrated that the Guinier law will underestimate the true R_g if there are highly extended structures present in the ensemble[76]. It is clear from the internuclear distance distributions shown in Figure 1b that there are highly extended structures present in the p53TAD ensemble and thus that the true R_g should be slightly larger than the value inferred from the Guinier law. Exactly how much larger the R_g value should be is not clear, but the fact the BEGR is giving a larger value than EOM suggest that it is doing a better job of selecting structures from the initial pool. It is also possible that BEGR has better structures to choose from since it generates all atom models and EOM is a coarse grain approach.

Identifying regions of local compactness or stiffness in the high weight structures

SAXS provides information about the distribution of conformations in the equilibrium ensemble at a resolution of about 15 Å. Higher resolution information is provided by complementary methods, such as NMR or molecular modeling. Using BEGR, in combination with NMR results presented above, possible regions of local compactness or stiffness along the polypeptide chain were identified. To determine whether the high weight structures in the 0 M and 8 M BEGR ensembles contain regions that are locally compact or stiff they were analyzed for correlations as described in the materials and methods. The correlation plots based on this analysis are shown in Figure 5. Positive correlations in these plots are expected for regions of the chain that experience similar positional variation between the individual structures in the BEGR ensembles. Positive correlations for sequential residues are an indication that this region of the chain is locally compact or stiff. The 50 top weighted structures from the 0 M BEGR ensemble have strong positive correlations at the two termini, the MDM2 binding site (S17-N29), and the RPA70 binding site (M40-D60). These correlations are weaker but remain positive for the 100 and 200 top weighted structures. The positive correlations observed at the MDM2 and RPA70 binding sites are probably an indication of local compactness that is induced by the presence of transient secondary structure. This is consistent with the NMR data presented in Figure 3, which shows that the largest structural changes upon urea unfolding are localized to these binding sites, suggesting the transient secondary structure present in these binding sites is destabilized by urea. The positive correlations observed at the two termini are probably an indication of local stiffness that is induced by the presence of several prolines. The top weighted structures from the 8 M BEGR ensemble also show positive correlations, but the pattern is different when compared to the 0 M BEGR ensemble. In particular, the distribution of positive correlations for the 8 M BEGR ensemble is more localized to the backbone and more uniformly distributed along the chain. This indicates stiffening along the entire chain, which is consistent with the Krakty-Porod analysis of the 8 M SAXS data and is expected for a linear polymer with an excluded volume effect. Positive correlations observed at the two termini of the 8 M BEGR ensemble are similar to the 0 M ensemble and are relatively unaffected by the addition of urea. If the local stiffening at the two termini is being induced by the prolines, then this behavior does not appear to be very sensitive to the presence of urea.

Similar patterns of positive correlations were observed for five independently generated BEGR ensembles used to fit the 0 M and 8 M SAXS data (data not shown). While it is not reasonable to make any claim of accuracy based on this result it does appear that the BEGR method can generate realistic and reproducible ensembles for IDPs. No consistent positive

correlations were observed when the same analysis was performed on 50, 100, and 200 randomly selected structures from the initial pool of 100,000 structures.

Conclusions

This study presents the most detailed investigation to date characterizing the chemical unfolding of an IDP. While it is well established that urea is capable of inducing the unfolding and expansion of ordered proteins, the ability of this destabilizing osmolyte to affect the dynamic structures of IDPs has not been thoroughly examined. The hydrodynamic properties of the native and chemically unfolded forms of p53TAD were investigated using SAXS, SEC, DLS, and NMR. The native state of p53TAD has hydrodynamic dimensions consistent with the values expected for a random coil, and the chemically unfolded state of p53TAD has hydrodynamic dimensions that are significantly greater than the values expected for a random coil. Auton and Bolen recently showed that newly exposed side chains are responsible for 75% of the increase in surface area when ordered proteins are unfolded with urea, but the energetics of unfolding is dominated by interactions between the denaturant and the backbone.[77] If this relationship can be applied to IDPs, then the observed increase in the hydrodynamic dimensions of p53TAD is due to the exposure of new side chains. For this to happen, the side chains had to be buried in the first place. We recently showed that two segments of p53TAD, corresponding to the MDM2 and RPA70 binding sites, make transient long-range contacts.[31] These transient contacts were not observed in a structural ensemble based on residual dipolar coupling data.[45] These two segments contain four aromatic amino acids as well as other nonpolar amino acids. Exposure of all or part of the side-chains for these residues could be responsible for the observed increase in the hydrodynamic dimensions of chemically unfolded p53TAD. The data presented in Fig. 3 provides strong support for this mechanistic interpretation.

This study also demonstrates that a large pool of simulated structures with broad conformational properties can be used to determine a structural ensemble for an IDP. There is no feature of the BEGR approach that is specific to the p53TAD system and it should therefore be applicable to other IDPs. The BEGR algorithm provided structures with larger R_g values than observed using other methods and we think this is an important feature for generating more realistic structural ensembles of IDPs. We also demonstrated that SAXS data could be used to identify possible regions of local compactness or stiffness in an ensemble model of a disordered protein. The analysis developed to identify these features is unique and is not related to the standard measures for comparing structures. For instance, clustering structures by backbone RMSD or Φ-Ψ dihedral angles did not identify regions of local compactness or stiffness (data not shown). The identification of locally compact or stiff regions in the BEGR ensemble demonstrates that the shape envelope of an IDP can be deconvoluted to identify local structural features. A recent report also demonstrates that this level of molecular detail can be extracted from SAXS data[78].

Acknowledgments

References

1. Wright PE, Dyson HJ. J Mol Biol. 1999; 293:321–331. [PubMed: 10550212]

2. Daughdrill, GW.; Pielak, GJ.; Uversky, VN.; Cortese, MS.; Dunker, AK. Protein Folding Handbook. Buchner, J.; Kiefhaber, T., editors. Vol. 3. WILEY-VCH; Darmstadt: 2005. p. 275-357.

3. Vendruscolo M. Curr Opin Struct Biol. 2007; 17:15–20. [PubMed: 17239581]

4. Uversky VN. Eur J Biochem. 2002; 269:2–12. [PubMed: 11784292]

5. Uversky VN. Protein Sci. 2002; 11:739–756. [PubMed: 11910019]

6. Dyson HJ, Wright PE. Nat Rev Mol Cell Biol. 2005; 6:197–208. [PubMed: 15738986]

7. Eliezer D. Methods Mol Biol. 2007; 350:49–67. [PubMed: 16957317]

8. Eliezer D. Curr Opin Struct Biol. 2009; 19:23–30. [PubMed: 19162471]

9. Jensen MR, Markwick PR, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernado P, Blackledge M. Structure. 2009; 17:1169–1185. [PubMed: 19748338]

10. Mittag T, Forman-Kay JD. Curr Opin Struct Biol. 2007; 17:3–14. [PubMed: 17250999]

11. Shortle D. Adv Protein Chem. 2002; 62:1–23. [PubMed: 12418099]

12. Tompa P. FEBS Lett. 2005; 579:3346–3354. [PubMed: 15943980]

13. Tompa, P. Structure and Function of Intrinsically Disordered Proteins. 1. Taylor and Francis Group; Boca Raton: 2010.

14. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. BMC Genomics. 2008; 9(Suppl 2):S1.

15. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Proteins. 2001; 42:38–48. [PubMed: 11093259]

16. Marsh JA, Forman-Kay JD. Biophys J. 2010; 98:2383–2390. [PubMed: 20483348]

17. Flory, PJ. Principles of Polymer Chemistry. Cornell University Press; Ithaca, N.Y.: 1953.

18. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW. Proc Natl Acad Sci U S A. 2004; 101:12491–12496. [PubMed: 15314214]

19. Millett IS, Doniach S, Plaxco KW. Adv Protein Chem. 2002; 62:241–262. [PubMed: 12418105]

20. Tanford C, Kawahara K, Lapanje S. J Biol Chem. 1966; 241:1921–1923. [PubMed: 5947952]

21. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Biochemistry. 1999; 38:16424–16431. [PubMed: 10600103]

22. Choy WY, Forman-Kay JD. J Mol Biol. 2001; 308:1011–1032. [PubMed: 11352588]

23. Marsh JA, Forman-Kay JD. J Mol Biol. 2009; 391:359–374. [PubMed: 19501099]

24. Bernado P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M. J Am Chem Soc. 2005; 127:17968–17969. [PubMed: 16366524]

25. Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, Mandelkow E, Zweckstetter M. PLoS Biol. 2009; 7:e34. [PubMed: 19226187]

26. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. Proc Natl Acad Sci U S A. 2005; 102:17002–17007. [PubMed: 16284250]

27. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. J Am Chem Soc. 2007; 129:5656–5664. [PubMed: 17411046]

28. Jha AK, Colubri A, Freed KF, Sosnick TR. Proc Natl Acad Sci U S A. 2005; 102:13099–13104. [PubMed: 16131545]

29. Francis CJ, Lindorff-Larsen K, Best RB, Vendruscolo M. Proteins. 2006; 65:145–152. [PubMed: 16862593]

30. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. J Am Chem Soc. 2005; 127:476–477. [PubMed: 15643843]

31. Lowry DF, Hausrath AC, Daughdrill GW. Proteins. 2008; 73:918–928. [PubMed: 18536020]

32. Lowry DF, Stancik A, Shrestha RM, Daughdrill GW. Proteins. 2007; 71:587–598. [PubMed: 17972286]

33. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. J Mol Biol. 2006; 362:1043–1059. [PubMed: 16935303]

34. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Biochemistry. 2005; 44:12454–12470. [PubMed: 16156658]

35. Gely S, Lowry DF, Bernard C, Jensen MR, Blackledge M, Costanzo S, Bourhis JM, Darbon H, Daughdrill G, Longhi S. J Mol Recognit. 2010; 23:435–447. [PubMed: 20058326]

36. Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, Blackledge M. J Am Chem Soc. 2010; 132:8407–8418. [PubMed: 20499903]

37. Vise P, Baral B, Stancik A, Lowry DF, Daughdrill GW. Proteins. 2007; 67:526–530. [PubMed: 17335006]

38. Bargonetti J, Manfredi JJ. Curr Opin Oncol. 2002; 14:86–91. [PubMed: 11790986]

39. Kaustov L, Yi GS, Ayed A, Bochkareva E, Bochkarev A, Arrowsmith CH. Cell Cycle. 2006; 5:489–494. [PubMed: 16481746]

40. Woods DB, Vousden KH. Exp Cell Res. 2001; 264:56–66. [PubMed: 11237523]

41. Vise PD, Baral B, Latos AJ, Daughdrill GW. Nucleic Acids Res. 2005; 33:2061–2077. [PubMed: 15824059]

42. Bell S, Klein C, Muller L, Hansen S, Buchner J. J Mol Biol. 2002; 322:917–927. [PubMed: 12367518]

43. Dawson R, Muller L, Dehner A, Klein C, Kessler H, Buchner J. J Mol Biol. 2003; 332:1131–1141. [PubMed: 14499615]

44. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH. J Biol Chem. 2000; 275:29426–29432. [PubMed: 10884388]

45. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR. Proc Natl Acad Sci U S A. 2008; 105:5762–5767. [PubMed: 18391200]

46. von Ossowski I, Eaton JT, Czjzek M, Perkins SJ, Frandsen TP, Schulein M, Panine P, Henrissat B, Receveur-Brechot V. Biophys J. 2005; 88:2823–2832. [PubMed: 15653742]

47. Guinier, A.; Fournet, F. Small Angle Scattering of X-rays. Wiley Interscience; New York: 1955.

48. Svergun DI. Journal of Applied Crystallography. 1992; 25:495–503.

49. Perez J, Vachette P, Russo D, Desmadril M, Durand D. J Mol Biol. 2001; 308:721–743. [PubMed: 11350171]

50. Uversky VN. Biochemistry. 1993; 32:13288–13298. [PubMed: 8241185]

51. Parmar AS, Muschol M. Biophys J. 2009; 97:590–598. [PubMed: 19619474]

52. Ball V, Ramsden JJ. Biopolymers. 1998; 46:489–497.

53. Berne, BJ.; Pecora, R. Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics. Wiley; New York: 1976.

54. Jakeman, E. Photon correlation. Plenum Press; New York: 1973.

55. Brown, W. Dynamic light scattering: The method and some applications. Oxford University Press; New York: 1993.

56. Kuehner DE, Heyer C, Ramsch C, Fornefeld UM, Blanch HW, Prausnitz JM. Biophys J. 1997; 73:3211–3224. [PubMed: 9414232]

57. Muschol M, Rosenberger F. Journal of Chemical Physics. 1995; 103:10424–10432.

58. Neal DG, Purich G, Cannell DS. Journal of Chemical Physics. 1984; 80:3469–3477.

59. Ytreberg FM, Zuckerman DM. Proc Natl Acad Sci U S A. 2008; 105:7982–7987. [PubMed: 18544653]

60. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. Nucleic Acids Res. 2000; 28:235–242. [PubMed: 10592235]

61. Simons KT, Bonneau R, Ruczinski I, Baker D. Proteins. 1999; (Suppl 3):171–176. [PubMed: 10526365]

62. Svergun DI, Barberato C, Koch MHJ. Journal of Applied Crystallography. 1995; 28:768–773.

63. Kirkpatrick S, Gelatt CD Jr, Vecchi MP. Science. 1983; 220:671–680. [PubMed: 17813860]

64. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Journal of Chemical Physics. 1953; 21:1087–1091.

65. Donoho DL, Tanner J. Proc Natl Acad Sci U S A. 2005; 102:9446–9451. [PubMed: 15976026]

66. Huang A, Stultz CM. PLoS Comput Biol. 2008; 4:e1000155. [PubMed: 18725924]

67. Fisher CK, Huang A, Stultz CM. J Am Chem Soc. 2010; 132:14919–14927. [PubMed: 20925316]

68. Moncoq K, Broutin I, Craescu CT, Vachette P, Ducruix A, Durand D. Biophys J. 2004; 87:4056–4064. [PubMed: 15465854]

69. Shell SS, Putnam CD, Kolodner RD. Mol Cell. 2007; 26:565–578. [PubMed: 17531814]

70. Garcia P, Serrano L, Durand D, Rico M, Bruix M. Protein Sci. 2001; 10:1100–1112. [PubMed: 11369848]

71. Lindorff-Larsen K, Kristjansdottir S, Teilum K, Fieber W, Dobson CM, Poulsen FM, Vendruscolo M. J Am Chem Soc. 2004; 126:3291–3299. [PubMed: 15012160]

72. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. J Am Chem Soc. 2009; 131:17908–17918. [PubMed: 19908838]

73. Rozycki B, Kim YC, Hummer G. Structure. 2011; 19:109–116. [PubMed: 21220121]

74. Gillespie JR, Shortle D. J Mol Biol. 1997; 268:170–184. [PubMed: 9149150]

75. Gillespie JR, Shortle D. J Mol Biol. 1997; 268:158–169. [PubMed: 9149149]

76. Bernado P. Eur Biophys J. 2010; 39:769–780. [PubMed: 19844700]

77. Auton M, Holthauzen LM, Bolen DW. Proc Natl Acad Sci U S A. 2007; 104:15317–15322. [PubMed: 17878304]

78. Rozycki B, Kim YC, Hummer G. Structure. 19:109–116. [PubMed: 21220121]
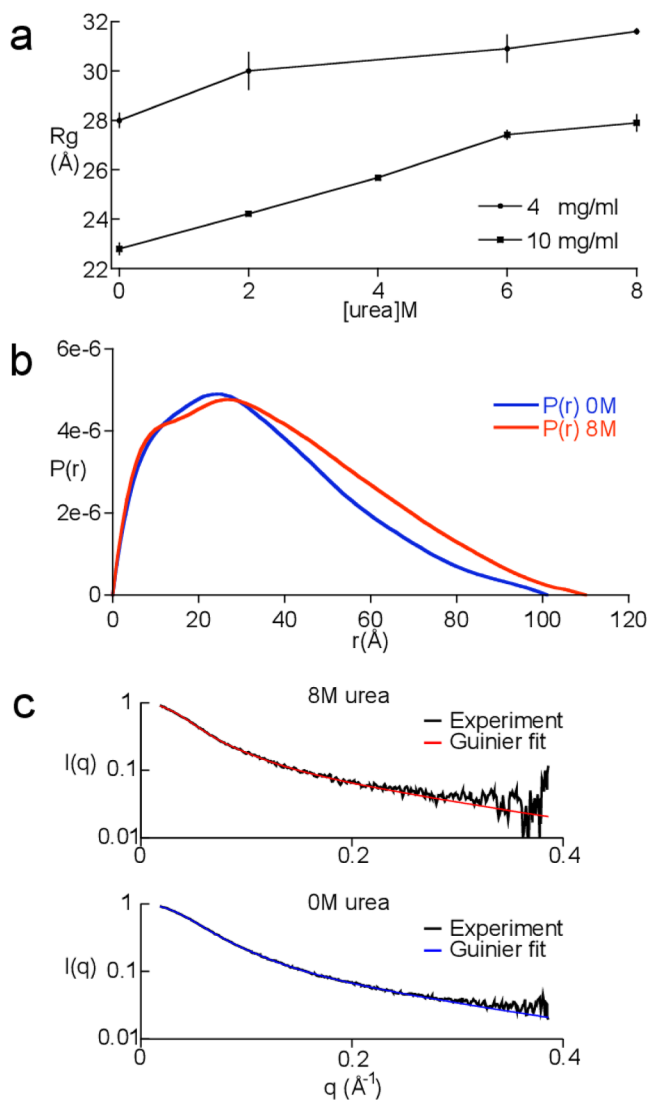
Fig. 1.
SAXS data for p53TAD. a. R_g values for p53TAD at 0, 2, 4, 6, and 8 M urea. R_g values are shown for 4 and 10 mg/ml samples. b. Internuclear distance distributions for 4 mg/ml p53TAD samples in 0 (blue curve) and 8 M (red curve) urea. c. Fits to the scattering curves in reciprocal space. Plot shows the scattering intensity, $I(q)$, as a function of the scattering vector, q. The data is shown in black and the fits are shown in blue and red for 0 and 8 M urea, respectively.
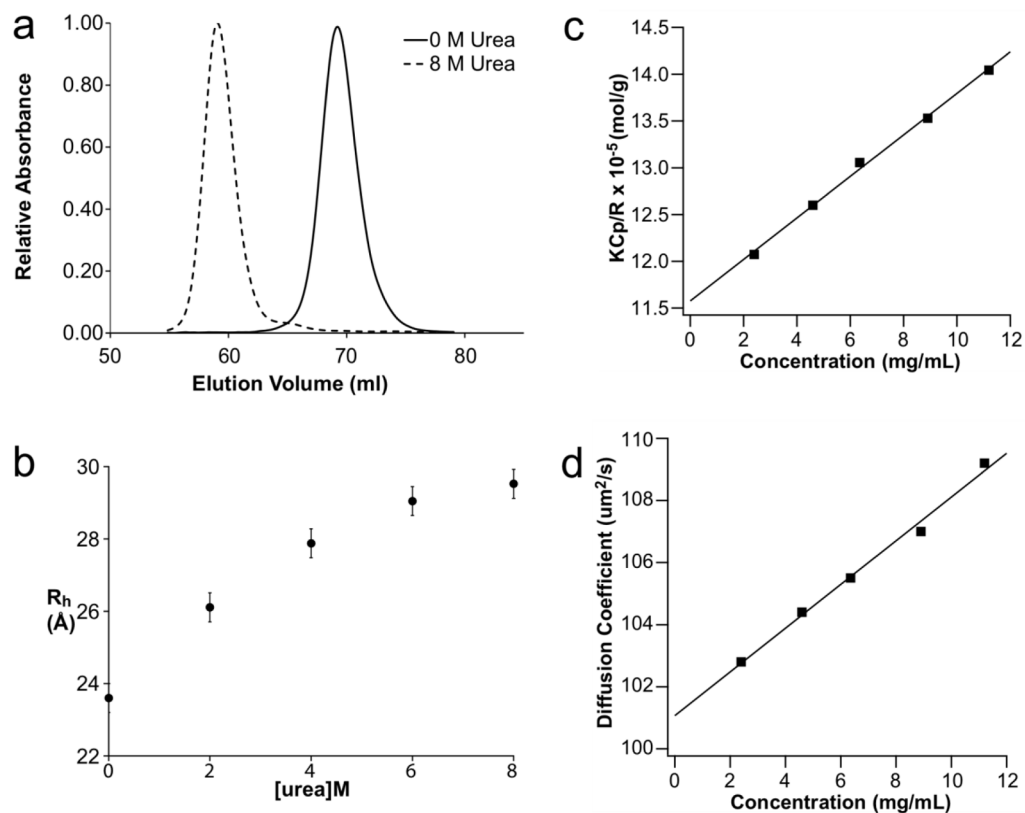
Fig. 2.
Hydrodynamic radii from size exclusion chromatography and static and dynamic light scattering. a. Chromatographic traces at 0 and 8 M urea. b. R_h values calculated at the different urea concentrations. c. The Debye Ratio (KCp/R) versus p53TAD concentration. d. Mutual diffusion coefficient D_m *vs.* p53TAD concentration. For c and d, the squares are individual measurements and the solid lines are linear fits through these data. According to eqns (2) and (4), the intercepts and slopes of these curves yield the following parameters for p53TAD: MW = 8.6 kDa, B_{22} = 11.2 x 10^{-4} mL mol/g^2, D_0 = 101.1 ± 0.3 μm^2/s, and R_h = 23.9 ± 0.1 Å.
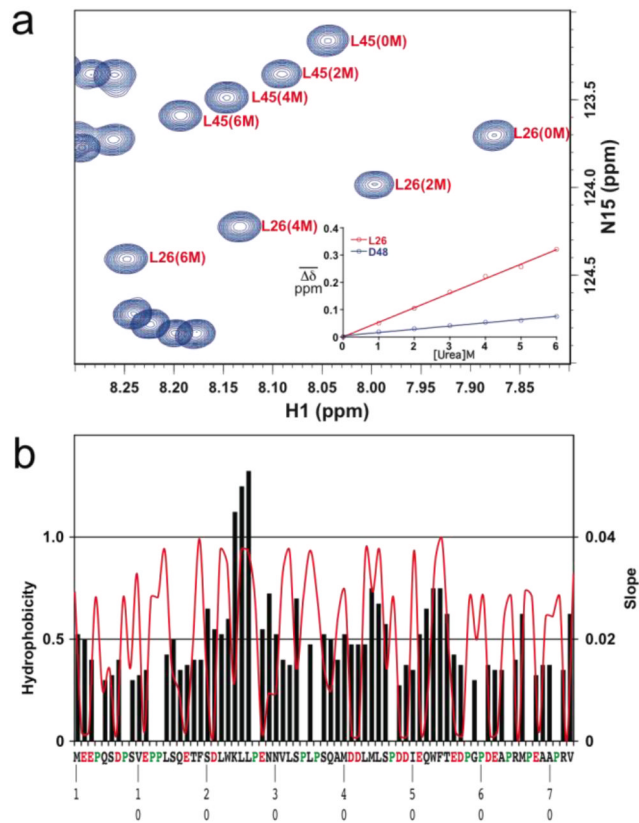
Fig. 3.
Residue specific chemical shift changes as a function of urea. a. Overlay of a selected region from the ^{1}H-^{15}N HSQC spectra at 0, 2, 4, and 6 M urea. Inset shows plots for two residues (L26 and D48). b. black bars show slope values and red line shows a hydrophobicity plot.
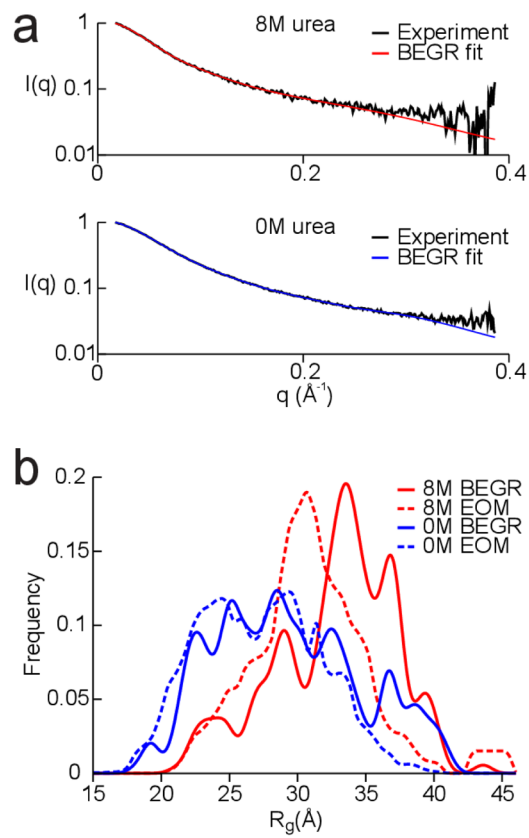
Fig. 4.
Fitting the SAXS data using the BEGR ensembles. a. Plot of the scattering intensity decay, $I(q)$, as a function of the scattering vector, q. Black lines show the SAXS data for 0 M and 8 M urea and blue and red lines show the fits using the BEGR ensembles. b. Plot of the R_g distribution for structures from the BEGR (solid lines) and EOM (dashed lines) ensembles.
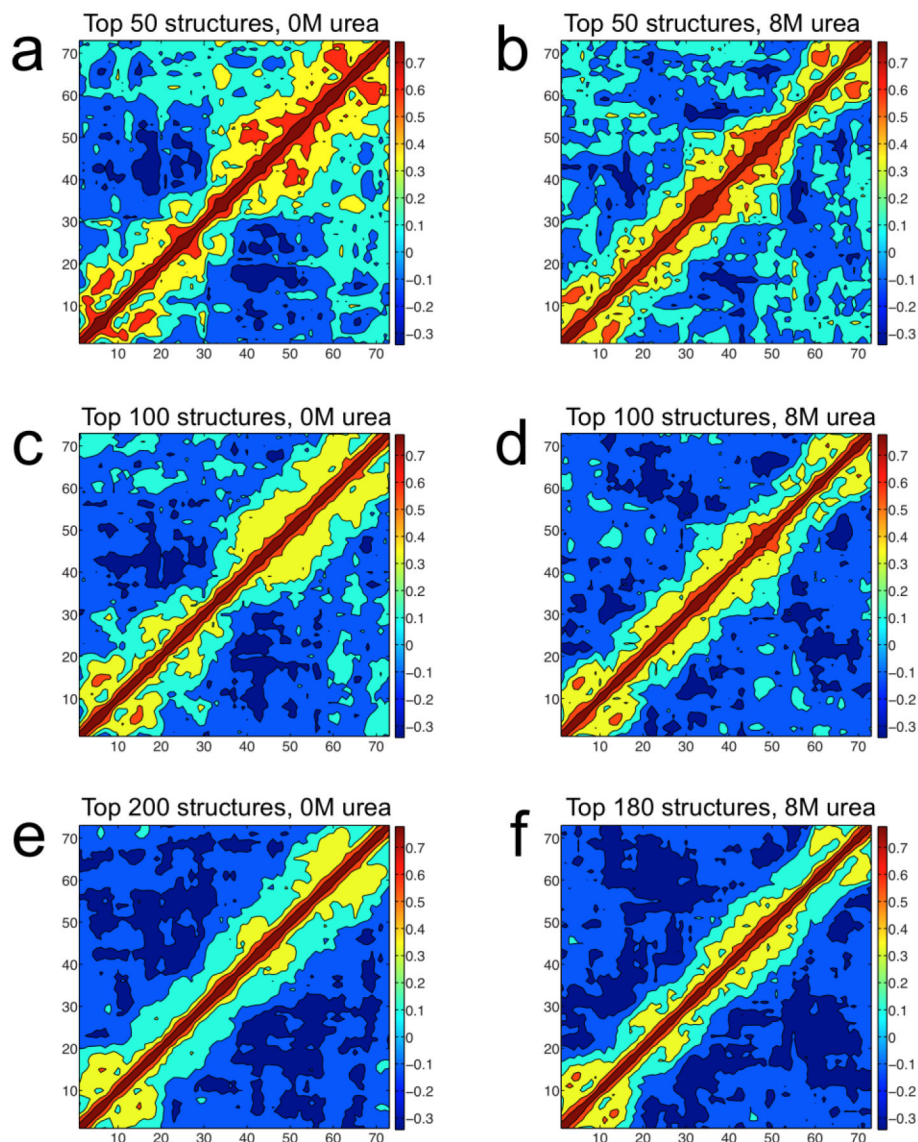
Fig. 5.
Correlation plots for number of neighbors using the top weighted structures from the 0 and 8 M BEGR ensembles. The observed correlations vary from −0.3 (dark blue) to 0.7 (dark red) and a scale bar is shown to the right of each plot. Amino acid position is shown on both axes. Correlation plots for a) 50, c) 100, and e) 200 top weighted structures from a total of 227 structures in the 0 M BEGR ensemble, representing representing 50.0%, 78.5%, and 99.5% of the total weight, respectively. Correlation plots for the b) 50, d) 100, and f) 180 top weighted structures from a total of 189 structures in the 8 M BEGR ensemble, representing 55.4%, 83.2%, and 99.8% of the total weight, respectively.