Fractal dimension of an intrinsically disordered protein: Small-angle X-ray scattering and computational study of the bacteriophage λ N protein

Daniel Johansen,¹ Jill Trehwella,^{2,3} and David P. Goldenberg^{1*}

¹Department of Biology, University of Utah, Salt Lake City, Utah

²Department of Chemistry, University of Utah, Salt Lake City, Utah

³School of Molecular Biosciences, The University of Sydney, Sydney, Australia

Received 5 July 2011; Revised 26 August 2011; Accepted 27 August 2011

DOI: 10.1002/pro.739

Published online 20 September 2011 proteinscience.org

Abstract: Small-angle X-ray scattering (SAXS) was used to characterize the bacteriophage λ N protein, a 107 residue intrinsically disordered protein (IDP) that functions as a transcriptional antitermination factor. The SAXS data were used to estimate both the average radius of gyration and the fractal dimension, a measure of the protein's internal scaling properties, under a variety of solution conditions. In the absence of denaturants, the radius of gyration was 38 ± 3.5 Å and the fractal dimension was 1.76 ± 0.05 , slightly larger than the value predicted for a well-solvated polymer with excluded volume (1.7). Neither the radius of gyration nor the fractal dimension changed significantly on the addition of urea, further indicating that the protein is extensively unfolded and well solvated in the absence of denaturant. The addition of NaCl or D₂O was found to promote aggregation, but did not appear to affect the properties of the monomeric form. The experimental SAXS profiles were also compared with those predicted by a computational model for a random-coil polypeptide, with an adjustable solvation energy term. The experimental data were well fit to the model with the solvation energy close to zero. These results indicate that the λ N protein is among the more expanded members of the broad class of IDPs, most likely because of its high content of charged residues and a large net charge (+15 at neutral pH). The expanded nature of the conformational ensemble may play a role in facilitating the interactions of the protein with other components of the dynamic transcriptional complex.

Keywords: unfolded proteins; small-angle X-ray scattering; fractal dimension; scaling exponent; solvation; radius of gyration

Introduction

Over the past decade, there has been increasing awareness of the importance of unfolded states of proteins *in vivo*.^{1–3} These partially or fully unfolded

states include the nascent forms of proteins that fold to stable three-dimensional structures, states that arise from transient unfolding, and the functional forms of some proteins that have little or no tendency to fold under physiological conditions. This latter class of proteins, referred to as intrinsically disordered proteins (IDPs), has attracted particular attention, as they appear to contradict the traditional paradigm of protein structure–function relationships.^{4–8}

One of the defining features of unfolded proteins as a class is the scaling of the average dimensions of the molecules with chain length. Hydrodynamic

Grant sponsor: U.S. National Science Foundation; Grant number: MCB-0749464 to (D.P.G.); Grant sponsor: Australian Research Council Discovery Project Scheme; Grant number: DP0770631 to (J.T.); Grant sponsor: University of Sydney (International Visiting Research Fellowship to D.P.G.).

*Correspondence to: David P. Goldenberg, Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112-0840. E-mail: goldenberg@biology.utah.edu

measurements by Tanford in the 1960s demonstrated a simple power-law relationship between the overall sizes of denaturant-unfolded proteins and their chain lengths.⁹ If the average size is expressed as the radius of gyration (R_g), the relationship has the form:

$$R_g = R_0 N^\nu, \quad (1)$$

where N is the number of amino-acid residues, R_0 is a constant, and ν is a scaling exponent. Relationships of this form are characteristic of disordered polymers, and ν is commonly referred to as the Flory exponent.^{10,11} In the intervening decades, this relationship has been confirmed and extended for a large number of proteins, using a variety of experimental techniques.¹² Small-angle X-ray scattering (SAXS) has proven particularly useful, as this method yields a more direct measurement of the radius of gyration than do hydrodynamic methods. For proteins unfolded in denaturants, such as guanidinium chloride (GuHCl) and urea, current estimates of ν place its value at about 0.6, very close to the value predicted from polymer theory for well solvated random coils with excluded volume effects accounted for.¹³

At present, the relationship between average dimensions and chain length for intrinsically disordered proteins is less well defined than for denaturant-unfolded globular proteins. Relatively few SAXS studies of IDPs have been published,^{14–16} but the hydrodynamic radii (R_h) of more than 30 IDPs have been measured by either size-exclusion chromatography or NMR-based diffusion measurements.^{17,18} For both denaturant-unfolded proteins and IDPs, R_h has been found to follow a power-law relationship with respect to chain length, similar to that observed for the radii of gyration for denatured proteins. Direct comparison of the two parameters is not straight forward, however, as they depend differently on molecular shape and the distribution of conformations in an ensemble.^{19,20} For the IDPs, there is significantly more deviation from the correlation, and many of these proteins appear to be more compact than denaturant-unfolded proteins of the same length. On the other hand, Kohn *et al.*¹³ identified five IDPs for which the radius of gyration (determined from SAXS measurements) was larger than for comparably sized denaturant-unfolded proteins and excluded these proteins from the data set used to estimate ν . These observations have led to the view that there may be multiple subclasses of IDPs with different degrees of compaction.^{17,18}

The range of results observed for IDPs highlights a general difficulty in interpreting the average dimensions of individual unfolded proteins. If the measured value of R_g or R_h for a given protein is found to be consistent with the measurements of other proteins, it is tempting to conclude that the molecule is well described as a random coil. However, a discrepancy between observed and predicted dimensions may be

due to the differences in experimental methods or conditions, aggregation, or other experimental artifacts. In addition, it is now widely appreciated that the average dimensions of unfolded molecules can be quite insensitive to the presence of nonrandom local structure.²¹

One way to address these difficulties is to examine the internal scaling relationships of individual proteins, which can be measured without reference to other molecules. If the distance between two residues in a disordered chain is R_{ij} , then the (root mean square (RMS)) average distance is predicted to follow the relationship:

$$\langle R_{ij} \rangle \propto N_{ij}^\nu, \quad (2)$$

where ν is the same exponent as in Eq. (1), provided that the number of intervening residues, N_{ij} , is sufficiently large. This relationship can be tested, and the value of ν estimated, by experimentally measuring distances between individual atom pairs within a protein, for instance by Förster resonance energy transfer (FRET).^{22,23} Carrying out measurements of this type for a series of labeled proteins is a major undertaking, however, and the individual measurements may be overly sensitive to local structures or interactions with the fluorescent probes.

An alternative approach to assess the internal scaling exponent is based on small-angle scattering by X-rays or neutrons (SANS). In most published scattering studies with unfolded proteins, the data have been used to estimate only the average radius of gyration, usually from Guinier plots derived from the scattering at very small angles, but the scattering from larger angles contains valuable additional information. At the very small angles, the intensity of scattering from a particle of any shape is predicted to follow a relationship of the form:

$$I(q) = K e^{-q^2 R_g^2/3}, \quad (3)$$

where K is a constant, and q represents the scattering angle expressed as the scattering vector magnitude, $q = (4\pi \sin \theta)/\lambda$; θ is one-half of the scattering angle, and λ is the X-ray or neutron wavelength.²⁴ This relationship is valid for $q \lesssim 1/R_g$, depending somewhat on the shape of the particle. In the Guinier plot, $\ln(I)$ is plotted versus q^2 , and the average radius of gyration is estimated from the slope of the linear region of the plot. For an ensemble of conformations, the radius of gyration derived from the Guinier plot represents an RMS average of the values for the individual molecules. At larger angles, scattering from many particle types is described by²⁵

$$I(q) \propto q^{-D_m}, \quad (4)$$

where D_m is a constant. This relationship gives rise to a linear region in a $\log(I)$ versus $\log(q)$, the negative

Figure 1. Fractal dimension as a descriptor of unfolded proteins. (A) The fractal dimension, D_m , is visualized as the exponent describing the increase in mass, m , of the protein encompassed by a sphere as the sphere radius, r , is increased. The polypeptide chain in the figure represents a random conformation of λN . (B) Simulated SAXS profiles for a globular protein (ribonuclease A, $M_r = 13,700$, solid curves) and a calculated ensemble of unfolded molecules (λN , $M_r = 12,300$, dashed curves). (C) Guinier plot of simulated SAXS profiles. The slope of the Guinier plot is used to estimate the radius of gyration. (D) Log-log plot of the simulated SAXS profiles. The slope of the linear region of the Log-log plot at mid- q values is the negative of the fractal dimension for nonglobular structures.

slope of which, D_m , reflects the distribution of interatomic distances on different length scales and can be interpreted as a mass fractal dimension.^{26–29}

The significance of the fractal dimension can be visualized by imagining a sphere enclosing a portion of an object of interest [Fig. 1(A)]. For fractal objects, that is, objects that display self-similarity over at least a limited range of scales, the mass, m , enclosed by the sphere increases with radius, r , according to a power-law relationship:

$$m \propto r^{D_m}. \quad (5)$$

For a one-dimensional object (a line), the exponent is 1; for a plane $D_m = 2$ and for a solid $D_m = 3$. The values of D_m are not limited to integers, however, and many fractal objects have D_m less than the number of spatial dimensions they occupy. An object described by a self-crossing random walk has a fractal dimension of 2. More generally, the fractal dimension for a polymer is related to the Flory exponent, ν , according to $D_m = 1/\nu$. (For globular particles, the exponent in the scattering relationship [Eq. (4)] is 4, rather than 3, because of contributions from the solvent interface.)

Because D_m can be directly estimated from the scattering profile, SAXS and SANS provide information about the internal scaling of interatomic distances, without reference to other proteins or specific models. Although the fractal dimension is not a measure of overall size, such as R_g and R_h , D_m is sensitive to favorable or unfavorable solvation of the chain, with larger values reflecting unfavorable interactions with solvent and more compact conformations.^{27,28,30–32}

The ability of SAXS or SANS to distinguish between folded and unfolded proteins is illustrated in Figure 1 (B–D) in which different representations of scattering data for a folded and an unfolded protein of similar molecular weights are shown. In each panel, the solid curve is the calculated SAXS profile for a small globular protein, bovine ribonuclease A (RNase A, $M_r = 13,700$), whereas the dashed profile was calculated for a simulated ensemble of the IDP used in this study, the bacteriophage λ N protein (λ N, $M_r = 12,300$) described below. In Panel B, the full scattering profiles are represented, showing the more rapid fall off of scattering intensity predicted for the unfolded protein. The Guinier representation is shown in Panel C, where the more negative slope for the unfolded protein represents the larger average radius of gyration for the unfolded ensemble (37 Å) than the globular protein (15 Å). A $\log(I)$ versus $\log(q)$ representation is shown in Panel D, showing the difference in the exponent, D_m . Importantly, the slope of the linear region at intermediate values of q is independent of the size of the molecule and reflects only the internal scaling exponent.

In this present study, SAXS was used to determine both the radius of gyration and the fractal dimension of λ N under a range of solution conditions. Previous studies have shown that λ N is an IDP, displaying very little structure in the absence of other molecules. *In vivo*, λ N functions as a transcriptional antitermination factor and interacts with a specific RNA sequence (the Box B segment) and multiple proteins in the transcription complex. On binding the Box B RNA, the amino terminal segment of λ N takes on an α -helical conformation, and another λ N segment binds to a bacterial host protein (NusA) in an extended conformation. Relatively little is known about the conformational properties of the other regions of λ N in the transcription complex, and it is possible that the disordered nature of the protein plays a role in facilitating structural changes in this dynamic complex. The SAXS data presented here, together with a computational simulation, confirm that the isolated protein is extensively unfolded, with the scaling exponent of a well-solvated polymer, and it is remarkably insensitive to changes in solution conditions.

Results

Small-angle X-ray scattering

SAXS data were recorded using a line-collimated beam from a sealed-tube source. This experimental arrangement provides good sensitivity with low X-ray beam intensities, but the directly recorded data represent a convolution of the beam profile with the scattering profile that would be observed with a point source, as each point on the detector records X-rays scattered at multiple angles. Two approaches were used to interpret the convoluted, or smeared, data. In the first, the observed profiles were corrected, or desmeared, using the iterative method of Lake,³³ and the corrected profiles were analyzed using the graphical methods described in the Introduction section. In the second approach, described further below, a computational model for the unfolded λ N protein was used to predict experimental SAXS profiles, which were convoluted with the experimental beam profile and then directly compared with the experimental data. This second approach uses information from the entire scattering curve and allows comparison with an explicit model of the disordered ensemble. This method also avoids the introduction of additional noise in the experimental scattering curves, which is an inevitable consequence of desmearing algorithms.

Figure 2(A) shows Guinier plots derived from the desmeared SAXS data for λ N at concentrations ranging from 3 to 20 mg/mL. Despite some noise in the data, which is enhanced by the desmearing procedure, the Guinier plots displayed good linearity up to $q = 0.04 \text{ Å}^{-1}$ and were used to estimate the average radius of gyration, as plotted in Figure 2(B). There was little systematic variation of the apparent radius of gyration with protein concentration, which, together with the linearity of the Guinier plots, indicates the absence of artifacts from aggregation or interparticle interference. The average radius of gyration from these measurements was $38 \pm 3.5 \text{ Å}$, significantly larger than the value predicted for a denaturant-unfolded protein of 107 residues ($31 \pm 0.1 \text{ Å}$ from Eq. (1) and the experimental data compiled by Kohn *et al.*¹³)

In Figure 3(A), SAXS data for λ N samples containing 0–4M urea are shown in the form of a $\log(I)$ versus $\log(q)$ plot. As expected, the region corresponding to $0.08 \leq q \leq 0.2 \text{ Å}^{-1}$ was nearly linear, and the values of the fractal dimension, D_m , estimated from the slopes are plotted in Figure 3(B). In the absence of urea, the estimate of D_m was 1.76 ± 0.05 , as compared to the values of 1.7 and 2.0 predicted for, respectively, a well-solvated polymer (with excluded volume accounted for) and an idealized random-flight chain (which can cross itself). The addition of urea caused only small changes in either D_m or the radius of gyration [Fig. 3(C)], indicating

Figure 2. SAXS from λ N at the indicated concentrations, in 50 mM Na-phosphate buffer, pH 7, at 20°C. The SAXS profiles shown in this figure and Figures 3–5 were corrected for slit smearing as described in the Materials and Methods section. (A) Guinier-plot representation of the scattering data at very small angles. The data for individual experiments have been arbitrarily displaced along the vertical axes for clarity. The error bars represent the relative uncertainties in the intensity measurements based on counting statistics. The straight lines represent least-squares fits to the Guinier relationship [Eq. (3)], from which the radii of gyration were estimated. The deviations from the straight lines are due to experimental noise, which is enhanced by the desmearing algorithm. (B) Average radius of gyration as a function of λ N protein concentration. The filled symbols represent individual measurements, and the error bars represent standard errors from the fits to Eq. (3). The line is a least-squares fit to the data.

that the protein was already quite well solvated in the absence of the denaturant.

Like many IDPs, λ N contains a large fraction of polar and charged residues, including 13 Arg, 11

Lys, two Asp, and seven Glu residues. As a consequence, it might be expected that the overall dimensions and scaling exponent of the protein would be

Figure 3. SAXS from 20 mg/mL λ N in 50 mM Na-phosphate buffer, pH 7, and the indicated urea concentrations. (A) Log-log plots of the desmeared scattering profiles. The error bars represent the relative uncertainties in the intensity measurements based on counting statistics. The cyan lines represent least-squares fits of the experimental data to the power-law relationship of Eq. (4), from which the fractal dimension, D_m , was estimated. (B) Fractal dimension versus urea concentration. The error bars represent standard errors from the least-squares fit to Eq. (4). (C) Radius of gyration versus urea concentration, determined from Guinier plots as shown in Figure 1(A), as a function of urea concentration. In Panels B and C, the lines represent least-squares fits to the data. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 4. SAXS from 10 mg/mL λ N in 50 mM Na-phosphate buffer, pH 7, and the indicated concentrations of NaCl. (A) Guinier-plot representation of the desmeared scattering data at very small angles. The error bars represent the relative uncertainties in the intensity measurements based on counting statistics. For the samples containing 0 and 0.1 M NaCl, the data were fit to Eq. (3), as represented by the straight lines. For the sample containing 0.2 M NaCl, the data were fit to a double-exponential function to account for the presence of two populations of particles with different radii of gyration, as described in Ref. 34. From this fit, the radii were estimated to be 25 ± 7 and 102 ± 9 Å. (B) Radius of gyration versus NaCl concentration, determined from the Guinier plots shown in Panel A. For the sample containing 0.2 M NaCl, the indicated value corresponds to the smaller radius estimated from the double-exponential fit. (C) Fractal dimension versus NaCl concentration, estimated from Log-log SAXS plots, as shown in Figure 3(A).

sensitive to changes in ionic strength. Guinier plots of SAXS data for λ N in the presence of 0, 0.1, and 0.2 M NaCl are shown in Figure 4(A). At the higher NaCl concentration, the Guinier plot displayed distinct curvature, indicative of salt-induced aggregation. The data for this sample were fit to a double-exponential function, consistent with two populations with radii of gyration of 25 ± 7 and 102 ± 9 Å, the smaller of which was assumed to reflect the monomeric λ N. From the relative radii of gyration of the two particle classes and the intensities extrapolated to zero scattering angle, the monomeric form was estimated to represent $\sim 67\%$ of the total scattering protein (see the Materials and Methods section for details of this calculation). The fractal dimension estimated for the sample containing 0.2 M NaCl was 2.0 [Fig. 4(C)], significantly greater than observed for the other conditions examined,

but this value is likely influenced by the presence of aggregates. At a lower concentration of NaCl, 0.1 M NaCl, the radius of gyration and fractal dimension were similar to those measured in the absence of added NaCl.

The effects of D_2O on the conformation of λ N were also examined (Fig. 5). These measurements were motivated in part by our parallel SANS studies of molecular crowding of unfolded proteins, which require the use of D_2O for contrast matching.³⁴ The replacement of H_2O by D_2O is known to increase the stabilities of many globular proteins,^{35–37} and this effect is generally attributed to an enhancement of the hydrophobic effect in D_2O ,^{38,39} although intramolecular hydrogen bonding is also likely to play a role.⁴⁰ It might be expected, then, that more collapsed conformations of an unfolded protein would be favored in D_2O . The most notable effect of D_2O on λ N, however, was an increased tendency towards aggregation, as indicated by curvature of the Guinier plots, particularly for the sample containing 20% D_2O . In this case, the curvature was not so pronounced as to yield a well-defined fit to a double-exponential function, as used in Figure 4, but the standard Guinier fit resulted in an apparent radius of gyration of 57 Å. The data for a sample containing 60% D_2O showed evidence for less pronounced aggregation, with $R_g = 46$ Å. The greater extent of aggregation at the lower D_2O concentration was

Figure 5. SAXS from 10 mg/mL λ N in 50 mM Na-phosphate buffer, pH 7, and the indicated concentrations of D_2O . (A) Guinier plots of the desmeared data. The relative uncertainties, as calculated for the other figures, are smaller than the symbols. (B) Radius of gyration versus D_2O concentration. (C) Fractal dimension as a function of D_2O concentration. Details as in Figure 4.

Figure 6. Distributions of backbone dihedral angles in simulated conformations of λ N (gray-scale histogram), compared with the distributions observed in high-resolution crystal structures of native proteins, as compiled by Lovell *et al.*⁴² (contours). (A) Distribution for Ala12 of the λ N simulated ensemble and the crystallographically observed distribution for non-Gly, non-Pro residues (excluding residues preceding Pro). (B) Distribution for Gly87 of the λ N simulated ensemble and the crystallographically observed distribution for Gly residues. The gray scale indicates the probabilities of the dihedral angle pairs lying within a grid square 5 Å on each side, relative to that expected if all dihedral angle pairs were equally probable. The contours representing the dihedrals observed in native proteins enclose 90% (solid curves) and 99% (dashed curves) of the conformations.

most likely due to the differences in sample handling and the time dependence of aggregation. The sample containing 20% D₂O displayed a slightly larger fractal dimension (1.84 ± 0.03), as would be expected from a general collapse of the polypeptide chain, but the difference was small and was not observed at the higher D₂O concentration. These results are consistent with SANS measurements of λ N in 46% D₂O, from which the radius of gyration of

the monomeric form was estimated to be 38 ± 3 Å and the fractal dimension 1.7 ± 0.02 . Aggregation was also clearly visible from the SANS data. It thus appears that D₂O promotes aggregation, perhaps by enhancing the hydrophobic effect, but has little effect on the overall conformation of monomeric λ N.

Comparison with a computational model

The SAXS data for λ N under different solution conditions were also analyzed by direct comparison to the profiles predicted by a previously described computational model of unfolded proteins.^{20,28} In this model, the polypeptide chain is represented at an atomic level, excluding hydrogen atoms, with bond lengths and angles set to ideal values. The dihedral angles for all rotatable bonds are initially set to random values and then adjusted to minimize steric overlaps, both local and long range. No other energetic terms are incorporated, and the model implicitly assumes that the polypeptide interacts equally well with the solvent and itself. This assumption corresponds to the condition described as an athermal solvent in polymer theory. An extension of the model includes an implicit solvation energy term (based on accessible surface area, ASA) by introducing a Boltzmann weighting factor for the randomly-generated conformations.²⁸ Previous work has shown that this model does a quite good job of predicting the average dimensions of denaturant-unfolded proteins²⁰ and the SAXS profile for a reduced and unfolded form of RNase A.²⁸

Despite the consistency with experimental data, the computational model described above has been criticized because the sampling procedure leads to a significant number of sterically disfavored backbone dihedral angles, with $\sim 1/3$ of the non-Gly residues located in the right-hand side of the Ramachandran plot.⁴¹ To address this problem, a modified procedure was introduced, in which the initial dihedral angles were chosen only from the allowed regions, rather than from all possible values as in the original procedure. Subsequent adjustment to minimize steric overlaps was found to cause only small changes in the dihedral angles, with the result that the final distributions closely matched those from which the initial values were drawn. Examples of the distributions for Ala and Gly residues are shown in Figure 6, in which the observed distributions are represented as a gray-scale histogram overlaid with contour lines representing the distribution observed in high-resolution crystal structures, as compiled by Lovell *et al.*⁴²

This modified sampling procedure was tested by generating simulated ensembles of five proteins, with chain lengths from 26 to 415 residues. For each sequence, $\sim 200,000$ conformations were generated, and the average radii of gyration for these ensembles are plotted in Figure 7, along with values from

Figure 7. Average radii of gyration calculated from computational simulations of unfolded proteins, compared with experimental results for denaturant-unfolded proteins. The open diamonds represent the simulation results obtained using the improved dihedral-angle sampling protocol described in this article, whereas the open circles represent the results obtained previously using the original sampling procedure.²⁰ Filled circles represent experimental data for proteins unfolded in urea or GuHCl, as compiled by Kohn *et al.*,¹³ with the outliers identified by these authors excluded. The asterisk represents the radius of gyration of λ N (38 Å) determined in this study. The dashed line is a least-squares fit of the new simulation data to Eq. (1), with fit values $R_0 = 2.0 \pm 0.15$ Å and $\nu = 0.62 \pm 0.013$. The solid line is a fit of the experimental data to the same equation, with fit values $R_0 = 2.0 \pm 0.17$ Å and $\nu = 0.59 \pm 0.015$. The uncertainties represent standard errors from the least-squares fits.

simulations using the original sampling method and the experimental values for denaturant-unfolded proteins compiled by Kohn *et al.*¹³ As shown in the figure, the new procedure resulted in average dimensions that were slightly larger than those of the ensembles generated by the original procedure, as well as the experimentally determined values.

Thus, the presumably more realistic distribution of dihedral angles leads to a slightly poorer agreement with experimental results. Although this result might be taken to indicate that the original sampling is a better representation of the dihedral angles found in unfolded proteins, this interpretation seems unlikely given the prevalence of highly unfavorable residue conformations observed in those simulations. Rather, the very good agreement with experimental data observed previously may have been due to compensation between the sampling of unrealistic dihedrals and underestimation of other factors that favor more compact conformations, per-

haps including unfavorable solvation or local conformational preferences. Nonetheless, the agreement between the newer computational results and experimental data for denaturant-unfolded proteins was still quite good, and no further adjustments to the dihedral sampling were deemed to be justified at this point.

As noted above, the simulations assume that the interactions of the solvent with the polypeptide chain are energetically equivalent to interactions within the chain, and the agreement with the experimental results suggests that concentrated solutions of urea and GuHCl may approximate this idealization. In the absence of denaturant, however, solvation is expected to be less favorable and more compact conformations may be stabilized. To simulate this effect, the accessible surface area, ASA, of each of the simulated conformations was calculated, and a Boltzmann weighting factor for each was computed as:

$$w_j = \frac{e^{-\Delta G_{\text{solv}} \cdot \text{ASA}_j / RT}}{\sum_{j=1}^N e^{-\Delta G_{\text{solv}} \cdot \text{ASA}_j / RT}}, \quad (6)$$

where N is the total number of conformations and ΔG_{solv} is an assumed free energy per unit of surface area. These weights were then used to calculate distributions of the radii of gyration and average SAXS profiles. In Figure 8(A), the unweighted distribution for the simulated ensemble of λ N is shown as a solid line. The distribution includes conformations with radii of gyration ranging from 17 to 68 Å, with an average value of 36 Å. Weighted distributions calculated for ΔG_{solv} values from -4 to 4 cal/mol/Å² are shown as dashed lines, and the inset of Figure 8(A) shows the average radius of gyration as a function of ΔG_{solv} . Negative values of ΔG_{solv} have relatively little effect on the distribution, reflecting the fact that the majority of conformations in the original distribution have large ASAs.²⁸ Positive values of ΔG_{solv} , however, favor the much smaller number of more compact conformations with low ASAs, leading to much narrower distributions.

SAXS profiles for the individual simulated conformations were calculated using the program CRY-SOL,⁴³ with standard parameters for a hydration layer, and average profiles corresponding to different values of ΔG_{solv} are plotted in Figure 8(B). The SAXS profile is relatively insensitive to negative values of ΔG_{solv} , but positive values lead to a pronounced shift of the curves to larger scattering angles, reflecting the predominance of more compact conformations. For each of the average SAXS profiles, the fractal dimension was estimated from the slope of the linear region of $\log(I)$ versus $\log(q)$ plots, and these values are plotted as a function of ΔG_{solv} in the inset to Figure 8(B). For the unweighted distribution, $D_m = 1.66$, close to the value predicted from

Figure 8. Simulation of solvation effects on the unfolded λ N ensemble. (A) Distribution of radii of gyration. An ensemble of $\sim 200,000$ conformations of λ N was generated assuming random distributions of dihedral angles and constrained only by the elimination of steric overlaps. The solid curve represents the unweighted distribution, with a bin width of $2 \text{ \AA}$. The dashed curves represent distributions weighted according to an implicit solvation energy based on ASA [Eq. (6)], with values of δG_{solv} from -4 to $4 \text{ cal/mol/\AA}^2$ as indicated by the key in Panel B. The inset shows $\text{RMS}(R_g)$ as a function of ΔG_{solv} . (B) Simulated SAXS profiles for solvation-weighted distributions. The program CRYSOLE⁴³ used to predict the SAXS profiles of the individual λ N conformations, which were then weighted according to Eq. (6). The ΔG_{solv} values associated with the different curves are indicated in the key. The fractal dimensions for the weighted ensembles were estimated from log-log plots of the simulated SAXS curves and plotted as a function of ΔG_{solv} in the inset. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

polymer theory for a chain with excluded volume in an athermal solvent. Negative values of ΔG_{solv} led to a slight decrease in D_m , whereas positive values of ΔG_{solv} resulted in a quite pronounced increase in D_m . The calculated fractal dimension was ~ 2 for $\Delta G_{\text{solv}} = 2 \text{ cal/mol/\AA}^2$. This corresponds to the condition of a θ solvent, where unfavorable solvation exactly balances the excluded volume effect, and the scaling exponent is the same as for the idealized random-flight chain ($\nu = 0.5$). With still more unfavorable solvation, the fractal dimension increases further, to about 2.5 for $\Delta G_{\text{solv}} = 4 \text{ cal/mol/\AA}^2$. This is still much smaller, however, than the power-law exponent expected for a globular protein, 4.

The simulated SAXS profiles were compared with the experimental data to provide a further test of the computational model and to estimate the value of ΔG_{solv} corresponding to real solution conditions (Fig. 9). In this analysis, the simulated curves for $\Delta G_{\text{solv}} = -4$ to 4 were individually fit to the experimental data with a single adjustable parameter representing a scaling factor for the absolute scattering intensity. The value of ΔG_{solv} providing the best fit was identified by comparing the sums of the squared residuals, plotted as the reduced χ^2 statistic in the insets to Figure 9.

Under the standard buffer conditions, the experimental data were well fit by all of the simulated SAXS profiles with $\Delta G_{\text{solv}} \leq 1 \text{ cal/mol/\AA}^2$, as shown in Figure 9(A). With more positive values of ΔG_{solv} , the deviation between the experimental and simulated curves became more pronounced. The comparison, thus, confirmed that λ N in the absence of denaturant displays the overall dimensions and scaling exponent expected of a disordered and well-solvated polypeptide chain. Figure 9(B) shows that nearly identical fitting results were obtained for the sample containing 4 M urea.

The experimental data for the sample containing 60% D_2O displayed a pattern similar to those observed for the standard buffer solution and the 4 M urea sample, but the overall residuals were larger due to poor fits at the smallest q values. The larger than predicted scattering intensities at small angles is indicative of aggregation, as also inferred by the Guinier plots of Figure 5. Outside of this region, however, the fit to the simulated curves for $\Delta G_{\text{solv}} \leq 1 \text{ cal/mol/\AA}^2$ was excellent, again indicating that the deuterated solution did not lead to measurable compaction of the chain. This example demonstrates one of the benefits of fitting the entire scattering profile

Figure 9. Direct fitting of experimental SAXS profiles to profiles predicted for simulated ensembles of λ N. Each of the simulated profiles corresponding to different values of ΔG_{solv} , as shown in Figure 8(B), was convoluted with the beam profile for the experiments and fit to the uncorrected experimental data by introducing a single adjustable scaling factor representing a displacement of the curve along the logarithmic vertical axis. The data were fit over the range of $q = 0.02\text{--}0.3 \text{ \AA}^{-1}$, as indicated by the solid portions of the curves. The individual best fits to the experimental profiles were compared using the reduced χ^2 statistic, as plotted in the insets. (A) 50 mM Na-phosphate buffer, as in Figure 2. (B) 4 M urea, as in Figure 3 (C) 60% D₂O, as in Figure 5. (D) 0.2 M NaCl, as in Figure 4. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

to a model curve, allowing the data to be interpreted even though the Guinier region is strongly influenced by a small fraction of aggregated molecules.

Significant aggregation was also indicated by the scattering profile for the sample containing 0.2 M NaCl [Fig. 9(D)]. In addition, the best fit for this sample was found with $\Delta G_{\text{solv}} = 2 \text{ cal/mol/\AA}^2$, reflecting the larger apparent fractal dimension estimated from the $\log(I)$ versus $\log(q)$ plot [Fig. 4(C)]. It is

likely, however, that the larger D_m value is heavily influenced by the aggregated material, making it difficult to assess the state of the monomeric protein.

Together with the graphical analyses described in the previous section, the comparison of experimental data with the simulated scattering profiles indicates that λ N is well described by a random-coil model and undergoes very little change on the addition of either urea or D₂O to the solution. The

protein remains predominantly monomeric at neutral pH in the absence of denaturant, but tends to aggregate in the presence of D₂O and added NaCl.

These and previous studies have shown that the SAXS profiles for at least some disordered proteins can be fit to quite simple computational models.^{28,44,45} In other cases, especially those involving both globular and disordered domains, these simple models are unlikely to be adequate, and optimization methods have been used to derive ensemble models for such proteins from SAXS data.⁴⁶

Discussion

λN is an expanded IDP

The λN protein was first shown to be extensively unfolded by circular dichroism (CD) and NMR spectroscopy.^{47,48} The CD spectrum of the free protein is characterized by a deep trough at ~200 nm, as expected for a random coil, and there is relatively little chemical shift dispersion in ¹H NMR spectra. λN also displays some, but not all, of the sequence features characteristic of IDPs. Of 107 residues, the sequence includes 33 charged residues, with an excess of 15 positive charges. On the normalized Kyte–Doolittle scale,⁴⁹ used by Uversky *et al.*,⁵⁰ the average hydrophobicity of λN is 0.39 (where 1 represents the most hydrophobic residue and 0 the least). The charge and hydrophobicity of λN place it within the disordered region of the Uversky charge–hydrophobicity plot and yield a score of –0.183 for the related FoldIndex score, in which 0 represents the cutoff for the prediction of disordered and folded proteins. On the other hand, the λN sequence differs from that of many other IDPs in that it does not display unusually low sequence complexity (other than a six-residue Arg-rich segment near the N terminus), as calculated by either the Seg or CAST programs.^{51,52}

Hydrodynamic studies of other IDPs suggest that many of these proteins, in the absence of denaturants, are more compact than denaturant-unfolded globular proteins of the same length.^{17,18} λN, however, appears to be among the more highly expanded IDPs, with an average radius of gyration (38 ± 3.5 Å) slightly greater than that predicted by the correlation for denaturant-unfolded proteins (Fig. 7). Furthermore, the radius of gyration was insensitive to added urea, indicating that the polypeptide chain is already fully expanded in the physiological buffer. The fractal dimension of λN in phosphate buffer (1.76 ± 0.05) is also indicative of a well-solvated polymer, as discussed further below.

The factors that determine the relative degree of expansion of different IDPs are not yet well understood. In a recent analysis, Marsh and Forman-Kay have demonstrated a significant correlation between the hydrodynamic radii of IDPs and net charge (after correction for chain length).¹⁸

Larger hydrodynamic radii were also correlated with higher proline content, but there was no significant correlation with average hydrophobicity. As noted above, λN has a quite high net charge (+15 at neutral pH), which likely contributes to its relatively large radius of gyration, but does not have a particularly high content of Pro residues (6.5%). The other factor that likely influences the average dimensions of some IDPs is the presence of any persistent long-range contacts or stable tertiary structure, even at levels that may not be easily detected spectroscopically. The insensitivity of λN to urea at concentrations up to 4 M suggests that such structures are rarely present in this protein.

Although λN appeared to be quite soluble in aqueous solution, showing little aggregation at concentrations up to 20 mg/mL, its tendency to aggregate increased significantly with the addition of NaCl or D₂O. The increase in ionic strength may act to shield repulsive electrostatic interactions among the large number of positively charged residues, whereas D₂O may enhance the hydrophobic effect.^{38,39} Neither increased ionic strength nor D₂O appeared to cause significant compaction of the monomeric protein, however, suggesting that the expanded nature of the protein is quite robust. A predominance of expanded conformations may be functionally important for λN, which must form interactions with multiple components of the transcriptional complex to facilitate antitermination.

Fractal dimensions and the solvation of unfolded proteins

In the language of polymer theory, the interaction between a polymer and its surrounding solvent is described using the terms poor and good to define the solvent. In a poor solvent, the monomers of the chain interact more favorably with other monomers than with the solvent, on average, whereas in a good solvent, the polymer interacts either as well or better with the solvent than with itself. In the computational model used here, good and poor solvents correspond to positive and negative values of ΔG_{solv} , respectively. An athermal solvent is one in which the monomer–solvent and monomer–monomer interactions are energetically equivalent that is, $\Delta G_{\text{solv}} = 0$. Theory and experiment both show that the fractal dimension of a polymer is sensitive to solvation: poor solvents lead to more compact conformations and a larger fractal dimension, whereas good solvents lead to a smaller value of D_m .^{30,53}

Given the very different properties of the functional groups making polypeptide chains, it might be expected that no single parameter could describe the solvation of an unfolded protein. However, the scaling of R_g with chain length for denaturant-unfolded proteins indicates a fractal dimension very close to 1.7, the value predicted for a chain in an athermal

solvent.¹³ Thus, concentrated solutions of urea and GuHCl appear to effectively solvate a wide range of different protein sequences.

For physiological conditions, there is less experimental data from which to infer the solvation properties of unfolded proteins. Several denatured globular proteins have been observed to undergo collapse on dilution of the denaturant, but this does not appear to be a universal phenomenon.^{54–56} For a set of proteins that do undergo collapse, Ziv and Haran⁵⁷ have applied an analytical polymer model that, like the computational model used here, includes a general solvation energy term. Although the two energies are not exactly equivalent, the values estimated by these authors correspond to a value of ΔG_{solv} of ~ 4 cal/mol/Å² at low denaturant concentration. This result is consistent with the predictions of our computational model for unfavorable solvation [Fig. 8(A)], but not the experimental results with λ N, thus highlighting the very different behavior observed for the different proteins.

The very existence of IDPs indicates that some sequences, at least, can remain unfolded and moderately well solvated in the absence of denaturants. As noted above, however, the scaling of the overall dimensions with chain length appears to be significantly less consistent for IDPs than for denaturant-unfolded proteins, making it difficult to draw either general or specific conclusions regarding their solvation. The approach described here, in which the fractal dimension is measured directly from the scattering profile, provides a means of assessing the solvation of a particular unfolded protein without reference to measurements of other molecules. Direct comparison of the scattering profiles with those predicted by a computational model, which includes an implicit solvation parameter, further allows for a quantitative estimate of the average solvation free energy. In the case of λ N, the fractal dimension is only slightly larger than the value predicted for an athermal solvent, and the experimental data were well fit by the model assuming an athermal or good solvent, that is, with $\Delta G_{\text{solv}} \leq 1$ cal/mol/Å² [Fig. 9(A)]. The SAXS profile, and its fit to the model [Fig. 9(B)], also showed very little change with the addition of urea, again suggesting that the chain is well solvated without the denaturant.

The results for λ N can be compared with those obtained previously using the same approach with reduced and alkylated forms of RNase A.²⁸ The native state of RNase A is stabilized by four disulfide bonds, and reduction of the disulfides leads to extensive unfolding.^{58,59} For SAXS studies, the Cys thiols of reduced RNase A were modified by reaction with either iodoacetic acid or iodoacetamide to prevent reformation of the disulfides and to manipulate the net charge of the proteins. SAXS profiles confirmed that the modified proteins were extensively

disordered, but also revealed that the fractal dimension of this protein was sensitive to both urea concentration and the net charge of the polypeptide chain. Under conditions where the protein had a modest net charge ($\sim +5$), the measured fractal dimension was 1.9 ± 0.01 in the presence of 1 M urea and decreased to 1.6 ± 0.01 in 6 M urea. At lower pH, however, where the net charge was $\sim +18$, the behavior of the reduced and alkylated forms of RNase A was similar to that of λ N, with a fractal dimension of 1.6, which did not change significantly with urea concentration. The SAXS data for RNase A were also well fit by the computational model described here (using either the original sampling of backbone dihedral angles or the improved protocol), with the best fits obtained with ΔG_{solv} values from -2 to 2 cal/mol/Å².

Together, the results with λ N and reduced RNase A indicate that unfolded polypeptide chains can interact quite favorably with aqueous solvent under non-denaturing conditions, leading to expanded conformational ensembles with low fractal dimension. However, these two proteins may represent one end of a continuum. Full expansion of RNase A at low urea concentrations appears to depend on a relatively large net charge, which very likely contributes to the expansion of λ N as well. The important role of electrostatic interactions in determining the dimensions of unfolded proteins is also highlighted by a recent single-molecule FRET study.⁶⁰ In addition, both proteins have relatively low average hydrophobicities. As discussed above, studies of other IDPs suggest that there is considerable diversity in their degree of compaction, and the approaches described here should be useful in better defining the properties of individual unfolded proteins. Such studies will help identify the features that determine the behavior of these molecules and, ultimately, their functions *in vivo*.

Materials and Methods

Protein samples

Bacteriophage λ N protein was purified from *Escherichia coli* BL21(DE3) bacteria containing the expression plasmid pET-N1, described by Rees *et al.*⁶¹ Cultures were grown in the PG medium of Studier⁶² containing 100 μ g/mL ampicillin. In a typical preparation, a 500 mL culture was grown at 37°C in a 2-L flask with shaking, until the absorbance at 600 nm reached ≈ 1 . Expression of the λ N gene was then induced by the addition of isopropyl β -D-thiogalactoside to a final concentration of 2 mM, and the culture was incubated with shaking for an additional 4 h. The bacteria were harvested by centrifugation, and the pellets were stored frozen at -70°C .

The frozen bacteria were resuspended in lysis buffer (20 mM tris(hydroxymethyl)aminomethane

hydrochloride (Tris-Cl), 20 mM ethylenediaminetetraacetic acid (EDTA), 50 mM NaCl, 10% glycerol, 5 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride, and 0.1 mg/mL benzamidine), using 10 mL buffer per 1 g frozen cell pellet. The resuspended bacteria were lysed by sonication and centrifuged for 30 min at 12,000 rpm in a Beckman JA-17 rotor. The λ N protein was found in the pellet fraction, which was resuspended in 6 M GuHCl, 20 mM Tris-Cl, 0.1 mM EDTA and 1 mM DTT. The sample was centrifuged for 2 h at 30,000 rpm in a Beckman 75 Ti ultracentrifuge rotor, and the pellet fraction discarded. To fully reduce the thiol of the single Cys residue in the λ N protein, fresh DTT was added to a final concentration of 5 mM and incubated for 30 min at room temperature. The thiols were then blocked by adding iodoacetamide to a final concentration of 20 mM, followed by an additional 15-min incubation at room temperature. The guanidinium and excess thiol reagents were removed by dialysis against 20 mM Tris-Cl, 0.1 mM EDTA, and 1 mM DTT. Precipitates that were formed during dialysis were removed by centrifugation for 30 min at 12,000 rpm in a Beckman JA-17 rotor.

Following dialysis, the protein was applied to a column of Whatman Express-Ion S cation exchange resin equilibrated in chromatography buffer (50 mM Tris-Cl pH 8, 5 mM EDTA, and 50 mM NaCl). The column was washed with the same buffer until the absorbance at 280 nm of the eluent was 0.01 or less. The protein was eluted with a step gradient of 0.1–1.0 M NaCl in the chromatography buffer, in steps of 0.1 M, and the fractions containing λ N were identified by SDS gel electrophoresis and UV absorbance. From SDS gel electrophoresis, non-denaturing gel electrophoresis and reversed-phase high-performance liquid chromatography, the protein was estimated to be at least 95% homogeneous. The identity of the protein was confirmed by electrospray-ionization mass spectrometry.

For storage, the purified N protein was dialyzed against 0.1 M acetic acid and lyophilized. Prior to SAXS measurements, the lyophilized protein was dissolved in 6 M GuHCl, 50 mM pH 7 Na-phosphate buffer. The protein was then dialyzed against 50 mM pH 7 Na-phosphate buffer containing 0.5 mM benzamidine, as well as urea, NaCl or D₂O as indicated for individual experiments. SAXS curves recorded before and after lyophilization were indistinguishable.

Small-angle X-ray scattering

SAXS data were measured using an Anton Paar SAXSess instrument with a sealed-tube X-ray source and line collimation, as described by Jeffries *et al.*⁶³ X-ray intensities were recorded using two-dimensional phosphor image plates, which were read using a Perkin-Elmer Cyclone scanner. The X-ray wavelength

was 1.5418 Å (Cu K α); the sample to detector distance was 264.5 mm; the sample slit width was 10 mm. Data were recorded for 2–4 h. The image data were integrated using the program ImageJ (available at <http://rsb.info.nih.gov/ij/>; developed by Wayne Rasband, National Institutes of Health, Bethesda, MD) with locally written macros. The width of the integration area (detector slit width) was 10 mm. Relative errors were calculated as the square root of the integrated intensities. For each sample, a reference scattering curve was recorded using a sample blank with identical buffer composition, and the reference curve was subtracted directly from the sample curve before any further analysis. No additional background subtraction (e.g., “Porod correction”) was applied. The slit-smeared data were corrected using a numerical algorithm based on the iterative method of Lake,³³ with a smoothing step included at each iteration to minimize the introduction of excess noise. The correction incorporated both the experimental beam-length profile and the detector integration profile. No correction was introduced for the beam width, which was less than 5% of the beam length.

The slit-corrected experimental data were fit to the Guinier approximation or a power-law relationship by the methods of least squares, with the experimental data weighted by the inverse of the relative uncertainties. To ensure consistency in the analyses, data from the same range of scattering angles were used in the analyses of all of the samples. For the Guinier analysis, data for $q = 0.025$ – 0.04 Å^{−1} were used. For determining the fractal dimension, the power function was fit to scattering data for $q = 0.08$ – 0.2 Å^{−1}. The Guinier plot for the sample containing 0.2 M NaCl (Fig. 4) displayed two distinct phases, and the data were fit to a double exponential function, assuming the scattering arose from two classes of particles of different sizes. The relative mass concentrations of the two species (C_1 and C_2) were estimated using the following relationship:

$$\frac{C_1}{C_2} = \frac{K_1}{K_2} \left(\frac{R_{g,2}}{R_{g,1}} \right)^{1.7}, \quad (7)$$

where K_1 and K_2 represent the scattering intensities from the two particle classes extrapolated to zero scattering angle, and $R_{g,1}$ and $R_{g,2}$ are the radii of gyration of the two classes. This relationship, used previously in Ref. 34, assumes that K_1 and K_2 are each proportional to both the molecular mass and the mass concentration of the corresponding particles, and that the molecular masses are proportional to $R_g^{1.7}$, as expected for a well-solvated polymer. For the λ N sample containing 2 M NaCl, $R_{g,1} = 2$ Å; $R_{g,2} = 102$ Å; $K_1 = 0.23$; $K_2 = 1.2$. The ratio of mass concentrations, C_1/C_2 , was thus estimated to be 2.

The reported estimates of the radius of gyration and fractal dimension of λ N in the standard buffer

($R_g = 38 \pm 3.5 \text{ \AA}$ and $D_m = 1.76 \pm 0.05 \text{ \AA}$), represent the average of eight independent measurements, and the reported uncertainties are standard errors. The predicted value of R_g for a denaturant-unfolded protein of 107 residues ($31 \pm 0.1 \text{ \AA}$) was calculated from Eq. (1), using values of R_0 and v obtained from a least-squares fit of the data compiled Kohn *et al.*¹³ to this equation. The fit values were: $R_0 = 1.96 \pm 0.064 \text{ \AA}$ and $v = 0.591 \pm 0.007$ (with uncertainties representing the standard errors from the least-squares fit, which was calculated using the experimental standard errors as inverse weights). The uncertainty in the predicted radius of gyration was calculated by propagation of the uncertainties in R_0 and v , corrected for the covariance between the fit parameters (-0.00044).

Computational simulations and data fitting

Ensembles of protein conformations based on the λ N sequence were generated using the program CYANA,⁶⁴ as described previously.^{20,28} Protein sequences were represented with all non-hydrogen atoms, using special Cyana residue entries for the N- and C-terminal residues, as well as the carboxyamidomethylated form of Cys93 of λ N. For this study, and a recent simulation of molecular crowding,³⁴ an improved sampling procedure was introduced to generate more realistic distributions of backbone dihedral angles. For each simulated conformation, backbone dihedral angles were initially set by selecting randomly from lookup tables based on the values observed in high-resolution protein crystal structures, as compiled by Lovell *et al.*⁴² Separate tables were constructed for four classes of residues: Gly, Pro, β -branched (Ile, Thr, and Val), and all others. In addition, a second set of four tables were constructed for the special cases where these residue types were followed by proline. The lookup tables were refined iteratively to eliminate ϕ - ψ dihedral pairs that led to local steric overlaps during initial simulations. Examples of the final distributions obtained for Ala and Gly residues are shown in Figure 6. All side-chain values were initially set to random values between 0 and 360° . The dihedral angles were then adjusted to minimize the CYANA target function, using only steric overlap terms. Approximately 250,000 conformations were generated for each of the proteins examined: ω -conotoxin MVIIA-Gly (26 residues), bovine pancreatic trypsin inhibitor (58 residues), λ N (107 residues), the α -subunit of *E. coli* tryptophan synthase (268 residues), and yeast phosphoglycerate kinase (415 residues). The 10% of the conformations with the highest CYANA target functions were eliminated from further analysis.

The ASAs of the individual conformations were calculated using the algorithm of Lee and Richards,⁶⁵ as implemented in the program ACCESS by Richmond and the group radii of Chothia.⁶⁶ The probe ra-

dus was $1.4 \text{ \AA}$. Radii of gyration were calculated from the coordinates of the backbone α -carbons. SAXS profiles were calculated from atomic coordinates using the program CRY SOL with the default parameters, including a bulk solvent density of $0.334 \text{ electrons/\AA}^3$, a hydration layer $3\text{-\AA}$ thick, and a scattering contrast of $0.03 \text{ electrons/\AA}^3$ for the hydration layer.^{43,67} Weighted distributions of the radii and gyration and SAXS profiles were generated by applying Boltzmann weighting factors, calculated according to Eq. (6), to the values for the individual conformations.

For fitting to experimental data, the simulated SAXS curves were convolved with the experimental beam-length profile and the detector integration profile, so as to match the smearing effects due to the line-collimation geometry used in the experimental measurements. For each calculated curve, corresponding to a different value of ΔG_{soln} , a single adjustable parameter was introduced to scale the total scattering intensity, thereby accounting for differences in protein concentration. The experimental data were weighted according to the relative uncertainties estimated from counting statistics. To compare the individual fits to a given experimental data set, the reduced sum of squares of residuals was calculated as:

$$\chi_v^2 = \frac{1}{v} \sum_{i=1}^n \frac{(Y_i - Y_{i,\text{fit}})^2}{\sigma_i^2}, \quad (8)$$

where n is the number of data points; Y_i and $Y_{i,\text{fit}}$ are the experimentally observed and fit values for each data point; σ_i is the estimated uncertainty for that point; v is the number of degrees of freedom in the fit, $n - 2$ in this case.⁶⁸ Because the intensities recorded by the phosphor image plates are arbitrarily scaled and do not represent absolute photon counts, the uncertainties, σ_i , are also not absolute. As a consequence, χ_v^2 calculated from these measurements can only be interpreted as a relative measure of the goodness of fit.

All of the local software used for the SAXS analysis and simulations is available from Goldenberg on request.

Acknowledgments

The authors thank Andrew Steiner for his contributions to the development of the dihedral sampling algorithm, Prof. Peter von Hippel for the λ N expression plasmid, Profs. David and Jane Richardson for the dihedral angle distribution data from Ref. 42, and Drs. Don Parkin and Cy M.J. Jeffries for assistance with the SAXS measurements.

References

1. Rose G, Ed. (2002) Unfolded proteins, Advances in protein chemistry, Vol. 62. New York: Academic Press.
2. Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104: 3607–3622.

3. Shi Z, Chen K, Liu Z, Kallenbach NR (2006) Conformation of the backbone in unfolded proteins. *Chem Rev* 106:1877–1897.
4. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. *Prot Sci* 11:739–756.
5. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
6. Dunker AK, Silman I, Uversky VN, Sussman JL (2008) Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* 18:756–764.
7. Gsponder J, Babu MM (2009) The rules of disorder or why disorder rules. *Prog Biophys Mol Biol* 99:94–103.
8. Tompa P (2010) Structure and function of intrinsically disordered proteins. Boca Raton, FL: Chapman & Hall.
9. Tanford C (1968) Protein denaturation. *Adv Prot Chem* 23:121–282.
10. Flory PJ (1949) The configuration of real polymer chains. *J Chem Phys* 17:303–310.
11. Flory PJ (1953) Principles of polymer chemistry. Ithaca, NY: Cornell Univ. Press.
12. Millett IS, Doniach S, Plaxco KW (2002) Towards a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Adv Prot Chem* 62:241–262.
13. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiagarajan P, Sosnick TR, Hasan M, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 101:12491–12496.
14. Uversky VN, Gillespie JR, Millet IS, Khodyakova AV, Vasiliev AM, Chernovskaya TV, Vasilenko RN, Kozlovskaya GD, Dolgikh DA, Fink AL, Doniach S, Abramov VM (1999) Natively unfolded human prothymosin α adopts partially collapsed conformation at acidic pH. *Biochemistry* 38:15009–15016.
15. Uversky VN, Li J, Souillac P, Millett IS, Doniach S, Jakes R, Goedert M, Fink AL (2002) Biophysical properties of the synucleins and their propensities to fibrillate. *J Biol Chem* 277:11970–11978.
16. Permyakov SE, Millett IS, Doniach S, Permyakov EA, Uversky VN (2003) Natively unfolded C-terminal domain of caldesmon remains substantially unstructured after the effective binding to calmodulin. *Proteins* 53:855–862.
17. Uversky VN (2002) What does it mean to be natively unfolded? *Eur J Biochem* 269:2–12.
18. Marsh JA, Forman-Kay JD (2010) Sequence determinants of compaction in intrinsically disordered proteins. *J Mol Biol* 98:2383–2390.
19. Choy WY, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE (2002) Distribution of molecular size within an unfolded state ensemble using small-angle x-ray scattering and pulse field gradient NMR techniques. *J Mol Biol* 316:101–112.
20. Goldenberg DP (2003) Computational simulation of the statistical properties of unfolded proteins. *J Mol Biol* 326:1615–1633.
21. Fitzkee NC, Rose GD (2004) Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci USA* 101:12497–12502.
22. Navon A, Ittah V, Landsman P, Scheraga HA, Haas E (2001) Distributions of intramolecular distances in the reduced and denatured states of bovine pancreatic ribonuclease A. Folding initiation structures in the C-terminal portions of the reduced protein. *Biochemistry* 40:105–118.
23. McCarney ER, Werner JH, Bernstein SL, Ruczinski I, Makarov DE, Goodwin PM, Plaxco KW (2005) Site-specific dimensions across a highly denatured protein: a single molecule study. *J Mol Biol* 352:672–682.
24. Guinier A (1939) La diffraction des rayons X aux très petits angles, application à l'étude de phénomènes ultramicroscopiques. *Ann Phys* 12:161–237.
25. Porod G, General theory. In: Glatter O, Kratky O, Eds. (1982) Small-angle X-ray scattering. London: Academic Press, pp 17–51.
26. Teixeira J (1988) Small-angle scattering by fractal systems. *J Appl Crystallogr* 21:781–785.
27. Dewey TG (1997) Fractals in molecular biophysics. Oxford, UK: Oxford Univ. Press. pp 20–58.
28. Wang Y, Trewhella J, Goldenberg DP (2008) Small-angle X-ray scattering of reduced ribonuclease A: effects of solution conditions and comparisons with a computational model of unfolded proteins. *J Mol Biol* 377:1576–1592.
29. Beaucage G (2008) Toward resolution of ambiguity for the unfolded state. *Biophys J* 95:503–509.
30. de Gennes PG (1979) Scaling concepts in polymer physics. Ithaca, NY: Cornell Univ. Press.
31. Schmidt PW, Use of scattering to determine the fractal dimension. In: Avnir D, Ed. (1989) The fractal approach to heterogeneous chemistry. Chichester, U.K.: Wiley, pp 67–79.
32. Beaucage G (1996) Small-angle scattering from polymeric mass fractals of arbitrary mass-fractal dimension. *J Appl Crystallogr* 29:134–136.
33. Lake JA (1967) An iterative method of slit-correcting small angle X-ray data. *Acta Crystallogr* 23:191–194.
34. Johansen D, Jeffries CMJ, Hammouda B, Trewhella J, Goldenberg DP (2011) Effects of macromolecular crowding on an intrinsically disordered protein characterized by small-angle neutron scattering with contrast matching. *Biophys J* 100:1120–1128.
35. Makhataadze GI, Clore GM, Gronenborn AM (1995) Solvent isotope effects and protein stability. *Nat Struct Biol* 2:852–855.
36. Kuhlman B, Raleigh DP (1998) Global analysis of the thermal and chemical denaturation of the N-terminal domain of the ribosomal protein L9 in H₂O and D₂O. Determination of the thermodynamic parameters, ΔH° , ΔS° , and ΔC_p , and evaluation of solvent isotope effects. *Protein Sci* 7:2405–2412.
37. Efimova YM, Haemers S, Wierczinski B, Norde W, van Well AA (2007) Stability of globular proteins in H₂O and D₂O. *Biopolymers* 85:264–273.
38. Kresheck GC, Schneider H, Scheraga HA (1965) The effect of D₂O on the thermal stability of proteins. Thermodynamic parameters for the transfer of model compounds from H₂O to D₂O. *J Phys Chem* 69:3132–3144.
39. Oas TG, Toone EJ (1997) Thermodynamic solvent isotope effects and molecular hydrophobicity. *Adv Biophys Chem* 6:1–52.
40. Cho Y, Sagle LB, Iimura S, Zhang Y, Kherb J, Chilkoti A, Scholtz JM, Cremer PS (2009) Hydrogen bonding of β -turn structure is stabilized in D₂O. *J Am Chem Soc* 131:15188–15193.
41. Gong H, Rose GD (2008) Assessing the solvent-dependent surface area of unfolded proteins using an ensemble model. *Proc Natl Acad Sci, USA* 105:3321–3326.
42. Lovell SC, Davis IW, Arendall WB, III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins* 50:437–450.
43. Svergun D, Barberato C, Koch MHJ (1995) CRYSOLE—a program to evaluate X-ray solution scattering of

- biological macromolecules from atomic coordinates. *J App Crystallogr* 28:768–773.
44. Bernadó P, Blanchard L, Timmins P, Marion D, Rulgrok RWH, Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle X-ray scattering. *Proc Natl Acad Sci USA* 102:17002–17007.
 45. Jha AK, Colubri A, Freed KF, Sosnick TR (2005) Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc Natl Acad Sci USA* 102:13099–13104.
 46. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129:5656–5664.
 47. Van Gilst MR, Rees WA, von Hippel PH (1997) Complexes of N antitermination protein of phage λ with specific and nonspecific RNA target sites on the nascent transcript. *Biochemistry* 36:1514–1524.
 48. Mogridge J, Legault P, Li J, Van Oene MD, Kay LE, Greenblatt J (1998) Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage λ N protein. *Mol Cell* 1:265–275.
 49. Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 157:105–132.
 50. Uversky VN, Gillespie JR, Fink AL (2000) Why are “natively unfolded” proteins unstructured under physiological conditions. *Proteins* 41:415–427.
 51. Wootton JC, Federhen S (1993) Statistics of local complexity in amino acid sequences and sequence databases. *Comput Chem* 17:149–163.
 52. Promponas VJ, Enright AJ, Tsoka S, Kreil DP, Leroy C, Hamodrakas S, Sander C, Ouzounis CA (2000) CAST: an iterative algorithm for the complexity analysis of sequence tracts. *Bioinformatics* 16:915–922.
 53. Rubinstein M, Colby RH (2003) *Polymer physics*. Oxford, UK: Oxford Univ. Press.
 54. Plaxco KW, Millett IS, Segel DJ, Doniach S, Baker D (1999) Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nat Struct Biol* 6:554–556.
 55. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* 18:16–26.
 56. England JL, Haran G (2011) Role of solvation effects in protein denaturation: from thermodynamics to single molecules and back. *Annu Rev Phys Chem* 62:257–277.
 57. Ziv G, Haran G (2009) Protein folding, protein collapse and Tanford’s transfer model: lessons from single-molecule FRET. *J Am Chem Soc* 131:2942–2947.
 58. Harrington WF, Sela M (1959) A comparison of the physical chemical properties of oxidized and reduced alkylated ribonuclease. *Biochim Biophys Acta* 31:427–434.
 59. Jacob J, Dothager RS, Thiyagarajan P, Sosnick TR (2007) Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J Mol Biol* 367:609–615.
 60. Müller-Spätth S, Soranno A, Hirschfeld V, Hofmann H, Rügger S, Raymond L, Nettels D, Schuler B (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 107:14609–14614.
 61. Rees WA, Weitzel SE, Yager TD, Das A, von Hippel PH (1996) Bacteriophage λ N protein alone can induce transcription antitermination *in vivo*. *Proc Natl Acad Sci, USA* 93:342–346.
 62. Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* 41:207–234.
 63. Jeffries CM, Whitten AE, Harris SP, Trehwella J (2008) Small-angle X-ray scattering reveals the N-terminal domain organization of cardiac myosin binding protein C. *J Mol Biol* 377:1186–1199.
 64. Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 273:283–298.
 65. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55:379–400.
 66. Chothia C (1976) The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 105:1–14.
 67. Svergun DI, Richard S, Koch MHJ, Sayers Z, Zaccai G (1998) Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc Natl Acad Sci USA* 95:2267–2272.
 68. Bevington PR, Robinson DK (2003) *Data reduction and error analysis for the physical sciences*, 3rd ed. New York: McGraw-Hill. pp 194–197.