Published in final edited form as:

J Mol Biol. 2008 April 11; 377(5): 1576–1592. doi:10.1016/j.jmb.2008.02.009.

Small-Angle X-ray Scattering of Reduced Ribonuclease A: Effects of Solution Conditions and Comparisons with a Computational Model of Unfolded Proteins

Yuanyuan Wang¹, Jill Trehwella^{1,2}, and David P. Goldenberg^{3,*}

¹ Department of Chemistry, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112-0840, USA

² School of Molecular and Microbial Biosciences, The University of Sydney, NSW 2006 Australia

³ Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112-0840, USA

Abstract

The disulfide-reduced form of bovine ribonuclease A (RNase A), with the Cys thiols irreversibly blocked, was characterized by small-angle x-ray scattering (SAXS). To help resolve the conflicting results and interpretations from previous studies of this model unfolded protein, scattering profiles were measured using a range of solution conditions and were compared with the profiles predicted by a computational model for a random-coil polypeptide. Analysis of the simulated and experimental profiles reveals that scattering intensities at intermediate angles, corresponding to interatomic distances in the range of 5–20 Å, are particularly sensitive to changes in solvation and can be used to assess the internal scaling behavior of the polypeptide chain, expressed as a mass fractal dimension, D_m . This region of the scattering curve is also much less sensitive to experimental artifacts than is the very small angle regime (the Guinier region) that has been more typically used to characterize unfolded proteins. The experimental SAXS profiles closely matched those predicted by the computational model assuming relatively small solvation energies. The scaling behavior of the polypeptide approaches that of a well solvated polymer under conditions where it has a large net charge and at high urea concentrations. At lower urea concentrations and neutral pH, the behavior of the chain approaches that expected for θ -conditions, where the effects of slightly unfavorable interactions with solvent balance those of excluded volume, leading to scaling behavior comparable to that of an idealized random walk chain. Though detectable, the shift towards more compact conformations at lower urea concentrations does not correspond to a transition to a globule state and is associated with little or no reduction in conformational entropy. This type of collapse, therefore, is unlikely to reduce greatly the conformational search for the native state.

Keywords

Ribonuclease A; unfolded proteins; small-angle x-ray scattering; scaling laws; fractal dimension; solvation energy

* Corresponding author. Email address: goldenberg@biology.utah.edu (David P. Goldenberg).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 Introduction

Fifty years ago, Anfinsen and colleagues first described the reduction of the disulfides of bovine ribonuclease A and the concomitant loss of catalytic activity(1). Subsequent studies showed that the reduced protein had the properties of a disordered polypeptide chain(2) and could efficiently reform the native disulfides and recover activity under appropriate oxidizing conditions(3;4). Despite the great influence of these classic experiments, the intervening decades have seen a remarkable degree of controversy regarding the physical properties of reduced RNase A. Published estimates of its radius of gyration, for instance, range from from 20 to 50 Å(5;6;7;8). Although most small-angle scattering, spectroscopic and hydrodynamic studies are consistent with a random-coil model for the unfolded protein, Förster resonance energy transfer (FRET) experiments suggest the presence of both local and long range non-random structure(9;10).

This lack of consensus regarding the properties of reduced RNase A is representative of the general controversy that continues to surround the subject of unfolded proteins (11;12;13;14;15;16;17). The early work of Tanford and colleagues demonstrated that the overall dimensions of proteins unfolded in denaturants generally display a power-law dependence on chain length that is consistent with the behavior of a random-coil polymer (18;19). This general observation has been confirmed repeatedly, for a variety of proteins, using both hydrodynamic and x-ray scattering methods (20;21;22). At the same time, however, there is spectroscopic evidence for localized structures in unfolded proteins under a variety of conditions (17). Specific structure in the unfolded state could have important implications for protein folding, perhaps providing a mechanism for limiting the search for the native conformation. In addition, it is now apparent that many polypeptides, or parts thereof, may remain in a highly disordered state in vivo, and the physical properties of these “intrinsically disordered” proteins could have biological consequences under normal and pathological conditions (23;24;25).

Several factors contribute to the difficulty in characterizing unfolded proteins. By its very nature, the unfolded state of a protein is an ensemble of rapidly interconverting conformations that is not amenable to precise structural characterization. Although a variety of spectroscopic methods, including high-resolution NMR, can be used to characterize these molecules, the observed spectroscopic signals and derived structural parameters generally represent averages over the ensemble. A further source of difficulty lies in the likelihood that the distribution of conformations in an unfolded ensemble will shift in response to changes in solution conditions. Unfolded proteins are often most easily studied in the presence of chemical denaturants, which act to enhance the solvation of the chain, and the properties observed under these conditions could be significantly different from those of the same polypeptide chain under physiological conditions. Solution pH, through its effect on the net charge of a polypeptide, may also affect the unfolded ensemble.

Although intrinsically-disordered polypeptides can be characterized under a wide variety of solution conditions, polypeptide chains that fold to native structures are especially difficult to characterize under physiological conditions, where the native state predominates at equilibrium. One way of circumventing this difficulty is to introduce covalent modifications that selectively destabilize the native state, thereby making the unfolded state predominant. The removal of disulfide bonds in proteins that require these covalent cross-links for their stability, as exemplified by RNase A, is one means of achieving this end. Other modifications used to promote unfolding include deletions of small segments (26) and oxidation of buried methionine residues (27).

In order to address the general question of how the properties of unfolded proteins are influenced by changes in solution conditions, we have developed a computational model to

predict the small-angle x-ray scattering (SAXS) profiles of unfolded proteins under different solvation conditions and have compared these predictions with experimental data for reduced RNase A in the presence of 1 to 6 M urea and at pH values from 3 to 9. To ensure that the protein was entirely free of disulfides, the Cys thiols of the reduced protein were irreversibly blocked. Under all of the conditions examined, the observed scattering curves were very similar to those predicted by the model, and there was evidence for a progressive shift towards more expanded conformations at low pH, where the protein has a large positive net charge, and at higher urea concentrations. The results are consistent with a random-coil model for this unfolded protein and provide new insights into the scaling behavior of unfolded polypeptides and their response to changes in solvation.

2 Results

2.1 A computational model of solvation effects on unfolded proteins

The results of most previous SAXS studies of unfolded proteins (14;22) have been analyzed using Guinier plots (28), which yield an estimate of the average radius of gyration, R_g , from the scattering at very small angles, typically corresponding to values of Q (the scattering vector magnitude) less than about $0.05 \text{ \AA}^{-1}$, where $Q = (4\pi \sin \theta)/\lambda$; θ is one half of the scattering angle; and λ is the x-ray wavelength. Additional information about the distribution of interatomic distances for a molecular ensemble can be obtained from the scattering intensities measured at larger angles, and the Kratky plot, discussed further below, is a traditional means of obtaining information from the entire scattering curve. A potentially more powerful approach, however, is to compare observed scattering curves with those predicted by explicit models of potential structures or ensembles of structures(29;30).

To facilitate the interpretation of experimental data for unfolded proteins, we have developed a simple computational method for generating ensembles of calculated structures based on the assumption of random-coil behavior(31). This model incorporates polypeptide stereochemistry and explicitly accounts for excluded volume effects, so that the chain is not allowed to cross itself. The calculated structures are based on the sequences of real proteins and include all non-hydrogen atoms. Individual conformations are generated by first setting all of the dihedral angles to random values and then adjusting these angles to minimize steric overlaps. No other energy terms are introduced to account for attractive interactions among atoms. Using the computer program DYANA (32), ensembles of 100,000 or more independent conformations can be readily generated and used to predict a variety of average parameters for the ensemble. The distributions of backbone dihedral angles observed in the calculated conformations are very similar to those observed when only local steric factors are considered, thus satisfying a widely-utilized definition of a random-coil ensemble (19;33).

In Figure 1, the predicted RMS radii of gyration of four proteins are plotted as a function of chain length and compared with experimental measurements for 28 unfolded proteins in urea or GuHCl, as recently compiled by Kohn et al.(22). These authors have noted that the experimental values follow closely a power-law relationship with an exponent of 0.598 ± 0.028 , as compared to 0.6 predicted for a random coil with excluded volume. The curve shown in Figure 1 is a fit of a power function to the simulated values, with a fit exponent of 0.58 ± 0.02 . In addition to predicting accurately the exponent of the power-law relationship, the simulations do a remarkably good job of predicting the actual values of the radii of gyration over a wide range of chain lengths, without introducing any adjustable scaling parameters.

These simulations are based on the implicit assumption that interactions between the polypeptide and its solvent are, on average, energetically equivalent to those between protein atoms. In the language of polymer statistics, the model assumes an “athermal solvent”, as compared to “good” or “poor” solvents, in which interactions with the polymer are either more

or less favorable than those within the chain (34). The good agreement between the results of the simulations and experiments performed in the presence of strong denaturants suggests that these conditions may approximate an athermal solvent for polypeptides.

In order to consider how changes in solvation may influence the properties of unfolded proteins, we have extended this treatment by using a weighted histogram analysis. Boltzmann weighting factors are calculated on the assumption that the solvation free energy of an individual conformation is proportional to its total accessible surface area, an assumption supported by a variety of experimental data (35;36;37;38;39;40;41;42;43;44). The Boltzmann factors are used to generate weighted distributions, from which the average properties of the ensemble can be calculated for conditions that either favor or disfavor solvation, as illustrated in Figures 2 and 3 for the case of RNase A.

Panel a of Figure 2 shows a histogram of the accessible surface areas (ASA) of approximately 450,000 calculated conformations of RNase A, represented by the filled circles. The mean accessible surface area of the calculated ensemble was 14,700 Å², as compared to 7,000 Å² for the native protein and 16,300 Å² for a fully extended chain. Rather than applying weights directly to this histogram, which would give rise to artifacts due to the inevitable sparse sampling at the extreme values, the histogram was replaced with a smooth function, as described in Materials and Methods. The physically possible minimum and maximum values of the accessible surface area were assumed to be 7,000 and 18,000 Å², respectively, and the fit curve was smoothed to zero at these values.

Boltzmann weighting factors for individual bins in the histogram were calculated according to:

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

where ASA_j is the accessible surface area associated with bin j ; ΔG_{solv} is the solvation free energy per unit of surface area; n is the total number of bins; and g_j is the number of conformations associated with the bin in the unweighted distribution. The weighted population in each bin is then calculated as $w_j g_j$. Weighted histograms corresponding to values of ΔG_{solv} from -4 to 4 cal/mol/Å² are shown on a logarithmic scale in Figure 2b, along with bars representing the original unsmoothed and unweighted histogram. Unfavorable solvation ($\Delta G_{\text{solv}} > 0$) is seen to shift the distribution towards conformations with less accessible surface area, while negative values of ΔG_{solv} favor larger surface areas.

An inherent limitation of weighted histogram methods such as used here is that they are based on the assumption that the original conformational sampling is adequate to identify representative conformations that may predominate under the extrapolated conditions, or that the properties of the sampled conformations can be accurately extrapolated (45;46;47). As shown in Figure 2b, the distribution of accessible surface areas appears to follow the decay of a Gaussian distribution over at least four orders of magnitude on either side of the maximum. For the weighted histogram corresponding to $\Delta G_{\text{solv}} = 4$ cal/mol/Å², approximately half of the distribution of ASA values is represented by the sampled conformations, albeit quite sparsely. Similarly, about half of the histogram for $\Delta G_{\text{solv}} = -4$ cal/mol/Å² overlaps with the sampled distribution. Although this cutoff is somewhat arbitrary, we consider these values of ΔG_{solv} to represent the limits of what can reasonably be extrapolated from the sampled conformations. Within this range, the properties derived from weighted histograms were found to be relatively insensitive to the exact procedure used to smooth the original histograms.

To calculate the effects of solvation on properties of the unfolded ensemble other than accessible surface area, two-dimensional histograms correlating ASA and the property of interest were used, as illustrated in Figure 3a for the radius of gyration. The figure also shows the values of these properties for the native structure and a fully extended chain. The sample includes conformations that are nearly as compact as the native protein, but none that approach that of a fully extended chain. The histogram was smoothed as described in Materials and Methods, to give the histogram in Figure 3b. Panels c to f of Figure 3 show histograms weighted according to Equation 1 to represent the effects of favorable or unfavorable interactions between the polypeptide chain and solvent. Since larger average dimensions are associated with larger accessible surfaces, favorable interactions with solvent lead to a shift towards conformations with larger radii of gyration (panels d and f), while unfavorable interactions lead to a shift towards more compact conformations (panels c and e). The weighted histograms can be used to calculate the average radii of gyration under different conditions, as illustrated in Figure 4a. Although the resulting curve has a sigmoidal shape suggestive of a protein unfolding transition, it is important to emphasize that this treatment is not meant to simulate a cooperative phase transition, but rather a progressive shift in the population of conformations in the broad ensemble. With larger positive values of ΔG_{solv} , it is possible that the chain would undergo a more cooperative transition to a globule state, but the limits of our initial ensemble preclude the extrapolations needed to explore this regime.

The contraction of the chain under conditions of less favorable solvation is also reflected in a decrease in the average distances between specific atoms, and the effects can be used to compare quantitatively the results of this numerical simulation with those of more traditional treatments of polymers. From polymer theory, the root-mean-square (RMS) distance between a pair of atoms separated by N residues is predicted to follow a power-law relationship:

$$\text{RMS}(R) = AN^{\nu} \quad (2)$$

where A is a numerical constant, and ν is often referred to as the Flory exponent (48;49). For an idealized random-flight chain that can cross itself, $\nu = 1/2$. For a real chain with excluded volume, the exponent is predicted to be $3/5$ under conditions of favorable solvation, but less than that value in the presence of a poor solvent. To calculate the change in interatomic distances due to solvation, histograms analogous to those shown in Figure 3 for the radius of gyration were constructed for the distances between the sulfur atoms of six pairs of Cys residues, chosen to represent chain segment lengths from 7 to 84 residues. (Cys pairs were chosen because of the direct relationship between these distance distributions and the propensity to form disulfide bonds. (31)) The histograms were smoothed and weighted according to solvation free energies ranging from -4 to 4 cal/mol/Å². In Figure 4b, the average RMS distances are plotted as a function of the segment length for five values of ΔG_{solv} . Each set of RMS distances was well fit by a power function, and the fit values of the exponent, ν , are plotted as a function of ΔG_{solv} in Figure 4c.

For the unweighted distribution ($\Delta G_{solv} = 0$), the exponent was 0.56, a value intermediate between those expected for a random-flight chain and a well-solvated chain with excluded volume. As expected, the value of the exponent increased with negative values of ΔG_{solv} and decreased with positive values. For $\Delta G_{solv} = 2$ cal/mol/Å², $\nu = 0.5$. Under these conditions, unfavorable solvation balances the repulsive force due to excluded volume, and the chain's statistical properties match those of an ideal random-flight chain. This corresponds to the θ -condition defined by Flory. At negative values of ΔG_{solv} the exponent appears to approach a value of about 0.61, slightly larger than that predicted by the Flory theory. The computational

model thus appears to capture the important features generally expected of a disordered chain under different solvent conditions.

2.2 Simulation of small-angle x-ray scattering from unfolded proteins

The atomic coordinates of the calculated conformations were also used to predict directly the SAXS curves expected for ensembles under different conditions, as illustrated in Figure 5. For this calculation, the program CRY SOL was used to calculate the expected scattering curve for each of approximately 45,000 conformations of the RNase A sequence. In panel a, the unweighted average of these scattering curves, plotted as $\log(I)$ versus Q , is represented by the curve marked with filled circles, while the scattering curve calculated for native RNase A is marked with filled diamonds. Curves corresponding to positive and negative values of ΔG_{solv} are marked with filled or open symbols as indicated by the key.

In considering the simulated curves shown in Figure 5a, it is useful to divide the scattering profile into three regions, corresponding to very small, intermediate and large values of Q (50; 51). The differences that emerge in these regimes can be highlighted using different graphical methods. At sufficiently small angles, the x-ray scattering intensity from a particle of any shape can be approximated by a Gaussian function:

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

where K is the square of the scattering density difference between the particle and the solvent (28). The scattering behavior in this regime is readily visualized in the widely-used Guinier plot, in which $\ln(I)$ is plotted as a function of Q^2 , as illustrated in Figure 5b using the calculated scattering data for native RNase A and various unfolded ensembles. The approximation of Equation 3 is valid for $R_g \cdot Q$ less than about 1, depending somewhat on the shape and uniformity of the scattering particles. For native RNase A, $R_g = 14.3 \text{ \AA}$, so that the linear region of the plot is expected to extend to approximately $Q = 0.07 \text{ \AA}^{-1}$, or $Q^2 \approx 0.005 \text{ \AA}^{-2}$, as observed in the calculated curve for this protein. For the ensembles of calculated unfolded conformations, however, there is clear curvature in the plots at values of Q^2 greater than about $0.002 \text{ \AA}^{-2}$, or less for the more expanded ensembles. The simulations thus highlight the point that the Guinier analysis must be based on accurate data at quite small scattering angles and utilizes only a fraction of the available data. The data from the smallest scattering angles are also particularly sensitive to the effects of aggregation and interparticle interference, which may further complicate the analysis.

At scattering angles beyond the Guinier region, the scattering intensity typically decreases more rapidly and can often be characterized by a power-law function of Q (52). For a compact globular structure, the scattering at larger angles decays with the fourth power of Q (52). In contrast, the scattering from an idealized random-flight chain, is predicted to decrease with the second power of Q (53). For a chain in which the orientations of bond segments are correlated, such as the “worm-like” chain of Porod and Kratky, the scattering behavior of the random-flight chain is preserved at small and intermediate angles, reflecting the random orientation of chain segments over relatively large distances. At larger angles, however, the local structure of the chain segments dominate scattering. In this regime, the chain can be approximated as a collection of randomly oriented thin rods, and the scattering intensity is predicted to decrease with the first power of Q (51;54).

The power-law dependence of scattering in the intermediate- and large- Q regimes can be visualized by plotting $\log(I)$ versus $\log(Q)$, as shown in Figure 6a for the simulated scattering data for reduced RNase A with different values of ΔG_{solv} . Qualitatively, the calculated log-

log plots are similar to those predicted for random-flight chains, with a nearly linear region for values of $\log(Q)$ greater than about -1.2 ($Q > 0.06 \text{ \AA}^{-2}$). The plot shows that the intermediate- Q region is sensitive to solvation, with the slope of the linear segment becoming more negative with more positive values of ΔG_{solv} . The linear region is also shifted towards higher $\log(Q)$ values as ΔG_{solv} increases. The inset to Figure 6a shows the linear segments on a more expanded scale, and the slopes derived from these segments are plotted as filled circles in panel b (as the negative values).

The exponent in the power-law dependence of the scattering intensity can be interpreted as the negative of a mass fractal dimension, D_m , which relates the mass enclosed by a volume to the linear dimensions of that volume (55;56;57;58). For a polymer, the fractal dimension also corresponds to the inverse of the Flory exponent. Thus, the fractal dimension for an idealized random-flight chain is 2, corresponding to the slope of -2 in the linear portion of the $\log(I)$ versus $\log(Q)$ plot. For real polymers, however, the excluded volume effect changes the scaling between mass density and linear dimensions, resulting in a reduced fractal dimension and a less negative slope in the log-log scattering curve. For a chain with excluded volume in a good solvent, the fractal dimension is predicted to approach 5/3 (58). On the other hand, the compact conformations favored by a poor solvent are expected to have an increased fractal dimension, possibly approaching 3, the value predicted for a collapsed globule.

As illustrated by the filled circles in Figure 6b, the values of D_m derived from the calculated scattering curves are about 1.7 for $\Delta G_{\text{solv}} \leq -2 \text{ cal/mol/\AA}^2$, in agreement with the prediction for a well-solvated polymer, and increase to values greater than 2 for unfavorable solvation. The open squares in the figure represent the inverses of the Flory exponent derived from the dependence of the RMS interatomic distance on chain segment length (Figure 4). For values of ΔG_{solv} from -3 to 3 cal/mol/ $\text{\AA}^2$, there is good agreement between the values of D_m obtained by the two methods. While the two sets of values are derived from the same model, it should be noted that they reflect distance correlations on distinct scales: The intermediate Q -regime of the scattering curves corresponds to interatomic distances of about 5–20 $\text{\AA}$, while the average distances plotted in Figure 4 cover the range from 20–80 $\text{\AA}$. Thus, the scaling properties appear to hold over a wide range of distances and for a range of solvation conditions.

Another commonly used method for analyzing scattering from polymers is a plot of the product IQ^2 versus Q , the Kratky plot (51;54), as illustrated in Figure 7 for the simulated scattering data. This plot is particularly useful for distinguishing the scattering profiles of polymers from those of compact globular structures. Within the intermediate- Q regime where the scattering from a polymer decays with approximately the second power of Q , the product IQ^2 remains nearly constant, leading to a horizontal region in the Kratky plot. In contrast, the scattering at intermediate and large angles from a globular structure decays with the fourth power of Q , contributing to a pronounced peak in the plot, as illustrated by the calculated curve for native RNase A. The Kratky plot also serves to highlight the differences in the predicted curves in the intermediate Q range for the unfolded protein under different solvent conditions.

This analysis of the predicted scattering curves suggests that important information about unfolded proteins can be found in the intermediate region of the SAXS curves, which reflect the density of chain segments on the scale of about 5–20 $\text{\AA}$. This range of distances is likely to be particularly important in defining the response of the polypeptide chain to changes in solvent.

2.3 Experimental scattering curves for reduced RNase A

Samples for SAXS measurements were prepared by reducing the disulfides of native RNase A under strongly denaturing conditions and reacting the resulting thiols with either iodoacetate or iodoacetamide, to generate carboxymethylated (RCM) and carboxyamidomethylated

(RCAM) forms. The final samples contained 1, 2, 4 or 6 M urea, and the pH was adjusted to 3, 5, 7 or 9 with the addition of HCl or NaOH.

Scattering data for RCAM-RNase A in solutions of 1–6 M urea at pH 7 are shown in Figure 8, where the data are plotted as $\log(I)$ versus Q so that differences in total scattering intensity, reflecting variations in sample concentration, affect only the vertical displacement of the curves. The experimental data were compared to the predictions of our simulations by directly fitting the scattering data to calculated curves, with the only adjustable parameter being a vertical offset. For each experimental curve, the observed data were fit to each of 9 calculated curves, corresponding to values of ΔG_{solv} from -4 to 4 cal/mol/Å², and the resulting residuals (reduced $\chi^2 = \chi_v^2$) were compared to identify the curve that best fit the experimental data. Plots of χ_v^2 versus ΔG_{solv} are shown in the insets to Figure 8. All of the experimental scattering curves were well fit by calculated curves, using values of ΔG_{solv} close to zero.

For RCAM-RNase A in 1 M urea at pH 7, the scattering curve calculated for $\Delta G_{solv} = 1$ cal/mol/Å² yielded the best fit to the experimental data, as represented by the solid curve in Figure 8a. Although values 2 cal/mol/Å² less than or greater than this value gave fits that were nearly as good (as represented by the dashed curves), the plot of χ_v^2 versus ΔG_{solv} shows a distinct minimum, indicating that the experimental data are able to distinguish among the curves calculated assuming different solvation free energies. For 2 and 4 M urea (panels b and c of Figure 8), optimal fits were obtained with $\Delta G_{solv} = 0$ and -1 cal/mol/Å², respectively. For 6 M urea, however, there was not a distinct minimum in the χ_v^2 plot, and the data could be well fit by any of the curves corresponding to $\Delta G_{solv} < 0$.

The plots of χ_v^2 versus ΔG_{solv} indicate that there is a systematic shift in the scattering curves with urea concentration. Since the differences in the calculated curves are most pronounced in the intermediate Q regime, this region of the experimental curves was examined further in plots of $\log(I)$ versus $\log(Q)$, as illustrated in Figure 9. Within the range of Q values from 0.063 to 0.16, each of the log-log scattering curves appeared to be quite linear. Least-squares fits to the data for the pH 7 samples yielded slopes that ranged progressively from -1.88 ± 0.01 , for 1 M urea, to -1.65 ± 0.01 for 6 M urea. Though small, the differences in slope lie well outside of the uncertainties estimated from the least-squares fits. This trend is consistent with that shown by the fits to the entire scattering curves and suggests that there is a small but measurable expansion of the chain with increased urea concentration.

In contrast to the trend seen for the samples at pH 7, there was very little difference among the scattering curves measured for RCAM-RNase A at pH 3 and different urea concentrations, as shown in the log-log plots in Figure 9b. For these samples, the slopes were all in the range of -1.56 to -1.59, again with uncertainties of about ± 0.01 .

For each of the solution conditions examined, the scattering curves for the two forms of reduced RNase A were analyzed using log-log plots, and the slopes derived from these plots (expressed as the mass fractal dimension, $D_m = -\text{slope}$) are graphed as a function of urea concentration in Figure 10. For RCAM-RNase A at pH 3 and 5, there was little or no change in D_m with urea concentration, and a similar pattern was observed for the RCM form at pH 3. In contrast, for the RCAM form at pH 7 and pH 9, as well as the RCM form at pH 5, there is a clear trend of decreased values of D_m at higher urea concentrations. The patterns are less clear for the RCM form at pH 7 and 9, where there was a decrease in D_m between 1 and 2 M urea, but little or no further change at higher concentrations.

The patterns observed for the two proteins appear to be correlated with their predicted net charges (shown in parenthesis for each pH value in Figure 10). The three cases for which little

or no urea dependence was observed were also those with the largest predicted net charges: 18.2 (RCAM, pH 3), 17.9 (RCM, pH 3) and 9.9 (RCAM pH 5). On the other hand, the strongest dependence on urea concentration was observed for proteins with net charges of 1.6 (RCAM, pH 9), 3.5 (RCM, pH 5), and 4.9 (RCAM, pH 7). The net negative charges of RCM RNase A at pH 7 and 9 (-3.1 and -6.4) were associated with an intermediate degree of sensitivity to urea.

These results suggest that a high density of positive charge leads to a conformational ensemble that is, on average, slightly more expanded than predicted by our simulations assuming an athermal solvent. With this degree of expansion, most of the conformations would already have quite large accessible surface areas, and increased urea concentrations would not be expected to cause a significant change in the distribution of conformations (panels d and f of Figure 3). Lower charge densities, on the other hand, appear to be associated with more compact ensembles, as indicated by larger values of D_m , and greater sensitivity to urea concentration.

The experimental data were also examined using Guinier plots, which indicated that the RMS radius of gyration of the reduced proteins lay in the range of 30–40 Å under all of the conditions examined. Unfortunately, the very small-angle scattering data used for the Guinier analysis were heavily influenced by parasitic scattering from the beam stop, due to an instrument malfunction described in the Materials and Methods section. As a consequence, there were large uncertainties in the estimates of radii of gyration, and it was impossible to establish any correlation between these values and the solution conditions. Importantly, however, the parasitic scattering was limited to Q less than about 0.025 Å^{-1} , and the intermediate- Q regime data used to estimate the fractal dimension were unaffected by this artifact.

3 Discussion

3.1 The properties of reduced and unfolded RNase A

The work presented here extends several previous SAXS studies(5;7;22;8) of reduced RNase A in two important ways. First, the thiols of the reduced protein were irreversibly blocked to prevent the reformation of disulfides and allowing measurements at a wide range of pH values. Second, the experimental scattering curves were directly compared with those predicted by an explicit model for the unfolded ensemble, using data over the full range of scattering angles. The experimental data were found to be in excellent agreement with the scattering curves predicted assuming relatively small solvation energies, *i.e.* $\Delta G_{solv} = -2$ to 2 cal/mol/Å^2 . Data from the intermediate Q regime suggest that more expanded conformations are favored by higher urea concentrations and pH values that lead to high charge densities on the polypeptide chain.

The present results are largely consistent with those recently described by Jacob et al., who carried out a similar SAXS study of reduced (but otherwise unmodified) RNase A at pH 2.5, in the presence of 0 to 6 M GuHCl and at temperatures ranging from 20 to 90 °C(8). Using Guinier analysis, these authors estimated the radius of gyration to be 34–35 Å over the full range of temperatures and GuHCl concentrations examined. This value is very close to both the value predicted by our simulations for an athermal solvent (36 Å) and the value predicted from the regression of experimental measurements of other proteins in denaturants (Figure 1 and ref. (22)). From these data, Jacob et al. argue that reduced RNase is best described as a fully expanded random coil even in the absence of denaturants. These results are comparable to those obtained at pH 3 in the present study, conditions where the protein has a large positive net charge (+18) and the scattering curve was insensitive to denaturant concentration. The absence of an observed effect of GuHCl on the radius of gyration in the study by Jacob et al. is also likely a reflection of the low pH used in those experiments.

The SAXS data argue strongly against a highly compact unfolded ensemble under any of the conditions examined. The best matches between the experimental data and the computational model were found using calculated ensembles for which the RMS radius of gyration is in the range of 30–40 Å, corresponding to panels b, c and d of Figure 3. It should be noted that the conclusions regarding the average dimensions of the ensemble are not dependent on the details of the solvation model, which can be viewed as simply a mechanism for generating ensembles with different size distributions for comparison with the experimental data.

Although the simulations based on a random-coil model predict the observed SAXS profiles quite well, this agreement does not rule out the possibility of local non-random structure in the unfolded state, a point that has been discussed in several recent papers (22;14;59;60;29;61;17;62). It is also difficult, from the SAXS data alone, to rule out the possibility that a small fraction of molecules have compact native-like structures, as simulations indicate that the concentration of such a species would have to be about 10% before it would significantly alter the scattering curve. It is possible that such structures may account for the time-resolved FRET measurements of reduced RNase A, which indicate the presence of a collapsed structure in the C-terminal half of the polypeptide (9;10). On the other hand, the far-UV circular dichroism spectrum of RCM-RNase A shows no evidence of regular secondary structure, indicating that the level of any residual structure must be quite low(63).

3.2 Scaling properties and fractal dimensions of unfolded proteins

Scaling relationships play a central role in the theory of disordered polymers and are often used as criteria for classifying the behavior of polymers under different solution conditions (64;49;58;60;65;66). Most applications of this theory to the characterization of unfolded proteins have focused on the scaling between the polypeptide length and the average overall dimensions of the chain (20;21;22). As illustrated in Figure 1, SAXS measurements for a large number of proteins unfolded in strong denaturants follow the predicted relationship with the Flory exponent, ν , close to 0.6, the predicted limiting value for a chain with excluded volume in a good solvent. The same scaling exponent is predicted to describe the RMS distances between pairs of atoms within the same chain, as expressed by Equation 2.

Closely related to the Flory scaling exponent, the mass fractal dimension is also a concept widely used in polymer science (34). In the limit of a chain much longer than its monomers, a random-flight chain has the properties of a fractal object, *i.e.* it is self-similar over a wide range of scales. The mass dimension of a fractal, D_m , is the scaling exponent in the relationship between the mass, m , of a portion of the object and the linear size of a volume enclosing that mass (*e.g.* the radius of a sphere, r):

$$m \propto r^{D_m} \quad (4)$$

Thus, D_m represents the efficiency with which an object fills a volume and is constant over all scales for a true fractal object. For a solid, $D_m = 3$, while $D_m = 2$ for a random flight chain and $D_m = 1$ for a thin rod. For a real chain, the range over which fractal behavior can be observed is necessarily limited by the finite dimensions of the polymer repeat unit and the overall dimension of the chain. For our computational model of unfolded RNase A, the scaling relationship appears to hold over the range from 5 Å (from the computed scattering curves, Figure 5) to 80 Å (from the calculated RMS distances between atoms, Figure 4), with values of D_m ranging from 1.7 to greater than 2, depending on the energetics of solvation.

Small-angle scattering experiments provide a means of determining the fractal dimension of a polymer or, equivalently, the Flory exponent averaged over all of the atom pairs in an ensemble.

This relationship between scattering in the intermediate- Q regime and scaling behavior has been recognized since the classic treatments by Debye(53) and by Kratky and Porod (50), and it has been used extensively in the characterization of synthetic polymers (55;67;57;68;69). However, this relationship does not appear to have been used previously in the analysis of unfolded proteins, other than in a study of protein-detergent complexes (70).

The experimental data for reduced RNase A indicate that this polypeptide has a fractal dimension ranging from 1.6 to 2, with the larger value observed at low urea concentration and neutral pH. The former value is slightly smaller than expected for the limiting case of a “good solvent”, while a fractal dimension of 2 corresponds to the θ -condition, where the balance between poor solvation and excluded volume leads to behavior like that of an idealized random-flight chain. Thus, even the most unfavorable solvent conditions examined here still lead to scaling relationships corresponding to a coil state, as opposed to a collapsed globule, for which D_m is expected to approach the value 3.

3.3 Possible implications for the energetics and mechanism of protein folding

A general hydrophobic collapse has frequently been hypothesized to be the initial step in the refolding of proteins diluted from denaturant solutions, and experimental evidence for such a collapse has been reported for several proteins(71;72;73;74;75;76;77;78). The model and experimental data presented here are consistent with a shift in the unfolded ensemble, but they suggest that the energetic consequences may be relatively small.

For RNase A at neutral pH, our data and model suggest that $\Delta G_{solv} \approx -2$ cal/mol/Å² in the presence of 6 M urea and that the average accessible surface area is approximately 15,500 Å², for a total ΔG_{solv} of -31 kcal/mol. Under the conditions corresponding to 1 M urea, $\Delta G_{solv} \approx 1.5$ cal/mol/Å², and the model predicts a total ASA of about 14,000 Å², for a total ΔG_{solv} of 21 kcal/mol. Thus, decreasing the urea concentration from 6 to 1 M is predicted to lead to an increase of 52 kcal/mol in the total solvation energy of the unfolded state, with the collapse having only a small compensating effect. For the native protein, with an accessible surface area of approximately 7,000 Å², the total solvation energy would increase by about 24 kcal/mol.

Collapse of a polypeptide chain upon dilution from a denaturant solution is generally assumed to lead to a reduction of conformational entropy. The magnitude of this effect, ΔS_{conf} , can be estimated from the calculated conformational distribution and the following relationship (derived in Materials and Methods):

$$\Delta S_{conf} = -R \sum_{j=1}^n g_j w_j \ln w_j - R \ln \sum_{j=1}^n g_j \quad (5)$$

Here, ΔS_{conf} is defined as the entropy change associated with the transfer of the polypeptide from an athermal solvent to conditions defined by the Boltzmann weighting factors, w_j , calculated according to Equation 1 for a given value of ΔG_{solv} . R is the gas constant, and g_j is the number of conformations contained in bin j in the unweighted histogram. In Figure 11, the predicted change in conformational entropy for reduced and unfolded RNase A is plotted as a function of ΔG_{solv} , showing that for either positive or negative values of ΔG_{solv} , ΔS_{conf} is negative. For the particular values of ΔG_{solv} estimated for 1 and 6 M urea, 1.5 and -2 cal/mol/Å², the calculated values of ΔS_{conf} are both approximately -2.5 cal/deg-mol, predicting that there would actually be no change in conformational entropy associated with the change in urea concentration!

Although the exact value of the entropy change is sensitive to the assumed values of ΔG_{solv} , a relatively small effect is predicted for any of the conditions considered by this treatment. For instance, a change of ΔG_{solv} from 0 to 4 cal/mol/Å² is predicted to cause an entropy decrease of about 18 cal/deg-mol, or about 0.15 cal/deg-mol per residue. This is much smaller than the 4 cal/deg-mol per residue typically estimated for the change in conformational entropy for complete folding (see ref. (79) and references therein). This result suggests that a hydrophobic collapse of the unfolded state, of the type simulated by our model and observed experimentally for RNase A, is unlikely, by itself, to be sufficient to greatly narrow the search for the native conformation. For some sequences and conditions, however, it is possible that a more extreme collapse to a globule state would result in a significant reduction in entropy.

4 Materials and Methods

4.1 Computational methods and data analysis

The strategy used to generate ensembles of calculated conformations has been described previously (31). Briefly, structures were generated by initially setting all backbone and dihedral angles to random values and then adjusting the dihedral angles to minimize the steric overlap between all atom pairs, using the computer program DYANA (32). Because the DYANA “energy” function, or target function, utilizes a relatively soft repulsion potential, the atomic radii used for the calculations were expanded by 10%, thereby generating a distribution of backbone dihedral angles similar to that of the classic Ramachandran plot. To avoid using structures trapped in especially disfavored conformations, the 10% of structures with the largest target function values were discarded.

Accessible surface areas of the individual conformations were calculated using the algorithm of Lee and Richards (80), as implemented in the program ACCESS by T.J. Richmond, and the group radii of Chothia(81). The probe radius was 1.4 Å. Radii of gyration were calculated from the coordinates of the backbone α -carbons. Small-angle x-ray scattering profiles were calculated from atomic coordinates using the program CRY SOL with the default parameters, including a bulk solvent density of 0.334 electrons/Å³, a hydration layer 3 Å thick and a scattering contrast of 0.03 electrons/Å³ for the hydration layer. (82;83).

4.1.1 Weighted Histogram—Before applying Boltzmann weighting factors to the histograms derived directly from the calculated ensembles, the distributions of accessible surface areas, radii of gyration and interatomic distances were smoothed to avoid artifacts. For the histogram of ASA shown in Figure 2, the values corresponding to $ASA \leq 14425 \text{ Å}^2$ and those with $ASA \geq 14,470 \text{ Å}^2$ were fit separately to a Gaussian function. The two fit functions were then joined and the values for ASA less than 7,275 Å² or greater than 17,700 Å² were set to zero. A sliding average, with a window of 4 data points, was used to smooth the histogram function. The two-dimensional histogram shown in Figure 3a was smoothed in two steps: First a 10×10 Gaussian filter was applied to the 10,000 histogram bins. To constrain the histogram to physically plausible dimensions, values corresponding to bins with $ASA > 16,000 \text{ Å}^2$ or $R_g < 16 \text{ Å}$ were set to zero, as were those populated with less than 1 chain per 10⁶. The histogram was then re-smoothed with a 10×10 Gaussian filter, to generate the histogram shown in Figure 3b. A similar method was used to smooth histograms correlating the accessible surface area with interatomic distances, and used to generate the data of Figure 4b, except that 5×5 Gaussian filters were used for smoothing, and a lower distance cutoff of 2.5 Å was applied.

4.1.2 Calculation of entropy changes—The entropy change associated with a change in solvation free energy was calculated by treating the individual conformations in the calculated ensembles as representative microstates. From the Boltzmann definition, the entropy of the ensemble is calculated as:

$$S = -R \sum_{i=1}^N p_i \ln p_i \quad (6)$$

where p_i is the probability of microstate i , and N is the total number of microstates. The probability of each microstate is determined by the solvation free energy according to:

$$p_i = \frac{e^{-\Delta G_{\text{solv}} \cdot \text{ASA}_i / RT}}{\sum_{i=1}^N e^{-\Delta G_{\text{solv}} \cdot \text{ASA}_i / RT}} \quad (7)$$

The reference state is taken to be the initial unweighted distribution, for which $\Delta G_{\text{solv}} = 0$. For this distribution, the probabilities are all equal, and $p_i = 1/N$. Therefore, the reference-state entropy is given by:

$$\begin{aligned} S_0 &= -R \sum_{i=1}^N \frac{1}{N} \ln \frac{1}{N} \\ &= R \ln N \end{aligned} \quad (8)$$

The entropy change for the transfer of an unfolded chain from this condition to one characterized by a non-zero ΔG_{solv} is then given by:

$$\begin{aligned} \Delta S_{\text{conf}} &= S - S_0 \\ &= -R \sum_{i=1}^N p_i \ln p_i - R \ln N \end{aligned} \quad (9)$$

Rather than using this relationship and the calculated ensemble directly, the distribution was represented by the smoothed histogram of ASA values shown in Figure 2. In this histogram, the N calculated conformations are divided into n bins, with each bin containing g_j conformations, all with approximately the same accessible surface area and, therefore, the same solvation free energy and probability. The weighting factors calculated according to Equation 1 represent the individual probabilities of the conformations in the corresponding bins. Equation 6 can, therefore, be written in terms of the histogram parameters as:

$$S = -R \sum_{j=1}^n g_j w_j \ln w_j \quad (10)$$

For the reference state, where $\Delta G_{\text{solv}} = 0$,

$$w_j = \frac{1}{\sum_{j=1}^n g_j} \quad (11)$$

and

$$\begin{aligned}
 S_0 &= -R \sum_{j=1}^n \frac{g_j}{\sum_{j=1}^n g_j} \ln \frac{1}{\sum_{j=1}^n g_j} \\
 &= R \ln \sum_{j=1}^n g_j
 \end{aligned}
 \tag{12}$$

Therefore, the entropy change is:

$$\begin{aligned}
 \Delta S_{\text{conf}} &= S - S_0 \\
 &= -R \sum_{j=1}^n g_j w_j \ln w_j - R \ln \sum_{j=1}^n g_j
 \end{aligned}
 \tag{13}$$

4.1.3 Small-angle scattering curves—Simulated small-angle scattering curves were calculated for approximately 45,000 conformations of RNase A and were weighted directly using Boltzmann factors calculated according to Equation 1. Because the experimental data were more finely spaced than the calculated scattering intensities, the simulated data were first fit to 6th-degree polynomials, as shown by the curves in Figure 5a. The resulting polynomials were, in turn, fit to the experimental data using the method of least squares and allowing only the constant term to float. The experimental data were weighted according to the standard errors calculated from counting statistics. Residuals were calculated as:

$$\chi_v^2 = \frac{1}{v} \sum_{i=1}^n \frac{(Y_i - Y_{i,\text{fit}})^2}{\sigma_i^2}
 \tag{14}$$

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 standard error for that point; and v is the number of degrees of freedom in the fit, $n - 2$ in this case (84).

All curve fitting and data analysis operations were performed using the computer program Igor Pro 5.0 (WaveMetrics, Inc.)

4.2 Preparation of reduced and alkylated RNase A

Bovine ribonuclease A (type XII-A) was purchased from Sigma Chemical Co. and used without further purification. Lyophilized native protein was dissolved to a final concentration of approximately 1 mg/mL in a solution also containing 6 M GuHCl, 0.1M Tris-HCl pH 8, 10 mM EDTA and 10 mM dithiothreitol. This solution was incubated for approximately 1 h at room temperature to unfold and reduce the protein. The free thiols were then blocked by adding either iodoacetamide (for RCAM-RNase A) or Na-iodoacetate (for RCM-RNase A) to a final concentration of 30 mM. The reaction was incubated for another 30 min at room temperature, and the Tris buffer was then neutralized by adding 1 equivalent of HCl.

The samples were dialyzed against 0.01 M HCl and then against a solution of 6 M urea adjusted to pH 2.5. Samples were concentrated to approximately 5 mg/mL using Amicon Ultra-15 centrifugal filter units with Ultracel-10 cellulose membranes (10,000 Dalton cutoff). Samples were stored at 4 °C in this solution until just prior to making scattering measurements. The pH of the solutions were adjusted by adding small amounts of HCl and measured using a glass electrode, relying on the protein to act as a pH buffer. The final urea concentration was adjusted by diluting the sample with water, and then the samples were reconcentrated to 5 mg/mL by

centrifugation. Concentrations were determined by measuring UV absorbance at 280 nm, using an extinction coefficient of $9,600 \text{ M}^{-1}\text{cm}^{-1}$. The filtrate from the final concentration step was used as the reference sample for SAXS measurements.

4.3 Small-angle x-ray scattering

SAXS data presented here were acquired at the Stanford Synchrotron Radiation Laboratory (SSRL) using Beam Line 4-2L. A polycarbonate sample cell equipped with $25 \mu\text{m}$ thick mica windows providing a 1 mm path length cell was precisely and reproducibly inserted into the sample holder maintained at 25°C . Approximately $30 \mu\text{L}$ of sample was required to fill the cell. Synchrotron X-ray radiation with a wavelength of $1.378 \text{ \AA}$ was incident in a pinhole configuration to give a full width at half-maximum on the sample of $0.2 \times 1.5 \text{ mm}$. The scattering due to the protein was obtained as a difference between the sample (protein in solvent) and solvent scattered intensity profile normalized for the integrated incident beam intensity and scaled for the sample and buffer transmissions monitored on a separate detector channel. All data were recorded automatically using the standard facilities at beam line 4-2 and a MarrCCD 165 fiber optic two-dimensional quadrant detector positioned for a sample-to-detector distance of 0.4 m to obtain a measurable Q -range of $0.0126\text{--}0.36 \text{ \AA}^{-1}$. However, the beam monitor that is used to automatically position the beam stop during data acquisition was malfunctioning, and as a result, data below $0.025 \text{ \AA}^{-1}$ were sporadically contaminated with parasitic scattering from the beam stop and hence could not be included in the data analysis. The Q -values at individual detector channels were calibrated using a (100) reflection from the standard polycrystalline cholesterol myristate sample. The scattering data were collected as 10 image frames each of 15 second exposure. Each successive frame was monitored for potential radiation damage. Solvent scattering was measured in the same cell as close in time to the sample data acquisition as practical. The program MacParse was used to average image frames, subtract solvent scattering, and circularly average the scattering profile.

Initial SAXS experiments to evaluate sample quality and optimize solvent conditions were performed at the University of Utah using the instrument described in ref. (85). Experiments to evaluate protein samples for potential inter-particle interference effects were also performed at SSRL Beam Line 4-2L. For these measurements, a one-dimensional wire position sensitive proportional detector filled with 80/20% Xe/CO₂ mixture was used at a sample-to-detector distance of 1.25 m (Q -range $\approx 0.03\text{--}0.78 \text{ \AA}^{-1}$). Data reduction was carried out with the software packages SOPKO or OTOKO customized for SSRL. Data were acquired for four different concentrations of RCM- and RCAM-RNase A (2–10 mg/ml) at low pH with 0 and 6 M urea. There was no measurable concentration dependence to the data for any of these conditions.

Acknowledgements

This research was supported by a grant from the U.S. National Institutes of Health (no. GM-42494). JT is supported by an Australian Research Council Federation Fellowship (FF0457488). X-ray scattering facilities at the University of Utah were supported by the Office of Science (BER), U. S. Department of Energy, Grant No. DE-FG02-05ER64026 (to JT). The small-angle scattering experiments were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program. We thank Drs. Hiro Tsuruta and Marc Niebuhr for assistance with SAXS instrumentation at SSRL and David Jacques and David Langley for their help in around the clock data acquisition.

References

1. Sela M, White F Jr, Anfinsen CB. Reductive cleavage of disulfide bridges in ribonuclease. *Science* 1957;125:691–692. [PubMed: 13421663]
2. Harrington W, Sela M. A comparison of the physical chemical properties of oxidized and reduced alkylated ribonuclease. *Biochim Biophys Acta* 1959;31:427–434. [PubMed: 13628670]

3. Anfinsen CB, Haber E, Sela M, White F Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA* 1961;47:1309–1314. [PubMed: 13683522]
4. Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973;181:223–230. [PubMed: 4124164]
5. Sosnick TR, Trewhella J. Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry* 1992;31:8329–8335. [PubMed: 1525171]
6. Nöppert A, Gast K, Müller-Frohne M, Zirwer D, Damaschun G. Reduced-denatured ribonuclease A is not in a compact state. *FEBS Lett* 1996;380:179–182. [PubMed: 8603733]
7. Zhou JM, Fan YX, Kihara H, Kimura K, Amemiya Y. The compactness of ribonuclease A and reduced ribonuclease A. *FEBS Lett* 1998;430:275–277. [PubMed: 9688554]
8. Jacob J, Dothager R, Thiagarajan P, Sosnick T. Fully reduced ribonuclease A does not expand at high denaturant concentration or temperature. *J Mol Biol* 2007;367:609–615. [PubMed: 17292402]
9. Navon A, Ittah V, Landsman P, Scheraga HA, Haas E. 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* 2001;40:105–118. [PubMed: 11141061]
10. Navon A, Ittah V, Scheraga HA, Haas E. Formation of the hydrophobic core of ribonuclease A through sequential coordinated conformational transitions. *Biochemistry* 2002;41:14225–14231. [PubMed: 12450386]
11. Dill KA, Shortle D. Denatured states of proteins. *Annu Rev Biochem* 1991;60:795–825. [PubMed: 1883209]
12. Shortle D. The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J* 1996;10:27–34. [PubMed: 8566543]
13. Shortle D, Ackerman MS. Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 2002;293:487–489. [PubMed: 11463915]
14. Millett IS, Doniach S, Plaxco KW. Towards a taxonomy of the denatured state: Small angle scattering studies of unfolded proteins. *Adv Prot Chem* 2002;62:241–262.
15. Baldwin RL. A new perspective on unfolded proteins. *Adv Prot Chem* 2002;62:361–367.
16. Rose, G., editor. *Unfolded Proteins*, vol. 62 of *Advances in Protein Chemistry*. Academic Press; New York: 2002.
17. McCarney E, Kohn J, Plaxco KW. Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit Rev Biochem Mol Biol* 2006;40:181–189. [PubMed: 16126485]
18. Tanford C, Kawahara K, Lapange S. Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrochloride. *J Am Chem Soc* 1967;89:729–736.
19. Tanford C. Protein denaturation. *Adv Prot Chem* 1968;23:121–282.
20. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* 1999;38:16424–16431. [PubMed: 10600103]
21. Tcherkasskaya O, Uversky VN. Denatured collapsed states in protein folding: Example of apomyoglobin. *Proteins* 2001;44:244–254. [PubMed: 11455597]
22. Kohn J, Millett IS, Jacob J, Zagrovic B, Dillon T, Cingel N, Dothager RS, Seifert S, Thiagarajan P, Sosnick TR, Hasan M, Pande V, Ruczinski I, Doniach S, Plaxco KW. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc Natl Acad Sci USA* 2004;101:12491–12496. [PubMed: 15314214]
23. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry* 2002;41:6573–6582. [PubMed: 12022860]
24. Uversky VN. Natively unfolded proteins: A point where biology waits for physics. *Prot Sci* 2002;11:739–756.
25. Dyson H, Wright P. Intrinsically unstructured proteins and their functions. *Nature Rev Mol Cell Biol* 2005;6:197–208. [PubMed: 15738986]

26. Flanagan JM, Kataoka M, Shortle D, Engelman DM. Truncated staphylococcal nuclease is compact but disordered. *Proc Natl Acad Sci USA* 1992;89:748–752. [PubMed: 1731350]
27. Chugha P, Sage H, Oas TG. Methionine oxidation of monomeric λ repressor: The denatured state ensemble under nondenaturing conditions. *Protein Sci* 2006;15:533–542. [PubMed: 16452618]
28. Guinier A. La diffraction des rayons X aux très petits angles, application à l'étude de phénomènes ultramicroscopiques. *Ann Physique* 1939;12:161–237.
29. Bernadó P, Blanchard L, Timmins P, Marion D, Rulgrok R, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci USA* 2005;102:17002–17007. [PubMed: 16284250]
30. Bernadó P, Mylonas E, Petoukhov M, Blackledge M, Svergun D. Structural characterization of flexible proteins using small-angle x-ray scattering. *J Am Chem Soc* 2007;129:5656–5664. [PubMed: 17411046]
31. Goldenberg DP. Computational simulation of the statistical properties of unfolded proteins. *J Mol Biol* 2003;326:1615–1633. [PubMed: 12595269]
32. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 1997;273:283–298. [PubMed: 9367762]
33. Smith LJ, Fiebig KM, Schwalbe H, Dobson CM. The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold Des* 1996;1:R95–R106. [PubMed: 9080177]
34. Rubinstein, M.; Colby, R. Polymer physics. Oxford Univ. Press; Oxford, UK: 2003.
35. Myers JK, Pace CN, Scholtz JM. Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 1995;4:2138–2148. [PubMed: 8535251]
36. Qu Y, Bolen CL, Bolen DW. Osmolyte-driven contraction of a random coil protein. *Proc Natl Acad Sci U S A* 1998;95:9268–9273. [PubMed: 9689069]
37. Auton M, Bolen DW. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry* 2004;43:1329–1342. [PubMed: 14756570]
38. Auton M, Bolen DW. Predicting the energetics of osmolyte-induced protein folding/unfolding. *Proc Natl Acad Sci USA* 2005;102:15065–15068. [PubMed: 16214887]
39. Schellman JA. Solvent denaturation. *Biopolymers* 1978;17:1305–1322.
40. Schellman JA. Protein stability in mixed solvents: A balance of contact interaction and excluded volume. *Biophys J* 2003;108:108–125. [PubMed: 12829469]
41. Timasheff SN. Control of protein stability and reactions by weakly interacting cosolvents: The simplicity of the complicated. *Adv Prot Chem* 1998;51:355–432.
42. Courtenay E, Capp M, Saeker R, Record MT Jr. Thermodynamic analysis of interactions between denaturants and protein surface exposed on unfolding: Interpretation of urea and guanidinium chloride m -values and their correlation with changes in accessible surface area (ASA) using preferential interaction coefficients and the local-bulk domain model. *Proteins* 2000;41:72–85. [PubMed: 11013402]
43. Courtenay E, Capp MW, Record MT Jr. Thermodynamics of interactions of urea and guanidinium salts with protein surface: Relationship between solute effects on protein processes and changes in water-accessible surface area. *Prot Sci* 2001;10:2485–2497.
44. Street T, Bolen DW, Rose G. A molecular mechanism for osmolyte-induced protein stability. *Proc Natl Acad Sci USA* 2006;103:13997–14002. [PubMed: 16968772]
45. Ferrenberg A, Swendsen R. New Monte Carlo technique for studying phase transitions. *Phys Rev Lett* 1988;61:2635–2638. [PubMed: 10039183]
46. Ferrenberg A, Swendsen R. Optimized Monte Carlo Data Analysis. *Phys Rev Lett* 1989;63:1195–1198. [PubMed: 10040500]
47. Socci ND, Onuchic JN. Kinetic and thermodynamic analysis of proteinlike heteropolymers: Monte Carlo histogram technique. *J Chem Phys* 1995;103:4732–4744.
48. Flory, PJ. Principles of Polymer Chemistry. Cornell Univ. Press; Ithaca, NY: 1953.
49. de Gennes, PG. Scaling Concepts in Polymer Physics. Cornell Univ. Press; Ithaca, NY: 1979.
50. Kratky O, Porod G. Röntgenuntersuchung gelöster faden-moleküle. *Rec Trav Chim Pays-Bas* 1949;68:1106–1122.

51. Kratky O. Possibilities of x-ray small angle analysis in the investigation of dissolved and solid high polymer substances. *Pure Appl Chem* 1966;12:483–523.
52. Porod, G. General theory. In: Glatter, O.; Kratky, O., editors. *Small-angle x-ray scattering*. Academic Press; London: 1982. p. 17-51.
53. Debye P. Molecular-weight determination by light scattering. *J Phys Chem* 1947;51:18–32.
54. Kratky, O. Natural high polymers in the dissolved and solid state. In: Glatter, O.; Kratky, O., editors. *Small-angle x-ray scattering*. Academic Press; London: 1982. p. 361-386.
55. Teixeira J. Small-angle scattering by fractal systems. *J Appl Cryst* 1988;21:781–785.
56. Schmidt, P. Use of scattering to determine the fractal dimension. In: Avnir, D., editor. *The fractal approach to heterogeneous chemistry*. John Wiley & Sons; Chichester, U.K.: 1989. p. 67-79.
57. Beaucage G. Small-angle scattering from polymeric mass fractals of arbitrary mass-fractal dimension. *J Appl Cryst* 1996;29:134–136.
58. Dewey, T. *Fractals in molecular biophysics*. Oxford Univ. Press; Oxford, UK: 1997. p. 20-58.
59. Jha A, Colubri A, Freed KF, Sosnick TR. Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc Natl Acad Sci USA* 2005;102:13099–13104. [PubMed: 16131545]
60. Ding F, Jha R, Dokholyan NV. Scaling behavior and structure of denatured proteins. *Structure* 2005;13:1047–1054. [PubMed: 16004876]
61. Tran H, Wang X, Pappu R. Reconciling observations of sequence-specific conformational propensities with the generic polymeric behavior of denatured proteins. *Biochemistry* 2005;44:11369–11380. [PubMed: 16114874]
62. Fizkee N, Rose GD. Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci U S A* 2004;101:12497–12502. [PubMed: 15314216]
63. Galat A, Creighton TE, Lord RC, Blout ER. Circular dichroism, Raman spectroscopy, and gel filtration of trapped folding intermediates of ribonuclease. *Biochemistry* 1981;20:594–601. [PubMed: 7213596]
64. Flory, PJ. *Principles of Polymer Chemistry*. Vol. chap. XIV. Cornell Univ. Press; Ithaca, NY: 1953. p. 595-639.
65. Tran H, Pappu R. Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. *Biophys J* 2006;91:1868–1886. [PubMed: 16766618]
66. Vitalis A, Wang X, Pappu RV. Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories. *Biophys J* 2007;93:1923–1937. [PubMed: 17526581]
67. Beaucage G. Approximations leading to a unified exponential/power-law approach to small-angle scattering. *J App Cryst* 1995;28:717–728.
68. Beaucage G, Rane S, Sukumaran S, Satkowski MM, Schechtman LA, Doi Y. Persistence length of isotactic poly(hydroxy butyrate). *Macromolecules* 1997;30:4158–4162.
69. Gittings M, Cipelletti L, Trappe V, Weitz D, In M, Marques C. Structure of guar in solutions of H₂O and D₂O: An ultra-small-angle light-scattering study. *J Phys Chem B* 2000;104:4381–4386.
70. Chen SH, Teixeira J. Structure and fractal dimension of protein-detergent complexes. *Phys Rev Lett* 1986;57:2583–2586. [PubMed: 10033804]
71. Dill KA. Theory for the folding and stability of globular proteins. *Biochemistry* 1985;24:1501–1509. [PubMed: 3986190]
72. Chan HS, Dill KA. Polymer principles in protein structure and stability. *Annu Rev Biophys Biophys Chem* 1991;20:447–490. [PubMed: 1867723]
73. Houry W, Rothwarf D, Scheraga H. Circular dichroism evidence for the presence of burst-phase intermediates on the conformational folding pathway of ribonuclease A. *Biochemistry* 1996;35:1015–1013.
74. Colon W, Roder H. Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* 1997;7:15–28. [PubMed: 9032062]
75. Sosnick TR, Shtilerman MD, Mayne L, Englander SW. Ultrafast signals in protein folding and the polypeptide contracted state. *Proc Natl Acad Sci USA* 1997;94:8545–50. [PubMed: 9238013]

76. Kimura T, Akiyama S, Uzawa T, Ishimori K, Morishima I, Takahashi S. Specifically collapsed intermediate in the early stage of the folding of ribonuclease A. *J Mol Biol* 2005;350:349–362. [PubMed: 15935376]
77. Uzawa T, Kimura T, Ishimori K, Morishima I, Matsui T, Ikeda-Saito M, Takahashi S, Akiyama S, Fujisawa T. Time-resolved small-angle x-ray scattering investigation of the folding dynamics of heme oxygenase: Implication of the scaling relationship for the submillisecond intermediates of protein folding. *J Mol Biol* 2006;357:997–1008. [PubMed: 16460755]
78. Arai M, Kondrashkina E, Kayatekin C, Matthews CR, Iwakura M, Bilsel O. Microsecond hydrophobic collapse in the folding of *Escherichia coli* dihydrofolate reductase, an α/β -type protein. *J Mol Biol* 2007;368:219–229. [PubMed: 17331539]
79. Thompson JB, Hansma HG, Hansma PK, Plaxco KW. The backbone conformational entropy of protein folding: Experimental measures from atomic force microscopy. *J Mol Biol* 2002;322:645–652. [PubMed: 12225756]
80. Lee B, Richards FM. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol* 1971;55:379–400. [PubMed: 5551392]
81. Chothia C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 1976;105:1–14. [PubMed: 994183]
82. Svergun D, Barberato C, Koch M. CRY SOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J App Crystal* 1995;28:768–773.
83. Svergun D, Richard S, Koch M, Sayers Z, Zaccai G. Protein hydration in solution: Experimental observation by x-ray and neutron scattering. *Proc Natl Acad Sci USA* 1998;95:2267–2272. [PubMed: 9482874]
84. Bevington, P.; Robinson, D. Data reduction and error analysis for the physical sciences. Vol. 3rd. McGraw-Hill; New York: 2003. p. 194-197.
85. Heidorn D, Trehwella J. Comparison of the Crystal and Solution Structures of Calmodulin and Troponin C. *Biochemistry* 1988;27:909–915. [PubMed: 3365370]
86. Wlodawer A, Svensson LA, Sjolín L, Gilliland GL. Structure of phosphate-free ribonuclease A refined at 1.26 Å. *Biochemistry* 1988;27:2705–2717. [PubMed: 3401445]

Abbreviations used

RNase A	bovine ribonuclease A
SAXS	small-angle x-ray scattering
FRET	Förster resonance energy transfer
RCM	reduced and carboxymethylated
RCAM	reduced and carboxyamidomethylated

Figure 1.

Calculated and experimentally determined radii of gyration for unfolded proteins in strong denaturants. The filled circles represent experimentally-determined values recently compiled by Kohn et al.(22). The open circles represent the values calculated from simulated ensembles of unfolded chains based on four amino acid sequences: ω -conotoxin MVIIA-Gly (26 amino acid residues), bovine pancreatic trypsin inhibitor (58 residues), ribonuclease A (124 residues) and the α -subunit of tryptophan synthase (268 residues) (31). The curve is a least-squares fit of a power function to the simulated data. The fit exponent is 0.58 ± 0.02 .

Figure 2.

Distributions of accessible surface areas (ASA) of conformations in simulated ensembles of reduced and unfolded RNase A. (a) The unweighted histogram for an ensemble of approximately 450,000 conformations calculated under the assumption of an athermal solvent. The filled circles represent the number of chains per 1,000 which lie in a bin corresponding to a range of ASA values of $55 \text{ \AA}^2$. The solid curve was generated by separately fitting the left- and right-hand sides of the observed distribution to Gaussian functions and smoothing the combined curves.

(b) Weighted histograms representing simulated distributions under conditions of favorable or unfavorable solvation. The original histogram is represented by shaded bars, and the fit and smoothed function is represented by the thicker curve labeled “0”. The thinner curves represent weighted histograms calculated for the indicated values of ΔG_{solv} (in units of $\text{cal/mol/\AA}^2$), with the Boltzmann weighting factors calculated from the accessible surface area according to Equation 1.

Figure 3.

Histograms correlating accessible surface area and radius of gyration in unweighted and weighted ensembles of unfolded RNase A. (a) The histogram derived directly from the ensemble of approximately 450,000 calculated conformations. (b) A smoothed histogram generated as described in Materials and Methods. (b), (c), (d) and (e) Weighted histograms calculated according to Equation 1 and using the indicated values of ΔG_{solv} . Each histogram is composed of 10,000 bins, with widths of $0.6 \text{ \AA}$ in the R_g dimension and $70 \text{ \AA}^2$ in the ASA dimension. The occupancies of the histogram bins are represented by the color key shown in panel a. Contour lines representing the unweighted, but smoothed, distribution are drawn for reference in panels b–f. The contour lines are separated by a factor of 10 in the occupancy of the histogram bins.

Figure 4.

Simulated effects of solvation on the average radius of gyration and average intramolecular distances in unfolded RNase A. (a) RMS radius of gyration as a function of ΔG_{solv} . Cells in the smoothed histogram shown in Figure 3a were weighted according to the accessible surface area and the indicated values of ΔG_{solv} , using Equation 1, and the RMS average radius of gyration was calculated from the weighted distribution.

(b) RMS distance between pairs of Cys sulfur atoms as a function of loop size, calculated from weighted distributions assuming different values of ΔG_{solv} , as indicated by the key. Two-dimensional histograms similar to those shown in Fig. 3a, but with the radius of gyration axis replaced with the distances (R) between Cys sulfur atoms, were constructed for six of the

possible Cys-Cys pairs of RNase A: 58-65, 65-84, 40-72, 58-110, 26-95 and 26-110. The histograms were smoothed and Boltzmann weighting was applied assuming values of ΔG_{solv} from -4 to 4 cal/mol/Å². The curves represent least-squares fits of the calculated values to a power function, $RMS(R) = AN^{\nu}$, where $N = |i - j|$ is the chain segment length separating Cys residues i and j , and ν is the Flory exponent. The values of the fit parameters were: $A = 5.89 \pm 0.27$, $\nu = 0.615 \pm 0.011$ ($\Delta G_{solv} = -4$ cal/mol/Å²); $A = 5.95 \pm 0.27$, $\nu = 0.598 \pm 0.011$ ($\Delta G_{solv} = -2$ cal/mol/Å²); $A = 6.10 \pm 0.026$, $\nu = 0.564 \pm 0.010$ ($\Delta G_{solv} = 0$ cal/mol/Å²); $A = 6.54 \pm 0.24$, $\nu = 0.501 \pm 0.009$ ($\Delta G_{solv} = 2$ cal/mol/Å²); $A = 7.35 \pm 0.34$, $\nu = 0.43 \pm 0.012$ ($\Delta G_{solv} = 4$ cal/mol/Å²).

(c) The Flory exponent, ν , for the power dependence of the RMS distance between Cys sulfur atoms as a function of solvation free energy, ΔG_{solv} .

Figure 5.

Simulated small-angle x-ray scattering curves for native and unfolded RNase A. For each of approximately 45,000 calculated conformations of RNase A, an SAXS curve was calculated using the computer program CRY SOL(82). The average of these curves, plotted as $\log(I)$ versus the scattering vector Q , is shown by the filled circles in panel a. To simulate the effects of favorable or unfavorable solvation, weighted averages of the curves were calculated, with the weights calculated from the accessible surface areas of the individual conformations and values of ΔG_{solv} as indicated by the key in panel a. Each of the curves shown in panel a represents a fit of the discrete points in the calculated curve to a sixth-order polynomial, which were used for fitting to the experimental data, as in Fig. 8. A predicted scattering curve for native RNase A was calculated from the atomic coordinates from entry 7RSA of the Protein Data Bank (86). The predicted scattering data are also shown as a Guinier plot ($\ln(I)$ versus Q^2) in panel b.

Figure 6.

Log-log plots of simulated small-angle x-ray scattering curves for unfolded RNase A under different solvation conditions. (a) The simulated curves were calculated as described in the legend for Figure 5, using Boltzmann weighting factors calculated for values of ΔG_{solv} from -4 to 4 cal/mol/Å², as indicated by the key. The inset shows an expanded view of the intermediate scattering angle region, from which the slope of each curve was determined. For each curve, the linear region was identified by visual inspection, and the data in this region was fit to a line by the method of least squares, as indicated by the line segments drawn in the inset. (b) Fractal mass dimension, D_m , as a function of ΔG_{solv} . The filled circles represent the negative slopes derived from the linear regions of the log-log plots in panel a, while the open squares are the inverses of the Flory exponent (ν) estimated from plots of RMS(R) versus chain segment length (Figure 4).

Figure 7.

Calculated Kratky plots for native RNase A and the simulated unfolded ensembles under different solvation conditions. The simulated scattering curves were calculated as described in the legend to Figure 5. Filled circles identify the curve calculated for the unweighted distribution ($\Delta G_{solv} = 0$); The curves identified with filled squares and triangles correspond to ΔG_{solv} values of 2 and 4 cal/mol/ $\text{\AA}^2$, respectively; Open squares and triangles identify curves calculated for ΔG_{solv} equal to -2 and -4 cal/mol/ $\text{\AA}^2$, respectively.

Figure 8.

Experimental and fit SAXS curves for reduced and carboxyamidomethylated (RCAM) RNase A at pH 7. Samples contained the indicated urea concentrations and approximately 5 mg/mL RCAM-RNase A. The experimental data are shown as small points with error bars derived from counting statistics. The experimental data were fit to polynomials representing the simulated scattering curves by the method of least squares, with only the constant term allowed to vary to account for variations in the scattering intensity. Each experimental data set was fit to 9 individual simulated scattering curves, corresponding to values of ΔG_{solv} from -4 to 4 cal/mol/Å², and the insets of each graph show the reduced χ^2 value as a function of ΔG_{solv} . The best fit to each experimental scattering curve is shown as a solid curve, corresponding to ΔG_{solv} = 1, 0, -1 and -2 cal/mol/Å² for 1 M, 2 M, 4 M and 6 M urea, respectively. In each graph, the short dashes represent the fit to the curve calculated using a value of ΔG_{solv} equal to 2 cal/mol/Å² greater than that for the best fit, and the long dashes represent the curve with ΔG_{solv} equal to 2 cal/mol/Å² less than that of the best fit.

Figure 9.

Log-log plots of experimental scattering curves for RCAM-RNase A in solutions containing urea at the indicated concentrations, at pH 7 (a) and pH 3 (b). The graphs have been displaced by arbitrary amounts along the vertical axis for clarity, and Q has units of $\text{\AA}^{-1}$. The lines represent least-squares fits to the data range shown in the figure, weighted by the uncertainties derived from counting statistics. The slopes derived from the data for the pH 7 samples were: -1.88 ± 0.012 (1 M urea), -1.83 ± 0.012 (2 M urea), -1.77 ± 0.009 (4 M urea), -1.65 ± 0.012 (6 M urea). For the pH 3 samples, the slopes were: -1.59 ± 0.013 (1 M urea), -1.57 ± 0.008 (2 M urea), -1.59 ± 0.01 (4 M urea), -1.56 ± 0.016 (6 M urea).

Figure 10.

Fractal dimension of reduced and unfolded RNase A as a function of urea concentration. The fractal dimension of the RCAM (a) and RCM (b) forms of RNase A were determined as the negative slopes of $\log(I)$ versus $\log(Q)$ over the range of $\log(Q)$ values from -1.2 to -0.8, as illustrated in Figure 9. The uncertainties in D_m , derived from the least-squares fits, were all less than or equal to 0.02, represented by the approximate size of the symbols. The predicted net charge of the polypeptide is shown in parentheses next to the pH values.

Figure 11.

The predicted change in conformational entropy, ΔS_{conf} , for the transfer of reduced and unfolded RNase A from an athermal solvent to conditions of favorable or unfavorable solvation. ΔS_{conf} was calculated from Equation 11 and the smoothed histogram shown in Figure 2.