Small-Angle Neutron Scattering by a Strongly Denatured Protein: Analysis Using Random Polymer Theory

Andrei-José Petrescu,*,** Veronique Receveur,# Patrick Calmettes,# Dominique Durand,#§ Michel Desmadril,¶ Benoit Roux,‖ and Jeremy C. Smith**

*Institute of Biochemistry, 77000 Bucharest, Romania; #Laboratoire Léon Brillouin, CEA-Saclay, 91191 Gif-sur-Yvette cedex, France; §LURE, Université Paris-Sud, 91405 Orsay cedex, France; ¶Laboratoire d'Enzymologie, Groupe de Recherche du CNRS, Université Paris-Sud, 91405 Orsay cedex, France; ‖Département de Chimie, Université de Montréal, Montréal, Canada; and **SBPM/DBCM, CEA-Saclay, 91191 Gif-sur-Yvette cedex, France

ABSTRACT Small-angle neutron scattering profiles are presented from phosphoglycerate kinase, in the native form and strongly denatured in 4 M guanidinium chloride (GdnHCl) solution. The data are interpreted using a model in which the excess scattering density associated with the protein is represented as a finite freely jointed chain of spheres. The similarity of the model-derived scattering function to experiment increases asymptotically with the number of spheres. The improvement of the fit obtained with more than ~200 spheres (i.e., two residues per sphere) is insignificant. The effects of finite size of the scattering units and of scattering length variation along the polypeptide chain are examined. Improved agreement with experiment is obtained when these effects are taken into account. A method for rapid calculation of the scattering profile of a full, all-atom configuration is examined. It is found that a representation of the chain containing two scattering units per residue, placed at the backbone and side-chain scattering length centroids, reproduces the full, all-atom profile to within 2%.

INTRODUCTION

The folding of globular proteins is a central problem in molecular biology. For a complete understanding of folding pathways and thermodynamics, a physical characterization of both native and denatured states is required. Denatured states can be generated, in vitro, using temperature or denaturing chemicals. The degree of denaturation depends on the conditions used. When the denaturation is mild, a globular protein may adopt a compact nonnative conformation with a significant amount of residual structure (Kuwajima, 1989; Shortle, 1996; Miranker and Dobson, 1996). More extreme conditions result in an expansion of the chain, with further loss of structure (Tanford, 1968). The characterization of the configurational distribution of strongly denatured proteins can provide information on initial events in protein folding. Phosphoglycerate kinase under these strongly denaturing conditions is the object of the present investigation.

Whereas the detailed examination of native proteins, which are relatively structured, is often possible using x-ray crystallography or NMR spectroscopy, this level of information is not readily available for denatured states. However, recently a variety of experimental techniques have been applied to furnish information on the conformational and thermodynamic properties of denatured states of several proteins. For example, local structural features in denatured proteins have been investigated using NMR (Neri et al., 1992; Radford et al., 1992; Shortle, 1996; Miranker and Dobson, 1996). Complementary, global information can be obtained using solution small-angle scattering of x-rays (SAXS) and neutrons (SANS). The small-angle scattering techniques provide information on low- and medium-resolution features of the orientationally and configurationally averaged protein chain. SAXS has been employed to investigate denatured states in several proteins (Kataoka et al., 1993, 1995; Eliezer et al., 1995). SANS has an important advantage over SAXS in that contrast of the protein over the solvent can be optimized by exploiting the difference in sign of the scattering lengths of deuterium and hydrogen (Higgins and Benoit, 1994).

In a small-angle scattering experiment the scattered intensity $I(q)$ is measured as a function of the scattering wave vector q. $I(q)$ as $q \rightarrow 0$ can be used to determine the radius of gyration. For a strongly denatured protein, $I(q)$ when $q > R_g^{-1}$ gives information on the chain statistics. To interpret the whole scattering curve simultaneously, a statistical model of the denatured protein is necessary. In recent work, small-angle neutron scattering has been applied to examine the configurational distribution of phosphoglycerate kinase (PGK) strongly denatured in 4 M guanidinium chloride (GdnDCl) (Calmettes et al., 1993, 1994). Under these conditions the radius of gyration of the protein is nearly four times larger than in the native state. In the range 0.045 $Å^{-1} < q < 0.2$ $Å^{-1}$, the protein behaves like an excluded volume chain. An atomic-detail representation of sample configurations of the polypeptide was obtained by using molecular mechanics calculations to insert segments of the chain into a freely jointed chain of spheres (FJC) that had been fitted to the experimental data.

Here we investigate the model dependence of SANS scattering curves from strongly denatured PGK. We present new experimental neutron scattering data and reinterpret

Received for publication 17 April 1996 and in final form 22 October 1996.

Address reprint requests to Dr. Jeremy C. Smith, Section de Biophysique des Proteines et des Membranes, Department de Biologie Cellulaire et Moleculaire, C.E.A., Centre d'Etudes, Saclay, 91191 Gif-sur-Yvette Cedex, France. Tel.: 33-1-69088139; Fax: 33-1-69088139; E-mail: jeremy.smith@cea.fr.

0006-3495/97/01/335/08 $2.00

them using FJC models. An improved fit to the experimental data is obtained. The effects of finite size of the scattering units and of scattering length variation along the polypeptide chain are examined. Improved agreement with experiments is obtained when these effects are taken into account. Finally, a rapid and accurate method for approximating the scattering profile from a full, all-atom configuration of the denatured protein is presented.

MATERIALS AND METHODS

Experimental

Sample preparation

Recombinant phosphoglycerate kinase (PGK) was prepared as described by Minard et al. (1983). The activity of the enzyme was verified using a coupled assay with glyceraldehyde-3-phosphate dehydrogenase (Bucher, 1955). The protein concentration was determined both by adsorbance at 280 nm (ϵ = 0.49 ml/mg·cm) and BCA protein assay (Smith et al., 1985). Deuteration of Gdn-HCl was performed as described by Calmettes et al. (1994).

Neutron scattering experiments

The SANS measurements were performed with the PACE instrument at the Laboratoire Léon Brillouin (C.E.A., Saclay, France) as described by Calmettes et al. (1994). Scattering profiles were measured for samples at 5 mg/ml concentration at constant temperature, 4.5 ± 0.2°C; These conditions prevented any aggregation of the denatured protein. The data were not corrected for the effects of finite concentration. To do this, measurements can be performed at different concentrations and extrapolated. The expected effect is to slightly increase $I(0)$ and the radius of gyration relative to the published curve. These measurements and corrections are in progress. Scattering intensities were simultaneously recorded at different scattering angles using neutrons of wavelengths λ = 5.0 and 11.0 Å. Different sample-to-detector distances were selected to obtain wave vector transfers q ranging from 7×10^{-3} to 7×10^{-1} Å^{-1}. The scattering curves shown in Fig. 1 are the superposition of two sets of data. The errors are relatively large for the four points at lowest q, mainly because of the decreased detection of the solid angle at the corresponding scattering angles. The scattering intensities were corrected for nonuniformity of the detector response by normalization to the mainly incoherent scattering from a 1.0-mm-thick water sample, whose absolute differential cross section was measured to be 0.81 ± 0.02 cm^{-1}sr^{-1} at λ = 5.0 Å and 0.98 ± 0.02 cm^{-1}sr^{-1} at λ = 11.0 Å. The scattering spectrum of the solvent alone was subtracted in each case from the solution spectrum. The result was further corrected for the excess incoherent scattering from the unexchanged hydrogen atoms in the protein, such that the remaining intensity almost vanishes for values of $q > 0.55$ Å^{-1}.

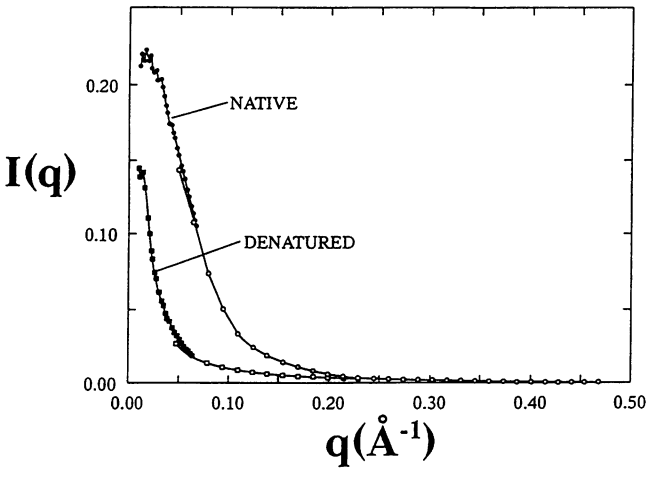

FIGURE 1 SANS profiles for PGK in the native form and denatured in 4 M guanidinium chloride solution. The contrast between the solvent and protein scattering has been maximized. Each scattering profile is a merger of two data sets, shown in filled and empty circles.

THEORETICAL ANALYSIS

The experimental conditions are such that the scattering profile is dominated by the excess scattering density associated with the protein chain and perturbed solvent. The coherent form factor $I(q)$ for a collection of n atoms separated by vectors r_{ij} and averaged over all orientations and configurations is given by (Higgins and Benoit, 1994)

$$I(q) = \frac{1}{n^2}\sum_{i,j}^{n}\langle e^{iq\cdot r_{ij}}\rangle = \frac{1}{n^2}\sum_{i,j}^{n}\left\langle\frac{\sin(qr_{ij})}{qr_{ij}}\right\rangle. \quad (1)$$

The scattering intensity at low q can be used to obtain the radius of gyration $R_g^2 = \sum_{ij}^{n}\langle r_{ij}^2\rangle/n$. In principle R_g can be obtained using the Guinier law. Expanding in a power series in q and retaining the first terms (i.e., for $q \to 0$), one obtains

$$I(q) \approx 1 - \frac{1}{3}R_g^2q^2. \quad (2)$$

In 1967 Tanford and colleagues suggested that denaturant-denatured states can be described as a random flight (Tanford et al., 1967). Consistent with this approach, to interpret the scattering profile over a wider q range, we use a model in which the excess scattering density associated with the denatured protein is treated as a finite, freely jointed chain of N identical spheres. The spheres are considered to be formed by material of equal, constant scattering density, composed of protein and associated solvent. They are connected to one another by rigid links such that the distance $2R$ between two adjacent centers is equal to their diameter, and consecutive spheres are in contact but do not overlap. The expression for the scattering profile can be derived analytically (Calmettes et al., 1993):

$$I(q) = \langle|F(q)|^2\rangle = \frac{1}{N^2}\sum_{i,j}^{N}|F^s(q)|^2\left[\frac{\sin(2qR)}{(2qR)}\right]^{|i-j|}, \quad (3)$$

were $F^s(q)$ is the form factor of an isolated sphere of radius R, and is given by

$$F^s(q) = 3\left[\frac{\sin(qR)}{(qR)^3} - \frac{\cos(qR)}{(qR)^2}\right]. \quad (4)$$

An alternative to the Guinier law for obtaining R_g is the Debye equation (Higgins and Benoit, 1994). The Debye model is equivalent to the above freely jointed chain model

when N is very large and R is zero. The coherent form factor $I(q)$ in Eq. 1 can be expressed as

$$I(q) = \frac{1}{n^2} \sum_{ij}^{n} \langle \omega_{ij}(q) \rangle, \tag{5}$$

where $\langle \omega_{ij}(q) \rangle$ is the Fourier transform of the probability $\langle p_{ij}(r) \rangle$ that two units i, j are separated by r:

$$\langle \omega_{ij}(q) \rangle = \int 4\pi r^2 \, dr \frac{\sin(qr_{ij})}{qr_{ij}} \langle p_{ij}(r) \rangle. \tag{6}$$

The Debye equation is based on the assumption that the scattering arises from an uncorrelated, freely jointed chain of material points separated by a constant distance R:

$$I(q) = \frac{1}{n^2} \sum_{ij}^{n} \langle \omega(q) \rangle^{|i-j|}, \tag{7}$$

where $\langle \omega(q) \rangle = \sin(qR)/qR$ is the Fourier transform of two adjacent units.

Assuming that the number of chain units is large ($n \rightarrow \infty$), the series can be explicitly summed giving the Debye equation:

$$I(q) = 2(e^{-x} - 1 + x)/x^2, \tag{8}$$

where $x = (qR_g)^2$. For $(qR_g)^2 < 9$, this function may be approximated to within 0.4% by the following expression:

$$\frac{1}{I(q)} = 1 + 0.359(qR_g)^{2.206}. \tag{9}$$

The radius of gyration may be obtained from the experimental data using either Guinier [$\log I(q) = f(q^2)$] or Debye [$1/I(q) = f(q^2)$] plots. It was found for the present denatured protein sample that the linear Guinier region encompasses only the first three to five experimental points. In contrast, the Debye approximation is valid for a much larger range of q (see Fig. 2). As the first points are subject to relatively large experimental error, it was decided to use the Debye approximation to obtain the radius of gyration. R_g obtained in this way is 71.5 ± 2.0 Å. This value can be compared with the previously published result of R_g = 78 Å, obtained using the Guinier approximation on a data set slightly affected by aggregation (Calmettes et al., 1993).

As can be seen from Eq. 3, the analytical form of the form factor for the finite freely jointed chain of spheres depends on only two parameters: N, the number of spheres, and R, their radius. To determine these two parameters Eq. 3 is fitted to the experimental data. This can be done in several ways. One can expand $I(q) = f(q;N, R)$ in q and perform a simultaneous linear fit for N and R in the region near $q = 0$. Another possibility would be to undertake nonlinear fits of R, for every integer value of N. The approach implemented in Calmettes et al. (1993) was different and consists of the following steps:

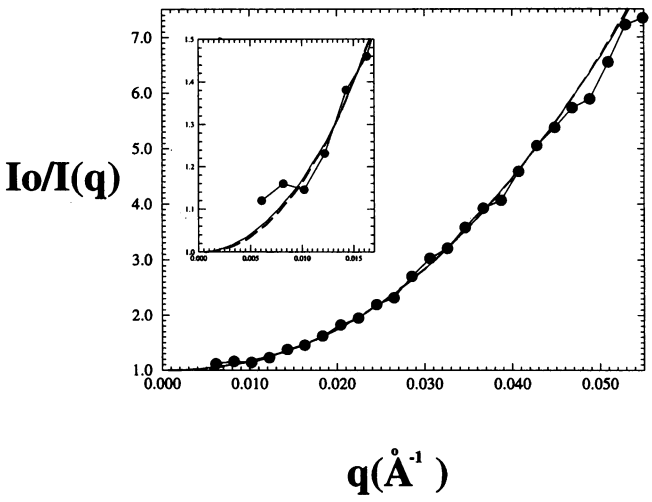

FIGURE 2 Fit of the Debye model to the scattering profile of denatured PGK. -●-, Experimental data; ——, Debye function; – – –, approximation to Debye function (Eq. 9) used to perform the fit.

1. Calculate for each N the form factor for radius $R = 1$, i.e.,

$$I(q) = f(q; N, 1).$$

2. Scale $I(q)$ such that for $q = 0$ the form factor matches the extrapolated intensity at $q = 0$, i.e.,

$$f(0; N, l) = I_o.$$

3. Calculate R by adjusting the first term in the expansion of $I(q)$:

$$I(q) = I_o - a_{exp} q^2 + \cdots \tag{10a}$$

Taking in account that

$$f(q; N, 1) = I_o - a_{model} q^2 + \cdots \tag{10b}$$

and

$$f(q; N, R) = I_o - a_{model} R^2 q^2 + \cdots \tag{10c}$$

and identifying terms in $I(q)$ and $f(q; N, R)$, one obtains

$$R = (a_{exp}/a_{model})^{1/2}. \tag{10d}$$

The accuracy of this procedure depends on the accuracy of the evaluation of $I(q)$ for small q. As in the present data set, the three to five experimental points at lowest q are subject to error and the evaluation of $I(q)$ at small q using an extrapolation to $q = 0$ of the experimental data leads to errors. In the present work the intensity at low q was determined by using the Debye equation. This is valid over a wide range of q, so the first four experimental points could be eliminated and the fitted scattering profile accurately extrapolated to $q = 0$ using the Debye function. The fitted Debye function was also used for calculating R according to Eq. 10d.

RESULTS

The SANS curves for native and denatured PGK are presented in Fig. 1. A fit of the Debye equation to the experimental $I(q)$ of the denatured protein is shown in Fig. 2. The equation fits the data well over the q range $0 < q < 3R_g^{-1}$. The fitted Debye equation was extrapolated to $q = 0$ to obtain $I(q)$ at small q. Subsequently, fits of the FJC model were made as described in Materials and Methods; these are shown for 1, 10, and 100 spheres in Fig. 3. $I(q)$ using 1 and 10 spheres falls off much more quickly with q than in the experiment. As in the previous work, the 100-sphere curve is in better agreement with experiment. However, the present model fits experiment much better than in the previously published work: the normalized mean-square deviation of the model curve from experiment is 0.834×10^{-4} compared with 14.22×10^{-4} for the previous 100-sphere fit (Calmettes et al., 1993). The increase in similarity was found to be due largely to a better fit to the data at low q, which in turn results from the improved representation of the low-q data using the Debye equation.

To investigate the deviations of $f(q; N, R)$ from $I(q)$, we calculate a similarity function defined as

$$S(q) = 100\left[1 - \left(\frac{|I_{exp}(q) - I_{model}(q)|}{I_{exp}(q)}\right)\right]. \quad (11)$$

$S(q)$ was calculated for FJC models with $N = 1$ to 6000 spheres. Fig. 4 *a*, which presents this function for 100 to 400 spheres, indicates that the FJC models reproduce best the low-q data ($q < 0.1$ Å^{-1}). The average similarity increases asymptotically with N.

A global measure of the difference in shape between experimental and model is D, given by

$$D = \sum_q \left(\frac{dI_{exp}}{dq} - \frac{dI_{mod}}{dq}\right)^2. \quad (12)$$

D as a function of N, given in Fig. 4 *b*, does possess a broad minimum, centered around 300 spheres. However, little variation in D is seen for $N > 200$.

By dividing the number of residues in the protein (415) by the number of spheres (N) in the FJC one obtains the number of residues per sphere. Equation 10d allows one to determine, for each value of N, the diameter $2R$ of the spheres. Results for several numbers of spheres are presented in Table 1. Also given are values of the maximum

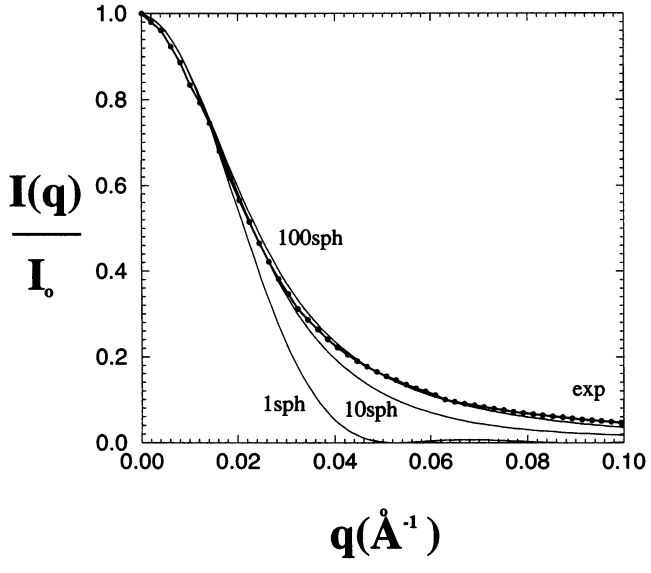

FIGURE 3 Comparison of experimental data with profiles calculated from fitted 1-, 10-, and 100-sphere FJC models.

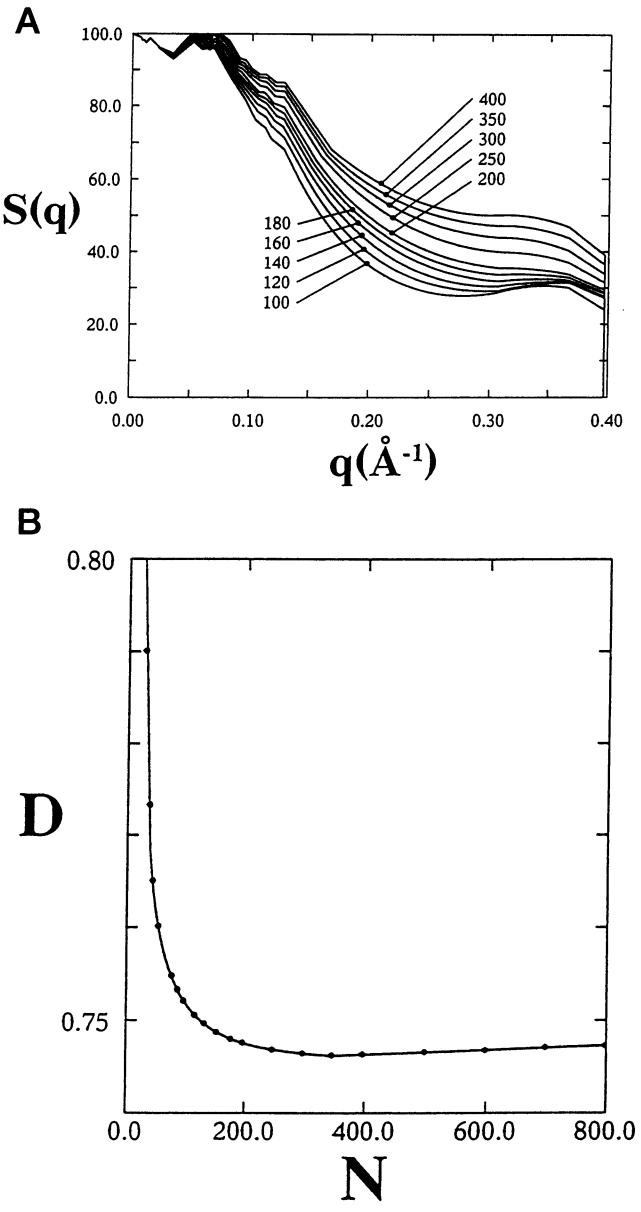

FIGURE 4 (*a*) Similarity function $S(q)$ for 100- to 400-sphere FJC models. (*b*) Shape difference function D for 1- to 1200-sphere FJC models.

TABLE 1 Diameter of FJC sphere size as a function of number of spheres, *N*

N	104 (four residues per sphere)	208 (two residues per sphere)	415 (one residue per sphere)	6000
Sphere diameter (Å)	15.9	11.3	8.0	1.0
Length of extended peptide chain (Å)	13.8	6.9	3.5	

Also given, for the cases where there are an integral number of residues per sphere, are the maximum dimensions of this number of residues in the extended conformation.

dimension L of the peptide backbone in the extended conformation, for equivalent integer numbers of residues. Clearly, a problem exists with the FJC models in that $L < 2R$. The difference between L and $2R$ is smallest for the 100-sphere model. Consequently, even though the average similarity function S is slightly lower for the 100-sphere chain than for FJC models with $N > 100$, the 100-sphere model was retained for further analysis. Most probably the overestimation of the sphere diameter is due to the fact that the FJC does not take into account excluded volume effects that are due to overlapping between third- and higher-order neighbor spheres. The fact that units may be partially or totally superposed in the model leads to an overestimation of the radius of individual spheres forming the chain, for a given radius of gyration. This intrinsic deficiency of FJC models disappears in the limit of the total number of atoms (~6000), as can be seen in Table 1.

A program was written to generate random configurations of a finite FJC of 100 spheres. Scattering curves from each configuration generated were calculated using the expression

$$I(q) = \frac{1}{N^2} |F^s(q)|^2 \sum_{i,j}^{N} b_i b_j \frac{\sin(qr_{ij})}{qr_{ij}}. \qquad (13)$$

In Fig. 5 the scattering profile calculated as the average over 50 individual FJC sphere configurations is compared with that calculated using Eq. 3, assuming an infinite number of configurations. The profiles are presented in Fig. 5 *a* as $I(q)$ versus q and in Fig. 5 *b* in a Kratky representation, $q^2I(q)$ versus q. The two profiles are closely similar, indicating that 50 is a sufficient number of configurations for convergence.

To evaluate effects due to the finite size of the scattering units on the intensity, $I(q)$ was calculated from configurations of points positioned in the centers of the spheres, with $F^s = 1$ in Eq. 13. The results are compared in Fig. 5 with $I(q)$ calculated from the chain of spheres. At large q the experimental curve is underestimated by "spheres" and overestimated by "points". The finite dimension of the scattering units results in a faster fall to zero of the intensity; this results from the convolution with the single-sphere form factor. These results suggest that the experimental scattering profile might be better reproduced using the same distance between sphere centers, but by reducing the sphere

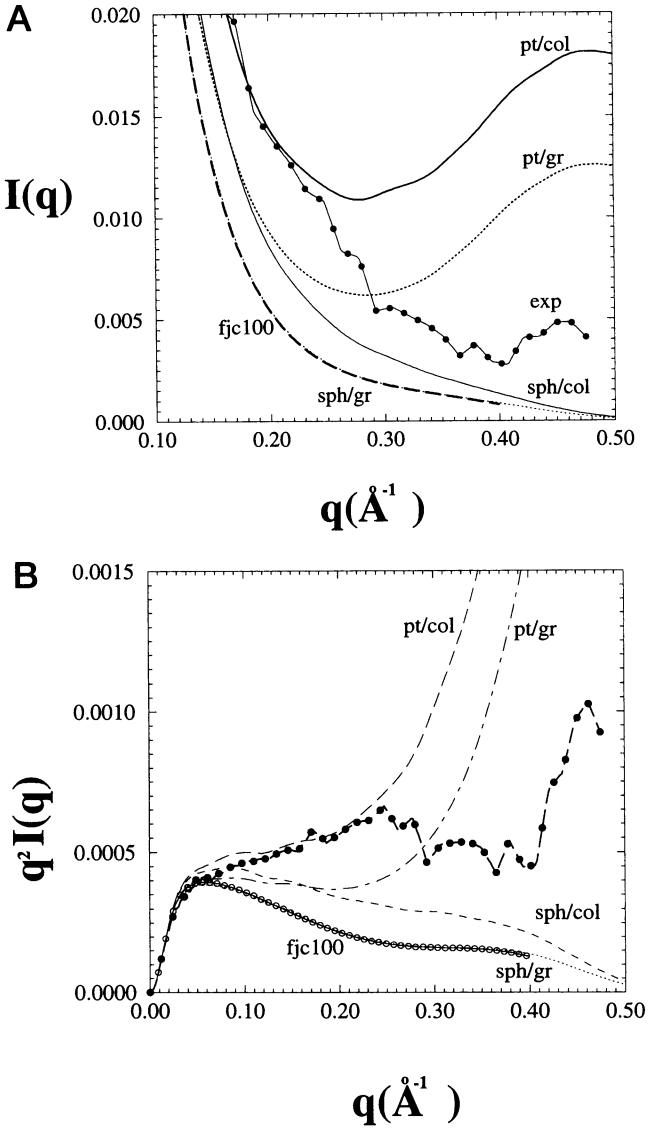

FIGURE 5 (*a*) Finite size and scattering length variation effects on calculated $I(q)$ for 100-sphere FJC model. gr uniform scattering length for each sphere; col, scattering length variation according to sequence; pts, scattering units are points at FJC sphere centers; sph, scattering units are the spheres. fjc100, calculation using Eq. 3. (*b*) Kratky plot of data in *a*.

radius. The physical interpretation of such a model would involve concentration of the scattering intensity around the centroid of each sphere. Although difficult to imagine for a native, close-packed, folded protein, this picture might be more reasonable for strongly denatured chains.

A second effect evaluated is related to the fact that the scattering factor is in reality not uniform over the chain, but varies from sphere to sphere according to the scattering factors of the constituent amino acids. The intensity curves calculated from chains with uniform scattering factors ($b_i = b_j = 1$ for all i, j in Eq. 13) are called "grey." Those from chains of spheres having different scattering factors are called "colored." As can be seen from Fig. 5, by "coloring"

the spheres the calculated intensity is closer to the experiment. In the cases of both points and spheres, the effect of taking into account the variations in the scattering length between individual spheres is to increase the intensity in the intermediate and high q regions. Finally, Fig. 5 *b* indicates that in low and intermediate q values the experimental profile is best approximated by the point-color approximation. However, this description rapidly diverges from experiment for $q > 0.25$ Å^{-1}. The deviation may be related to the fact that in the point-color approximation the individual scattering points are each assigned the mean scattering length of the spheres from which they are derived, i.e., the scattering length variation is included at a resolution an order of magnitude lower than atomic resolution. These results suggest that the scattering profile is essentially given by the FJC distribution in the range $q = 0-0.25$ Å^{-1}, whereas the profile at higher q depends critically on the higher resolution details of the chain.

The above analysis suggests that the use of models at atomic detail could improve agreement with the experimental data in the intermediate and high q ranges. However, these calculations are computer intensive. It is thus desirable and instructive to develop methods for approximating a full, all-atom scattering profile. To examine this, single configurations of the denatured polypeptide chain were generated using the method described by Calmettes et al. (1993). In this method, groups of int(N_{tot}/N) residues (where N_{tot} is the total number of residues) occupy each sphere of a single FJC configuration. Within each sphere the residues are considered initially to conserve the native structure of the protein. This biasing of the configuration generation is unlikely to alter the conclusions drawn from the present calculations. The native segments are oriented in each sphere and subjected to constrained energy minimization using the CHARMM program (Brooks et al., 1983).

At least two approaches may be used to rapidly compute the scattering curve from an all-atom configuration. A first possibility is to divide the space into boxes, precompute a distance matrix between the boxes, and then use it to calculate intensities from configurations. This replaces distance calculations by jumps to memory locations. An alternative way to approximate the scattering profile is to reduce the number of scattering points by grouping them according to a prerequisite scheme and to use this simplified configuration to approximate the intensity. This requires less memory space and was the method examined in the present work. Again, the general formula used for reconstructing the scattering curve is Eq. 13. Two levels of approximation were investigated. First, all atoms belonging to any given residue were grouped in a single scattering unit with a scattering factor equal to the sum of the constituents. Several positions of these scattering units were examined: on one of the heavy atoms of the backbone (C_α, C, N) or, according to the length of the side-chain involved, on a side-chain atom (C_β, C_γ, etc). As an illustration, in Fig. 7 results based on a single configuration of the unfolded state at full atomic details are examined, using several positions of the point scatterer, and are compared with the all-atom profile calculated for the same configuration. A picture of the configuration used is given in Fig. 6. The intensity curve for $q < 0.1$ Å^{-1} is found to be extremely sensitive to the position of the scattering units. The full, all-atom curve is considerably overestimated if scattering centers are positioned along the backbone. This might be related to regularities in nearest-neighbor distances along the backbone. The best fit at this level of approximation was obtained when units were placed at the scattering-length centroids of the residues.

A second level of approximation was examined by using two or three scattering units for each amino acid: one for the backbone atoms and one or two for the side chain, depending on its length. These results are also presented in Fig. 7. Again, several possibilities were tried with respect to the position of these scattering units, and again the best agreement with the all-atom curve is obtained when placing scattering units in the scattering-length centroid of the atoms forming a unit. This profile remains within 2% of the all-atom curve over the whole q interval and is 50–60-fold faster to compute than an all-atom profile.

DISCUSSION

To obtain a complete description of the chain statistics of strongly denatured proteins, it will be necessary to combine information from several experimental techniques sensitive to structure and dynamics, some of which give information on local interactions and others on global properties. SANS gives information on the global time-averaged configurational distribution. The work presented here investigates factors determining the SANS profile of a strongly denatured protein. To interpret the data in a consistent way requires a single model that can represent the configura-

FIGURE 6 Backbone tracing of single configuration of denatured PGK. Scattering profiles calculated from this configuration with different approximations are shown in Fig. 7.

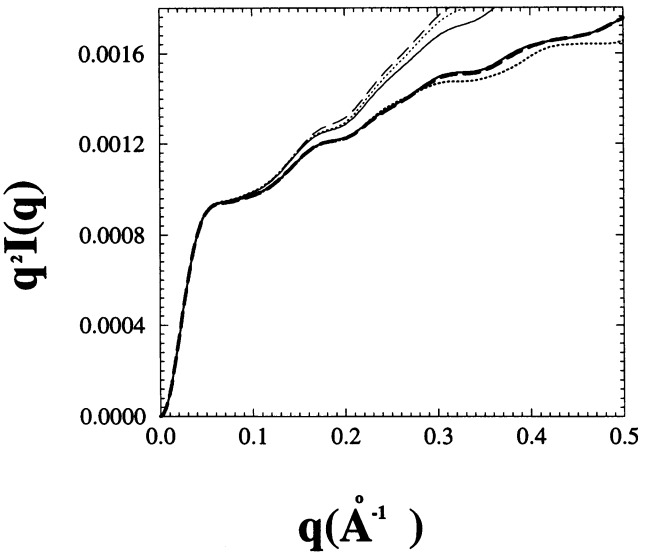

FIGURE 7 Kratky plot of scattering from all-atom representation of a single chain configuration (*bold solid line*); with scattering center placed on backbone Cα atom (*thin solid line*); with scattering center placed on backbone carbonyl carbon (*dotted line*); with scattering center placed on backbone nitrogen atom (*long dashes*); with scattering center placed on each residue centroid (*short dashes*); with scattering center placed on scattering centroids of backbone atoms (*bold dashes*) of each residue plus one or two side-chain scattering centroids depending on side-chain length.

tional statistics on the ~100 Å to ~5 Å length scales, i.e., from the overall size of the molecule down to the single residue size. Extension of the results to provide atomic-detail configurations is then feasible, although the data themselves do not directly furnish information at atomic resolution. The finite, freely jointed chain of spheres model is found to reproduce well the scattering curve over the q range 0–0.25 $Å^{-1}$, a range 20 times larger than that used to fit the model. At higher values of q the computed scattering profile diverges from experiment, suggesting that a more accurate representation of the high-resolution chain statistics is required. The similarity of the model-derived scattering function to experiment increases asymptotically with the number of spheres. The improvement of the fit obtained with more than ~200 spheres (i.e., two residues per sphere) is insignificant. Improved agreement with experiment is obtained when the effects of finite size of the scattering units and of scattering length variation along the polypeptide chain are taken into account.

Among the questions that arise are whether persistent, close-range interactions, such as hydrophobic clusters, exist between residues. The present model does not incorporate such interactions if they occur between residues far away in the sequence. However, the presence of local, persistent interactions may be, in principle, consistent with the model.

As yet we have not managed to generate a converged series of atomic-detail configurations with an average scattering profile in agreement with experiment over the whole accessible q range. However, some clues as to how to approach this are apparent from the results presented here. Scattering-length variations along the chain clearly play an important role in determining the high q intensity. Moreover, the finite size calculations suggest that the description of the excess scattering density associated with the protein chain should be somewhat more refined than the contiguous spheres representation used here. Excluded volume effects play a significant role in determining the scattering (Calmettes et al., 1994) and could be included in an improved numerical model. An accurate approximation of single all-atom configurations is possible, in which the scattering lengths associated with individual residue side-chain and backbone segments are concentrated on the corresponding scattering centroids. Scattering profiles at this level of approximation are rapid to compute and are better suited to take in account contrast effects than an all-atom calculation. This is due to the fact that the volume of a group of atoms can be better defined than for single atoms, and is of the same magnitude or larger than that of solvent molecules. This may make possible the relatively accurate subtraction of the background solvent scattering for these units.

This work was partially supported by a PECO grant awarded to AJP by the European Community.

REFERENCES

Brooks, S. R., R. E. Bruccoleri, S. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy minimisation and dynamics calculations. *J. Comput. Chem.* 4:187–217.

Bucher, T. 1955. Phosphoglycerate kinase from brewer's yeast. *Methods Enzymol.* 1:415–422.

Calmettes, P., D. Durand, M. Desmadril, P. Minard, V. Receveur, and J. C. Smith. 1994. How random is a highly denatured protein? *Biophys. Chem.* 53:105–114.

Calmettes, P., B. Roux, D. Durand, M. Desmadril, and J. C. Smith. 1993. Configurational distribution of denatured phosphoglycerate kinase. *J. Mol. Biol.* 231:840–848.

Eliezer, D., P. A. Jennings, P. E. Wright, S. Doniach, K. O. Hodgson, and H. Tsuruata. 1995. The radius of gyration of an apomyoglobin folding intermediate. *Science.* 270:487–488.

Higgins, J. S., and H. C. Benoit. 1994. Polymers and Neutron Scattering. Clarendon Press, Oxford.

Kataoka, M., Y. Hagihara, K. Mihara, and Y. Goto. 1993. Molten globule of cytochrome c studied by small angle x-ray scattering. *J. Mol. Biol.* 229:591–596.

Kataoka, M., I. Nishii, T. Fujisawa, T. Ueki, F. Tokunaga, and Y. Goto. 1995. Structural characteristics of the molten globule and native states of apomyoglobin by solution x-ray scattering. *J. Mol. Biol.* 249:215–228.

Kuwajima, K. 1989. The molten globule as a clue for understanding the folding and cooperativity of globular protein structures. *Proteins.* 6:87–103.

Minard, P., M. Desmadril, N. Ballery, N. Perahia, L. Hall, and J. M. Yon. 1983. Study of the fast-reacting cysteines in phosphoglycerate kinase using chemical modification, and site-directed mutagenesis. *Eur. J. Biochem.* 183:419–423.

Miranker, A., and C. M. Dobson. 1996. Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.* 6:31–42.

Neri, D., M. Billeter, G. Wider, and K. Wutrich. 1992. NMR determination of residual structure in urea-denatured protein, the 343 repressor. *Science.* 257:1559–1563.

Radford, S. E., C. M. Dobson, and P. A. Evans. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature.* 358:302–307.

Shortle, D. R. 1996. Structural analysis of non-native states of proteins by NMR methods. *Curr. Opin. Struct. Biol.* 6:24–30.

Smith, P. K., R. I. Krohm, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. O. Provenzo, E. K. Fujimoto, H. M. Goebe, B. J. Olson, and D. K. Klerk. 1985. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* 150:76–85.

Tanford, C. 1968. Protein denaturation. *Adv. Protein Chem.* 23:121–275.

Tanford, C., K. Kawahara, and S. Lapajne. 1967. Proteins as random coils. I. Intrinsic viscosities and sedimentation coefficients in concentrated guanidine hydrocloride. *J. Am. Chem. Soc.* 89:729–736.