Excluded volume in the configurational distribution of a strongly-denatured protein

ANDREI-JOSE PETRESCU,^{1,2} VERONIQUE RECEVEUR,² PATRICK CALMETTES,²
DOMINIQUE DURAND,^{2,3} AND JEREMY C. SMITH⁴

¹Institute of Biochemistry, 77700 Bucharest 17, Romania

²Laboratoire Leon Brillouin, CEA-Saclay, 91191 Gif sur Yvette, France

³LURE, Université Paris-Sud 91401 Orsay, France

⁴SBPM/DBCM, CEA-Saclay, 91191 Gif sur Yvette, France

(RECEIVED November 3, 1997; ACCEPTED February 23, 1998)

Abstract

The configurational distribution of phosphoglycerate kinase (PGK) strongly-denatured in 4M guanidine hydrochloride solution is investigated using small-angle neutron scattering (SANS) and Monte Carlo computer simulation. It is shown that the experimental scattering profile can be represented by a random flexible chain of spheres of excess scattering density with excluded volume interactions, the best agreement being achieved when partial sphere intersection is allowed. The radius of gyration of the chain increases by a factor of 4 on denaturation, whereas the average length of segments ~ 5 residues long increases by only $\sim 10\%$, consistent with a picture in which the large expansion on denaturation originates primarily from increased long-range flexibility of the polypeptide chain. The results provide a description of the chain statistics from which the construction of starting points for simulation studies of folding of the protein can be envisaged.

Keywords: denaturation; protein folding; simulation; small-angle neutron scattering

The understanding of protein folding poses a fundamental challenge to structural molecular biology. For many proteins folding *in vivo* is assisted by chaperones such as BiP (Kassenbrock et al., 1988) or GroEL (Schmidt & Buchner, 1992). These mechanisms are rather diverse and may involve, for example, lectin-type interactions mediated by monoglucosylated glycans attached to the protein chain (Zapun et al., 1997; Petrescu et al., 1997a). In most of these chaperonin-assisted cases, unfolding is irreversible, and cell control is essential. However, characterization of the reversible folding of small proteins *in vitro* remains a vital step toward understanding both assisted and unassisted folding *in vivo*.

Central to the understanding of the reversible *in vitro* process is the determination of the configurational space explored by a given protein at different stages during folding (Sali et al., 1994; Wolynes et al., 1996; Dill & Chan, 1997). One end state of this process, the native state, is located in a relatively limited region of the configurational space of the protein chain that can be conveniently probed using diffraction, spectroscopic, and simulation methods. In contrast, nonnative states may explore a wider range of configurations, and this leads to particular problems in their theoretical and experimental characterization.

Stable nonnative states can be induced *in vitro* using temperature, pressure, or chemical denaturants. Depending on the conditions and the protein, the degree of denaturation can vary widely, from compact states with significant amounts of residual structure to strongly denatured, expanded conformations (Tanford, 1968; Kuwajima, 1989; Miranker & Dobson, 1996). Compact nonnative states may give information on protein-folding intermediates and have been the subject of much recent experimental and theoretical work (Kim & Baldwin, 1990; Alexandrescu et al., 1993; Wright & Jennings, 1993; Lamy & Smith, 1996). However, it is also necessary to characterize strongly-denatured states, so as to provide a framework from which questions concerning both folding thermodynamics and the earliest stages in folding can be addressed. Of particular importance is the determination of the short- and long-range flexibility of strongly-denatured polypeptide chains. A related question is whether or not a protein, in a given strongly-denatured state, possesses conformational preferences, i.e., whether or not local or nonlocal structure exists.

The experimental investigation of nonnative states has employed a variety of complementary techniques. Spectroscopies such as fluorescence, circular dichroism, nuclear magnetic resonance, and quasielastic incoherent neutron scattering can be used to investigate local structural and dynamical properties (Radford et al., 1992; Chen et al., 1996; Shortle, 1996; Miranker & Dobson, 1996; Receveur et al., 1997). Some of these techniques have been used

Reprint requests to: Jeremy C. Smith, SBPM/DBCM, CEA-Saclay, 91191 Gif sur Yvette, France; e-mail: Jeremy.Smith@cea.fr.

to detect the presence of local structure in proteins under strongly-denaturing conditions (Neri et al., 1992; Gottfried & Haas, 1992). Complementary to the above methods are Small-Angle X-ray and Neutron Scattering (SAXS and SANS), which can be used to obtain information on the average structure of proteins in solution at low and medium resolutions (Calmettes et al., 1993a, 1993b, 1994; Kataoka et al., 1993, 1995; Eliezer et al., 1995; Chen et al., 1996; Kataoka & Goto, 1997; Petrescu et al., 1997b). The neutron method has the advantage, over its X-ray counterpart, of allowing the difference in sign of the scattering lengths of hydrogen and deuterium to be exploited so as to optimize the contrast between the protein and solvent, thus retaining good contrast in high concentrations of denaturant and permitting constant contrast to be obtained as a function of denaturant concentration.

Using small-angle scattering techniques one can expect to obtain information not only on the overall size of the protein (the radius of gyration), but also on the chain statistics. With this aim, SANS experiments have been performed on phosphoglycerate kinase (PGK) strongly-denatured in 4 M guanidinium chloride (GdmCl) (Calmettes et al., 1993a, 1993b, 1994; Petrescu et al., 1997b). The data were interpreted with analytical models in which the excess scattering density associated with the protein was represented as a freely-jointed chain (FJC) of contiguous uniform spheres. An FJC can take any configuration in space, subject to the constraint that the distance between the centers of geometry of nearest-neighbor spheres is fixed. Although the FJC models provide useful information on the chain statistics, they suffer from physical inconsistency at higher resolution, i.e., the theoretical scattering curves in best agreement with experiment possess spheres of unrealistically large diameters, such that the sum of the sphere diameters (the "contour" length) is significantly larger than the length of the equivalent polypeptide chain in a fully extended conformation (Petrescu et al., 1997b).

In the present article we present refined experimental scattering data on the 4 M GdmCl-denatured PGK state, corrected for finite concentration effects, and interpret them with computer models that present a consistent picture of the polypeptide chain configurational distribution at both medium and high resolutions. A likely explanation for the contour length problem with the FJC models is that there is a significant excluded volume effect in the denatured protein chain, i.e., that steric hindrance between segments of the protein has a nonnegligible effect on the scattering profile. The excluded volume effect was first introduced in the theoretical statistical physics of polymers (Flory, 1949; Fox & Flory, 1949; Chan & Dill, 1991). When excluded volume interactions are significant, the chain statistics are profoundly altered. One expected effect is that, for a chain of spheres with a given radius of gyration, the presence of excluded volume will lead to a compensating reduction of the sphere diameter relative to that of the corresponding FJC. This sphere reduction may lead to a description of the intrasphere polypeptide chain conformation (the local chain geometry) that is more physically realistic than that hitherto obtained (Petrescu et al., 1997b).

In the present work, the SANS data are analyzed at low resolution to obtain the radius of gyration, which is then used as a bound for the generation of medium-resolution excluded-volume chains (EVC). In contrast to the FJC models, analytical calculations of EVC are difficult. The difficulty can be overcome by generating computer EVC models, and here this is done using a Monte Carlo method. It is shown that, whereas both FJC and EVC can reproduce the low-resolution SANS data, only EVC succeeds

at higher resolution. Moreover, the EVC models do lead to a physically reasonable description of the local chain geometry. Finally, it is shown that partially interpenetrating spheres, including close nonlocal protein-protein interaction, lead to further improvement in the agreement with experiment. The results of the present work provide a description of the configurational distribution of the protein chain that is in accord with the whole accessible scattering profile. This description, in principle, can form a starting point for simulation of the folding of PGK.

Materials and methods

Experimental

The experimental methods are those previously reported (Petrescu et al., 1997b) with the difference that, in the present case, the effects of finite protein concentration were investigated. The concentrations used in the present experiments were 2.5, 3.5, 5.0, 7.0, and 10.0 mg/mL. In the Zimm approximation (Zimm, 1948), which is based on the concentration dependence of the osmotic pressure, the concentration dependence may be written as follows:

$$\frac{I(0,0)}{I(q,c)} = \frac{1}{P(q)} + 2A_2Mc + 3A_3Mc^2 + \dots \quad (1)$$

where $I(q,c)$ is the scattering function at concentration c , $P(q)$ is the structure factor of an isolated molecule, A_i is the i th virial coefficient, and M the molecular weight of the protein. The virial coefficients provide information on the intermolecular interaction in the solution. By performing experiments at different concentrations, the virial coefficients may be evaluated and $P(q)$ derived. Values of A_2 , the second virial, were obtained: $(11.5 \pm 1.5) \times 10^{-4} \text{ cm}^3 \text{ mol/g}^2$ for the strongly denatured and $(0.4 \pm 0.1) \times 10^{-4} \text{ cm}^3 \text{ mol/g}^2$ for the native proteins. The second virial coefficient is due to pairwise intermolecular interaction. The present results for A_2 indicate that there is significant intermolecular repulsion in the strongly-denatured state, this being much reduced in the native state. The value of A_2 for the strongly-denatured state is large enough to lead to nonnegligible variation of $I(0,c)$ with c over the concentration range examined here. In contrast, the small value of A_2 for the native protein leads to a weak concentration dependence and the extrapolation to $c = 0$ can be performed for this protein with very small error. This extrapolation, when corrected for the small change of contrast due to the GdmCl in the strongly-denatured protein sample, can also be used to obtain a value for $I(0,0)$ for the strongly-denatured protein. The value obtained in this manner was found to be the same, to within statistical error, as that obtained using Equation 1 directly from the strongly-denatured protein data. The above results are consistent with the absence of aggregation in the strongly-denatured protein sample and the presence of intermolecular repulsion. Further evidence for the absence of aggregation was obtained from direct calculation from $I(0,c)$, which, for $c \rightarrow 0$, can be used to obtain the molecular weight, m_w , from the equation $I(0,c) = cm_w/N_A(b - b_s)^2 \Phi^2$ where N_A is Avogadro's number, b the scattering length density of the protein, b_s that of the solvent, and Φ the partial specific volume of the protein. For all the concentrations examined here, the molecular weight obtained was close to that of a monomer (45.2 kDa). Further SANS experimentation showed that aggregation of the strongly-denatured protein does occur at the concentrations used here, but only after times ranging from

about 10 h at 2.5 mg/mL to 4 h at 5.0 mg/mL (Receveur, 1997). The data presented here were collected at times shorter than the above minimal aggregation times.

The finite concentration correction has a significant effect on the experimental scattering profile only at very low q ($< 0.04 \text{ \AA}^{-1}$) and, in practice, the only modified physical result is the radius of gyration (R_g). As discussed, the scattering profile of strongly-denatured PGK (Calmettes et al., 1994) is well described in the low- q region by the Debye function (Debye, 1947; Rawiso et al., 1987):

$$P(q) = \frac{2}{x^2} (x - 1 + e^{-x}) \quad (2)$$

where $x = (qR_g)^2$. The experimental profile is shown in Figure 1 together with the result of a Debye fit. The radius of gyration obtained from the fit is $R_g = 90.0 \pm 4.5 \text{ \AA}$ for the corrected profile, increased by $\sim 25\%$ compared to the value obtained from a non-corrected profile at a finite concentration of 5 mg/mL. This radius of gyration is greatly increased compared to that of the native protein ($23.8 \pm 0.3 \text{ \AA}$).

Theoretical

To interpret the data over the full accessible q -range ($0 < q < 0.4 \text{ \AA}^{-1}$) the protein was modeled as an excluded volume chain using computer simulation. In principle, Monte Carlo or molecular dynamics techniques, with full atomic-detail potential functions, can be used to do this. However, sufficient sampling of the atomic-detail configurational space of a strongly-denatured protein using molecular dynamics is computationally prohibitive, due mainly to the long dynamical relaxation times involved and the large size of the system when solvent is explicitly incorporated. For similar reasons, atomic-detail Monte Carlo simulation of strongly-denatured proteins may also be presently out-of-reach computationally, although we note, in passing, that Monte Carlo methods have re-

cently been applied with success to interpret SANS profiles of polyethylene and polystyrene (Destrée et al., 1996; Pedersen & Schurtenberg, 1996). A further problem is that the calculation of the atomic-detail scattering profile is also computationally intensive, due to the fact that a full pair function over all scattering units must be evaluated and then averaged over each chain configuration generated. A preliminary investigation into how to do this efficiently has been reported (Petrescu et al., 1997b).

Here we use a method that avoids the above complications arising from atomic-detail potential functions, and that is tailored to the information experimentally accessible. The method used simplifies the description of the denatured protein by considering a chain of N spheres of radius r_s and uniform excess scattering length b . Each sphere contains amino acid residues and associated solvent. Any two consecutive spheres are in contact, i.e., the distance between the centers-of-geometry of nearest neighbors is $2r_s$. This description is illustrated in Figure 2. The excluded volume condition forbids overlapping of a sphere with any other sphere. EVC configurations were generated using a Monte Carlo scheme based on the corresponding hard-sphere potential. To do this the position of the i th sphere center in the local frame of the $(i-1)$ th chain element (with axes z' along the $(i-2)-(i-1)$ link direction and x', y' in the perpendicular plane) was generated using two randomly sampled angles: θ with the z' axis and ω in the $x'Oy'$ plane. The coordinates in the laboratory frame were then back calculated using an Eyring transformation (Boyd & Phillips, 1993). The growth of the chain was stopped when the center of the last-generated sphere was found to be within $2r_s$ of any other sphere center. The overlap distribution as a function of chain length is shown in Figure 3. The probability of overlap is much higher for near than for far neighbors, resulting in frequent early chain termination. To overcome this problem a local overlap deletion procedure was applied, in which 2nd, 3rd, 4th, and 5th-order neighbor overlaps were avoided by deleting the two spheres concerned and restarting the chain generation from the preceding sphere. Chain generation was terminated definitively if overlaps between 6th or higher-order neighbors were detected, and the chain was then rejected. This correction procedure was found to alter the distribution of θ angles by less than 1% compared to a collection of chains generated without the local overlap deletion procedure. Furthermore, no difference was found between the average scattering profiles generated both with, and without, the local overlap deletion method. Thus, for the present purposes, no significant bias was introduced when using it. The length of the chains was further increased using a routine based on the dimerization algorithm (Alexandrowicz, 1969) in which chains are joined together end-to-end and a check for sphere intersection is performed.

Besides the local condition of reciprocal exclusion imposed by excluded volume, the procedure used incorporates a global constraint, i.e., that the scattering profile averaged over the whole set of configurations of an N -sphere model fits the experimental profile near $q = 0$, by adjusting the sphere radius using the scaling procedure (Petrescu et al., 1997b). The procedure for comparison of the models with experiment involves first generating a collection of N -sphere EVC chains, then scaling r_s to reproduce the experimental R_g , and finally comparing the model scattering profile with experiment over the whole accessible q -range. Therefore, the variables in the model are N and r_s . These two variables are not independent: as N increases r_s will decrease to maintain the radius of gyration. As the protein contains 415 residues, the number of residues in each sphere can be deduced from N .

Fig. 1. Debye fit to experimental concentration-corrected low q scattering profile. The black dots are the experimental data, and the solid line is the fitted Debye function.

Excluded Volume Chain

**Partially-Interpenetrating Spheres
Excluded Volume Chain**

Fig. 2. Illustration of two very similar configurations of the models of chains of spheres of excess scattering density. The spheres contain protein and associated solvent. The chains of spheres can take up any configuration subject to the excluded volume condition. In the EVC model all of each sphere is excluded from penetration by any other sphere. In the partial-interpenetration EVC model, the excluded volume is that associated with a fraction f of the sphere radius, and is represented by the internal spheres. Best agreement with experiment is obtained with 5 residues per sphere, with no interpenetration of nearest-neighbor spheres allowed, whereas second- and higher neighbors are allowed to partially interpenetrate, with $f = 0.7$.

Scattering profiles from individual configurations of spheres were computed as $P(q; N, r_s) = I(q; N, r_s)/I(0)$ where $I(q; N, r_s)$ is given by

$$I(q; N, r_s) = |F^{sph}(q; r_s)|^2 \sum_{i=1}^N \frac{\sin(qr_{ij})}{qr_{ij}} \quad (3)$$

where r_{ij} is the distance between the centers of geometry of spheres i, j . $|F^{sph}(q; r_s)|$ is the scattering factor of a sphere, given by the Rayleigh formula:

Fig. 3. Sphere contacts in Monte Carlo chain generation. N is the number of sphere contacts in a typical Monte Carlo run and $\Delta n_{contact}$: the number of spheres separating the contacting spheres.

$$|F^{sph}(q; r_s)| = 3 \left[\frac{\sin(qr_s)}{(qr_s)^3} - \frac{\cos(qr_s)}{(qr_s)^2} \right] \quad (4)$$

The scattering profile was calculated from the model by averaging over the scattering profiles of the individual EVC configurations using the fitted sphere radius (r_s). The scattering from corresponding FJC models was calculated as described in Calmettes et al. (1993).

Results

An important question addressed here is whether the presence of “forbidden regions” of space, which the strongly-denatured PGK polypeptide chain cannot access, can be detected in the SANS profile. These forbidden regions, modeled here by excluded volume, are likely to be due in part to the impossibility of superposition of different parts of the chain (steric hindrance), but might also include other effects such as difficulty in dislodging strongly-bound solvent molecules. To investigate the consequences of incorporating excluded-volume effects in models describing the strongly-denatured protein chain, we compare the scattering profiles from FJC and EVC models. In previous work it was found that the most satisfactory FJC model for the strongly-denatured protein was found to be that with 4 residues/sphere ($m = 4$), i.e., about 100 spheres (Calmettes et al., 1993). In Figure 4 the $m = 4$ FJC model is compared to EVC models in a Kratky plot: $q^2 P(q)$ vs. q . The Kratky representation allows a convenient visualization of the scattering at $q > 0.05 \text{ \AA}^{-1}$. For $q < 0.02 \text{ \AA}^{-1}$ the profiles are sensitive only to the overall size of the protein, i.e., R_g and both the FJC and EVC models are in accord with experiment. Differences are seen for $q > 0.02 \text{ \AA}^{-1}$. The $m = 4$ EVC scattering profile is clearly qualitatively different from the corresponding FJC profile, indicating that excluded volume has a marked effect on the calculated profile. Furthermore, the EVC models are closer to experiment than the best FJC profile.

For $q > 0.1 \text{ \AA}^{-1}$ the scattering profile is sensitive to the local chain dimensions and stiffness. In the present model this is manifested by the form factor of the spheres, which strongly influences

Fig. 4. Kratky plot of experimental and theoretical SANS. The errors bars give the experimental error in $q^2P(q)$; the error in q is negligible. Fjc4 is the 4 residues/sphere FJC model. M1^{–10} are scattering profiles calculated from the EVC Monte Carlo model, with 1–10 residues/sphere.

the high q behavior, where smaller spheres lead to too high, and larger spheres to too low, intensity. This effect is evident in comparing the EVC models with different m , shown in Figure 4. The experimental profile is not well defined for $q > 0.3 \text{ Å}^{-1}$ due to the presence of incoherent scattering. Taking this into account, the $m = 6$ model appears to be in best agreement with experiment.

All the EVC profiles in Figure 4 possess the characteristic that $q^2P(q)$ is more intense than experiment in the region $0.05 < q < 0.2 \text{ Å}^{-1}$. This difference is probably significant, as the experimental data in this region are expected to be relatively accurate. Therefore, a refinement of the model was introduced in which the excluded volume condition is modified so as to allow partial interpenetration of the spheres. This was achieved by setting the excluded volume of the spheres to be that associated with a fraction (f) of the sphere radius. Tests showed that best agreement with experiment was obtained when no interpenetration of nearest-neighbor spheres was allowed, whereas second- and higher neighbors were allowed to partially interpenetrate, with $f = 0.7$. In Figure 5 are shown the interpenetrating sphere EVC models in best agreement with experiment together with the best $f = 1.0$ (noninterpenetrating sphere) model ($m = 6$). Clearly, the incorporation of limited sphere interpenetration al-

Fig. 5. Kratky plot of experimental data and best-fitting interpenetrating sphere EVC models. m3_0.7 refers to the three residues/sphere $f = 0.7$ model, and so on. The best fitting noninterpenetrating sphere model ($m = 6$) is also shown, for comparison.

lows profiles to be generated that are in improved agreement with experiment. The closest agreement, almost to within experimental error over the whole q -range, is obtained with $p = 5$.

The sphere diameters ($d = 2r_s$) determined for the 1–7 sphere EVC models are presented in Table 1. This shows that for $m > 1$ the EVC spheres have physically reasonable diameters. This is in contrast to the FJC sphere diameters, which were found to be consistently larger than the length of corresponding extended peptide backbones (Petrescu et al., 1997). For $m = 1$ the sphere diameter is still too large, due to the additional restrictions on the θ angle imposed by the geometry of the peptide bond. When these are included, the optimized sphere size for $m = 1$ was found to approach the C_α – C_α distance of 3.8 Å .

The change in sphere size on denaturation, i.e., the change in the local chain density, is of particular interest. For the model best in agreement with experiment, i.e., $m = 5$, the increase in sphere diameter on strong denaturation is $\sim 10\%$. When the sphere diameter increase is expressed as a fraction of the maximum possible increase, i.e., that corresponding to full extension of the linear

Table 1. Diameters of unit spheres, maximal length of the corresponding extended peptide, and average length in native PGK

	m^a						
	1aa	2aa	3aa	4aa	5aa	6aa	7aa
$d_{\text{Native PGK}}^b (\text{Å})$	3.80	6.00	7.41	9.10	10.85	12.33	13.53
$d_{\text{EVC sphere}}^c (\text{Å})$	4.70	6.95	8.65	10.80	12.10	13.75	14.90
$L_{\text{max}}^d (\text{Å})$	3.80	7.30	10.90	14.60	18.20	21.80	25.50

^a m = number of residues per sphere.

^b $d_{\text{Native PGK}}$ = average length of corresponding number of residues in native PGK.

^c $d_{\text{EVC sphere}}$ = diameter of corresponding EVC sphere.

^d L_{max} = length of corresponding fully-extended peptide chain

polypeptide chain, a value of $\sim 17\%$ is obtained for the 5-residue-per-sphere model. This result suggests that in strongly-denatured PGK the dramatic increase of R_g , from ~ 23 Å in the native state to ~ 90 Å, is associated mainly with long-range flexibility of the chain, i.e., at distances > 25 Å, rather than a strongly-increased short-range flexibility that would lead to larger decreases in the local chain density.

Discussion

The amino acid sequence of a protein can be expected to influence its unfolded states and folding pathways. Proteins may aggregate or remain compact even under the most extreme conditions of heat and chemical denaturant. However, the present and previous data indicate that PGK unfolds into a random coil-like structure in 4 M GdmCl with a large increase in the radius of gyration (from ~ 24 Å in the native state to ~ 90 Å), without aggregating under the conditions studied here. It may therefore provide a convenient system for examining the properties of a "fully-unfolded" protein chain.

Among the properties of interest is excluded volume. The present work investigates the effect of excluded volume on the configurational distribution and SANS profile of the denatured PGK chain. The major conclusions are that excluded volume does play an important role and that, in contrast to the simpler, freely-jointed chain description, the microscopic model of the protein obtained is consistent with known geometric limitations imposed by the covalent geometry of the peptide chain. Both EVC and FJC fit the experiment at low q (< 0.02 Å $^{-1}$), i.e., that part of the profile corresponding to long length to scale correlations. However, while EVC is also able to describe the experimental profile at higher q , FJC fails to do so. The reason for this is that the sphere diameter strongly influences the form of the scattering profile at high q . As a consequence of the too-large sphere diameters, FJC models with the experimental R_g underestimate the high- q SANS intensity.

The applicability of EVC is consistent with examination of the data using analytical random polymer theory, in which the chain is described as being made up of rigid rods of length b . For $3R_g^{-1} < q < b^{-1}$ the scattering profile then follows the relation $P(q) = P_\infty x^{-1/2\nu}$ where $x = (qR_g)^2$ and ν is the "excluded volume exponent" (Flory, 1953). The values of P_∞ and of ν are sensitive to the chain statistics, and have been found to be 1.3 and 0.588 for an excluded-volume chain (Le Guillou & Zinn-Justin, 1977; Rawiso et al., 1987; Destree et al., 1996; Pedersen & Schurtenberg, 1996). Fitting to the present SANS data for strongly-denatured PGK over $0.02 < q < 0.1$ Å $^{-1}$ gives values of $P_\infty = 1.3 \pm 0.15$ and $\nu = 0.58 \pm 0.02$, consistent with the theoretical excluded volume behavior (Calmettes et al., 1994). The agreement with experiment of the present finite EVC spheres model is consistent with this result.

The excluded volume effect, which we have discussed here primarily in terms of intrachain interactions, also acts between chains; this leads to significant repulsion between the different protein molecules, which prevents the chains from interpenetrating mutually. This is characterized by the second virial coefficient, which, for excluded volume chains, is given by $A_2 = 0.96\pi^{3/2}N_A R_g^3/m_w^2$ (Lapp et al., 1992). Using the monomeric mass for PGK and the present experimental R_g leads to a value of 11.9×10^{-4} cm 3 mol/g 2 for the strongly-denatured protein, in accord with the present experimental result for A_2 of $(11.5 \pm 1.5) \times 10^{-4}$ cm 3 mol/g 2 . Finally, we note that the excluded volume chain properties of the present denatured

state may help explain the rather large change in R_g on denaturation relative to smaller proteins that have been examined by SAXS (Semisotnov et al., 1996). For a compact sphere R_g is proportional to $N^{0.33}$ where N is the number of residues. The radius of gyration of an excluded volume chain, R_{gE} , is proportional to $N^{0.60}$ (Flory, 1953). Consequently, the ratio, R_{gE}/R_g is expected to increase with N . As PGK has a relatively high molecular weight, the increase in R_g on full denaturation is expected to be proportionally larger than for smaller proteins.

The present results clearly indicate that the configurationally-averaged spatial distribution of scattering density in a strongly-denatured protein is well described by the EVC chain of spheres. However, small but significant differences exist between the non-interpenetrating EVC profiles and experiment in the region 0.08 – 0.20 Å $^{-1}$. Refinement of the EVC model so as to allow partial interpenetration of the spheres results in a profile in agreement with experiment over the whole accessible q -range. Physically, the partial interpenetration can be viewed as displacement of some of the solvent or protein atoms of one sphere by some of the atoms of another. This may be due, for example, to side-chain:side-chain interactions. In the best-fitting model, this interpenetration is present only for third- or higher-order neighbors of a $m = 5$ chain, i.e., between segments of the chain at least 11 residues apart. The partial sphere interpenetration results, therefore, may be evidence for nonlocal interaction between the protein chains at long range in the sequence. This is random in the present model and does not include any persistent sequence-specific nonlocal interaction. If present, this chain interaction does not seem to be sufficient to allow significant aggregation of the protein molecules.

Experimentally, the solution scattering profile is spherically averaged, leading to a corresponding loss of information content relative to the hypothetical nonaveraged case. Moreover, the present data are limited to low resolution, leading to a lack of information on the details of the interaction present. Consequently, the model presented here is not necessarily the only one that can explain the data. For example, one simple modification might involve the use of an elastic soft-sphere potential instead of a reduced hard-core. It is also conceivable that nonspherical geometries, such as cylinders, might also be in accord with experiment. Among the modifications of the present model that might be applied is the variation of the sphere radius along the polypeptide chain, according to the sizes of the side chains present, although the experimental errors in the high- Q region are such that this is unlikely to significantly affect the accuracy of the reproduction of the experimental data in the present case. Another approach would be to adopt ideas from random polymer theory in which local geometries can be represented by a finite chain cross section and a persistence length. However, all models in accord with the present experiment can be expected to possess the basic excluded volume and chain flexibility properties of the present model. The model, in principle, can be tested using other experimental techniques, such as NMR spectroscopy and fluorescence energy transfer. In particular, the present model does predict the absence of persistent nonlocal interactions, i.e., stable interactions among residues distant in the sequence. Such interactions can lead to NMR or fluorescence signals significantly different from the random case (Neri et al., 1992; Gottfried & Haas, 1992).

The results here indicate a significant increase in chain flexibility on denaturation, as evidenced by the number, and hence size, of spheres in the chain. Whereas the native protein scattering can be approximately reproduced with only two or three spheres (i.e.,

$\sim 10^2$ residues per sphere) (Calmettes et al., 1993), the best-fitting models for the strongly-denatured state are 4 residues per sphere for FJC and 5 for EVC. Consequently the significant chain flexibility present in the fitted FJC models is maintained in the present EVC description. The results suggest that in strongly-denatured PGK the dramatic increase of R_g is associated mainly with the introduction of chain flexibility at distances >25 Å (twice the sphere diameter) rather than strongly-increased shorter-range flexibility that would lead to larger decreases in the local chain density than derived here.

A final question concerns the possible evolution of the present modeling toward an all-atom description of the strongly-denatured protein chain. This work is in progress and may enable several questions to be addressed. One of these is whether the presence of internal conformational preferences within the EVC spheres can influence the form of the scattering profile. Another concerns the sensitivity of the scattering curves to the solvent distribution around the protein chain. Atomic-detail models of strongly-denatured states might finally allow the buried surface area on protein folding to be quantitatively determined; whereas the surface area is reasonably well defined for the native state it is highly model-dependent for unfolded proteins (Creamer et al., 1997). Atomic-detail models would also be useful as starting points for explicit, all-atom molecular dynamics simulations of early events in protein folding. Finally, we note that, in principle, the present EVC description might be used as bounds for starting low-resolution protein folding simulations of PGK, using, for example, lattice models (Wolynes et al., 1996; Dill & Chan, 1997).

Acknowledgments

One of the authors (A.-J.P.) was partially financed by a PECO grant obtained by the Laboratoire Léon Brillouin; he wishes to acknowledge the European Community mobility schemes, which made possible this collaboration.

References

- Alexandrescu AT, Evans PA, Pitkeathly M, Baum J, Dobson CM. 1993. Structure and dynamics of the acid-denatured molten globule state of alpha lactalbumin—A two-dimensional NMR study. *Biochemistry* 32:1707–1718.
- Alexandrowicz Z. 1969. Monte Carlo of chains with excluded volume: A way to evade sample attrition. *J Chem Phys* 51:561–565.
- Boyd RH, Phillips PJ. 1993. *The science of polymer molecules*. Cambridge: Cambridge University Press.
- Bucher T. 1955. Phosphoglycerate kinase from brewer's yeast. *Methods Enzymol* 1:415–422.
- Calmettes P, Durand D, Desmadril M, Minard P, Receveur V, Smith JC. 1994. How random is a highly denatured protein? *Biophys Chem* 53:105–114.
- Calmettes P, Durand D, Smith JC, Desmadril M, Minard P, Douillard R. 1993a. Structure of proteins unfolded by guanidinium chloride. *J Phys IV (France)* 3:C8–253.
- Calmettes P, Roux B, Durand D, Desmadril M, Smith JC. 1993b. Configurational distribution of denatured phosphoglycerate kinase. *J Mol Biol* 231:840–848.
- Chan HS, Dill KA. 1991. Polymer principles in protein structure and stability. *Annu Rev Biophys Biomol Str* 20:447–490.
- Chen L, Hodgson KO, Doniach S. 1996. A lysosome folding intermediate revealed by solution X-ray scattering. *J Mol Biol* 261:568–671.
- Creamer TP, Srinivasan R, Rose G. 1997. Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility. *Biochemistry* 36:2832–2835.
- Debye P. 1947. Molecular weight determination by light scattering. *J Phys Colloid Chem* 51:18–32.
- Des Cloizeaux J, Jannink G. 1990. *Polymers in solution: Their modeling and structure*. Oxford: Clarendon Press.
- Destrée M, Lyulin A, Ryckaret JP. 1996. Monte Carlo prediction of the structure factor of polyethylene in good and θ solvents. *Macromolecules* 29:1721–1727.
- Dill KA, Chan HS. 1997. From Levinthal pathways to funnels. *Nature Struct Biol* 4(1):10–19.
- Eliezer D, Jennings PA, Wright PE, Doniach S, Hodgson KO, Tsuruta H. 1995. The radius of gyration of an apomyoglobin folding intermediate. *Science* 270:487–488.
- Flory PJ. 1949. The configuration of real polymer chains. *J Chem Phys* 17:303–308.
- Flory PJ. 1953. *Principles of polymer chemistry*. Ithaca, NY: Cornell University Press.
- Fox TG Jr, Flory PJ. 1949. The configuration of real polymer chains. *J Phys Colloid Chem* 53:197–212.
- Gottfried DS, Haas E. 1992. Nonlocal interactions stabilise compact folding intermediates in reduced unfolded BPTI. *Biochemistry* 31:12353–12359.
- Higgins JS, Benoit HC. 1994. *Polymers and neutron scattering*. Oxford: Clarendon Press.
- Kassenbrock CK, Garcia PD, Walter P, Kelly RB. 1988. Heavy chain binding protein recognizes aberrant polypeptides translocated in vitro. *Nature* 333:90–93.
- Kataoka M, Goto Y. 1997. X-ray solution scattering studies of protein folding. *Folding Design* 1:107–114.
- Kataoka M, Hagiwara Y, Miahara K, Goto Y. 1993. Molten globule of cytochrome *c* studied by small angle X-ray scattering. *J Mol Biol* 229:591–596.
- Kataoka M, Nishii I, Fujisawa T, Ukey T, Tokunaga F, Goto Y. 1995. Structural characteristics of the molten globule and native states of apomyoglobin by solution X-ray scattering. *J Mol Biol* 249:215–228.
- Kim PS, Baldwin RL. 1990. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* 59:631–660.
- Kremer K. 1996. Computer simulation methods for polymer physics. In: Binder K, Ciccotti G, eds. *Monte Carlo and molecular dynamics of condensed matter systems*. Bologna: SIF, pp 673–723.
- Kleinert H. 1990. *Path integrals in quantum mechanics, statistics and polymer physics*. Singapore: World Scientific.
- Kuwajima K. 1989. The molten globule as a clue for understanding the folding and cooperativity of globular protein structure. *Proteins Struct Funct Genet* 8:1–103.
- Lapp A, Motin M, Strazielle C, Brosetta D, Liebler L. 1992. Experimental evidence for the strong interpenetration of two polymer chains very different in size. *J Phys II (France)* 2:1247–1256.
- Lamy A, Smith JC. 1996. Denaturation of truncated *Staphylococcal* nuclease in molecular dynamics simulation at 300 K. *J Am Chem Soc* 118:7326–7328.
- Le Guillou JC, Zinn-Justin J. 1977. Critical exponents for the n -vector model in three dimensions from field theory. *Phys Rev Lett* 39:95.
- Miranker A, Dobson CM. 1996. Collapse and cooperativity in protein folding. *Curr Opin Struct Biol* 6:31–42.
- Neri D, Billeter M, Wider G, Wuthrich K. 1992. NMR determination of residual structure in a urea-denatured protein; the 434 repressor. *Science* 257:1559.
- Pedersen JS, Schurtenberg P. 1996. Scattering function of semiflexible polymers with and without excluded volume effects. *Macromolecules* 29:7602–7612.
- Petrescu AJ, Butters TD, Reikensmeier G, Petrescu SM, Platt FM, Dewk RA, Wormald MR. 1997a. The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *EMBO J* 16:4302–4310.
- Petrescu AJ, Receveur V, Calmettes P, Durand D, Desmadril N, Roux B, Smith JC. 1997b. Small-angle neutron scattering by a strongly-denatured protein: Analysis using random polymer theory. *Biophys J* 72:335–342.
- Radford SE, Dobson CM, Evans PA. 1992. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 358:302–307.
- Rawiso M, Duplessix R, Picot C. 1987. Scattering function of polystyrene. *Macromolecules* 20:630–648.
- Receveur V. 1997. Etude structurale et dynamique des états dénaturés de la phosphoglycerate kinase de levure par diffusion des neutrons et des rayons X [Thèse de Docteur en Sciences]. Orsay: Université Paris-Sud.
- Receveur V, Calmettes P, Smith JC, Desmadril N, Coddien G, Durand D. 1997. Picosecond dynamical changes on denaturation of yeast phosphoglycerate kinase revealed by quasielastic neutron scattering. *Proteins Struct Funct Genet* 28:380–387.
- Sali A, Shakhnovitch E, Karplus M. 1994. How does a protein fold? *Nature* 369:248–251.
- Schmidt M, Buchner J. 1992. Interaction of GroEL with an all-beta-protein. *J Biol Chem* 267:16829–16835.
- Semisotnov GV, Kihara H, Kotova N, Kimura K, Amemiya Y, Wakabayashi K, Serdyuk I, Timchenko AA, Chiba K, Nikaido IN, Ikura T, Kuwajima K.

1996. Protein globularization during folding: A study by synchrotron small-angle X-ray scattering. *J Mol Biol* 262:559–574.
- Shortle DR. 1996. Structural analysis of non-native states of proteins by NMR methods. *Curr Opin Struct Biol* 6:24–30.
- Tanford C. 1968. Protein denaturation. *Adv Protein Chem* 23:121–275.
- Wolynes PG, Luthey-Schulten Z, Onuchic JN. 1996. Fast folding experiments and the topography of protein folding energy landscapes. *Chem Biol* 3:425–432.
- Wright PE, Jennings PA. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* 262:892–895.
- Zapun A, Petrescu SM, Rudd PM, Dwek RA, Thomas DY, Bergeron JJM. 1997. Conformation independent binding of monoglucosylated ribonuclease B to calnexin. *Cell* 88:29–38.
- Zimm BH. 1948. The scattering of light and the radial distribution function of high-polymer solutions. *J Chem Phys* 16:1093–1105.