Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques †

ARTICLE *in* BIOCHEMISTRY · JANUARY 2000

Impact Factor: 3.02 · DOI: 10.1021/bi991765q · Source: PubMed

6 AUTHORS, INCLUDING:

Jonathan A Jones
University of Oxford
88 PUBLICATIONS **4,920** CITATIONS

SEE PROFILE

Articles

Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques†

Deborah K. Wilkins, Shaun B. Grimshaw, Véronique Receveur,‡ Christopher M. Dobson, Jonathan A. Jones,§ and Lorna J. Smith*

Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, England

Received July 28, 1999; Revised Manuscript Received October 4, 1999

ABSTRACT: Pulse field gradient NMR methods have been used to determine the effective hydrodynamic radii of a range of native and nonnative protein conformations. From these experimental data, empirical relationships between the measured hydrodynamic radius (R_h) and the number of residues in the polypeptide chain (N) have been established; for native folded proteins $R_h = 4.75N^{0.29}$Å and for highly denatured states $R_h = 2.21N^{0.57}$Å. Predictions from these equations agree well with experimental data from dynamic light scattering and small-angle X-ray or neutron scattering studies reported in the literature for proteins ranging in size from 58 to 760 amino acid residues. The predicted values of the hydrodynamic radii provide a framework that can be used to analyze the conformational properties of a range of nonnative states of proteins. Several examples are given here to illustrate this approach including data for partially structured molten globule states and for proteins that are unfolded but biologically active under physiological conditions. These reveal evidence for significant coupling between local and global features of the conformational ensembles adopted in such states. In particular, the effective dimensions of the polypeptide chain are found to depend significantly on the level of persistence of regions of secondary structure or features such as hydrophobic clusters within a conformational ensemble.

In the characterization of native or nonnative states of proteins in solution, a measurement of the molecular dimensions of the system is of considerable importance. For native folded proteins this is particularly valuable when dimers or higher oligomers are formed or when the system is highly anisotropic (*1*−*3*). In the case of nonnative states an ensemble of interconverting conformers will in general be populated (*4, 5*). Here a measure of the average dimensions of the conformational ensemble can give information about the nature of the structures adopted by the polypeptide chain (*6*−*9*). This provides an opportunity for probing the coupling of the local and global conformational properties of these unfolded and partly folded species. This is of importance for understanding the role of these conformationally disordered states in protein folding, stability, and aggregation (*10*−*12*).

Most studies of molecular dimensions have relied on measurements of the effective radius of gyration by small-angle scattering studies using neutrons (SANS)[1] or X-rays (SAXS) (*6, 8, 13*). Alternatively, the hydrodynamic or Stokes radii have been determined via translational diffusion coefficient measurements using dynamic light scattering (*14*). Both of these approaches have been applied to the study of a number of proteins in their native states and also in a variety of nonnative conformations. The systems studied include denatured states of phosphoglycerate kinase (*15, 16*) and ribonuclease A (*17, 18*) in guanidine hydrochloride, and partially folded molten globule states of apomyoglobin, bovine α-lactalbumin, and cytochrome *c* (*19*−*24*). Recently time-resolved measurements by these techniques have also been used to study the changes in dimensions during protein folding (*21, 23, 25, 26*).

We have previously reported the effective dimensions of both the native and nonnative states of lysozyme determined by pulse field gradient (PFG) NMR techniques (*27*). We showed that through diffusion measurements performed with these techniques the changes in the average hydrodynamic radius of the protein resulting from urea denaturation could be followed. Similar NMR methods have been reported by

† This is a contribution from the Oxford Centre for Molecular Sciences, which is supported by the U.K. Engineering and Physical Sciences Research Council, the Biotechnology and Biological Sciences Research Council, and the Medical Research Council. The research of C.M.D. is supported in part by the Wellcome Trust and by an International Research Scholars award from the Howard Hughes Medical Research Institute. L.J.S. and J.A.J. are Royal Society University Research Fellows. V.R. was supported by an EMBO fellowship.

* Corresponding author: Phone: +44−1865−275961; FAX +44−1865−275921; E-mail lorna.smith@chemistry.oxford.ac.uk.

‡ Present address: AFMB−CNRS, 31 Chemin Joseph Aiguier, 13402 Marseille, France.

§ Present address: Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, England.

[1] Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; NOE, nuclear Overhauser enhancement; PFG, pulse field gradient; PGK, phosphoglycerate kinase; PG-SLED, pulse gradient stimulated echo longitudinal encode−decode; R_h, hydrodynamic radius; RMSD, root-mean-square difference; SANS, small-angle neutron scattering; SAXS, small-angle X-ray scattering.

10.1021/bi991765q CCC: $18.00
Published on Web 11/24/1999

Table 1: Hydrodynamic Radii for Native and Denatured States of Proteins Measured by the PFG NMR Method[a]

(A) Folded Proteins

protein	R_h (Å)	no. of residues
bovine pancreatic trypsin inhibitor	15.8	58
SH3 domain of PI3 kinase	18.6	90[b]
horse heart cytochrome *c*	17.8	104
hen lysozyme	20.5	129
horse myoglobin	21.2	153
S. equisimilis streptokinase	36.0	414
yeast triosephosphate isomerase	29.7	494 (dimer)

(B) Polypeptide Chains without Disulfide Bridges under Strongly Denaturing Conditions

peptide or protein	conditions	R_h (Å)	no. of residues
residues 49−64 from hen lysozyme	8 M urea, pH 2	10.4	16
residues 17−37 from D3 of fibronectin binding protein[c]	8 M urea, pH 7	13.3	22
residues 2−38 from D3 of fibronectin binding protein[c]	8 M urea, pH 7	15.5	32
SH3 domain of PI3 kinase	3.5 M GuHCl, pH 7.2[d]	27.5	90[b]
horse heart cytochrome *c*	pH 2	32.6	104
hen lysozyme	8 M urea, pH 2, reduced and methylated	34.6	129
yeast triosephosphate isomerase	2.5 M GuHCl, pH 7.2[d]	49.7	247 (monomer)

[a] The errors in the hydrodynamic radii values estimated from repeat measurements are ±0.25 Å. [b] Sequence includes a two-residue N-terminal extension and a four-residue C-terminal extension to the native SH3 sequence. [c] Sequence taken from the third D repeat in the fibronectin binding domain of *S. aureus* fibronectin binding protein variant A (FnBPA) (*59*). [d] The SH3 domain sample contained 20 mM H_3PO_4, and the triosephosphate isomerase sample contained 5 mM Tris and 100 mM NaCl.

Pan et al. (*9*), who measured the dimensions of native BPTI and partially folded and unfolded variants of the protein that contain one or no disulfide bridges. The PFG NMR method has also been used in a number of studies that have probed aggregation and the formation of oligomers in protein samples (*2*, *3*, *28*, *29*). The NMR approach has particular appeal as the measurements of hydrodynamic radius can be made on samples identical to some of those used to characterize the local properties of the polypeptide chain. This is particularly important for nonnative states, where recent studies have shown that, under favorable conditions, multidimensional heteronuclear NMR techniques can give a wealth of data about local conformational preferences and the dynamics of the polypeptide chains in these disordered states (*30*, *31*).

Here we report results from PFG NMR methods designed to measure the hydrodynamic radii of a range of native and denatured proteins. In this work dioxan was included in the protein solutions to act as a viscosity probe, enabling values of the hydrodynamic radii to be determined relative to a protein of known dimensions (*27*). By studying a number of proteins both in their native folded state and under strongly denaturing conditions, it has been possible to characterize the relationships between the effective hydrodynamic radii and the length of the polypeptide chain for the different states. These results have been compared with other protein radius data reported in the literature and with predictions from theory. We demonstrate that the results provide an approach in which hydrodynamic radius measurements made for partially folded protein species can be interpreted.

MATERIALS AND METHODS

Protein and Peptide Samples. Hen lysozyme, yeast triosephosphate isomerase, horse heart cytochrome *c*, horse myoglobin, and bovine pancreatic trypsin inhibitor were purchased from Sigma Chemical Co. (U.K.). Streptokinase from *Streptococcus equisimilis* and fibronectin binding protein D1−D4 from *Staphylococcus aureus* were supplied by SmithKline Beecham Pharmaceuticals. The SH3 domain of PI3 kinase was expressed in *Escherichia coli* as described previously (*32*). The peptides with sequences corresponding to residues 49−64 from hen lysozyme and residues 17−37 from D3 of fibronectin binding protein were synthesized on an Applied Biosystems 430A automated peptide synthesizer, by standard 9-fluorenylmethoxycarbonyl (Fmoc) methodology, and were purified by reverse phase HPLC. The purity and integrity of the peptides were checked by electrospray mass spectrometry. The peptide with a sequence corresponding to residues 2−38 from D3 of fibronectin binding protein was obtained from Bachem California. Hen lysozyme with the disulfide bonds reduced and S-methylated was prepared by methods described previously by Ching-Li and Atassi (*33*) and was characterized by mass spectrometry.

Protein concentrations of approximately 1 mM were used for the NMR samples unless otherwise stated, 20 μL of 1% dioxan in D_2O being added to all the samples. For the native-state samples the following conditions were used: BPTI, pH 4.5; cytochrome *c*, pH 2.0 (holoprotein); hen lysozyme, pH 2.0; myoglobin, pH 6.0 (holoprotein); SH3 domain, pH 7.2 in 20 mM H_3PO_4; triosephosphate isomerase, pH 7.2 in 5 mM Tris and 100 mM NaCl; streptokinase, pH 7.0 in 20 mM H_3PO_4. The conditions used for the protein samples under highly denaturing conditions are given in Table 1. Where used, the concentration of urea in solutions was determined by measurement of the refractive index (*34*). For the sample of hen lysozyme in which the disulfide bonds were reduced and methylated, an increase in the apparent protein hydrodynamic radius was observed with increasing protein concentration (in the range 0.25−2 mM) in both the absence and presence of 8 M urea. This result could reflect the presence of some aggregated species within the lysozyme sample solution. The values of the hydrodynamic radii for this system were therefore obtained by extrapolation to zero protein concentration.

NMR Experiments. All NMR spectra were recorded on home-built spectrometers at the Oxford Centre for Molecular Sciences with ^1H operating frequencies of 500.15 and 600.20 MHz. The spectrometers are equipped with Oxford Instruments Co. magnets (Oxford, U.K.), Omega software and digital control equipment (Bruker Instruments), home-built triple-resonance pulsed-field-gradient probe heads (*35*) and

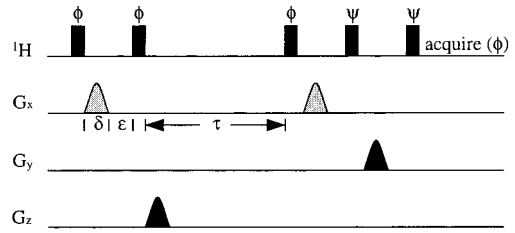

FIGURE 1: PG-SLED sequence. The phase cycle ($\phi = x, x, y, y, -x, -x, -y, -y$; $\psi = x, -x, y, -y, -x, x, -y, y$) combines the standard CYCLOPS cycle, to remove acquisition artifacts, with phase alternation of the longitudinal encode sequence, which causes longitudinal relaxation to be manifested as an exponential decay (*61*). The diffusion labeling gradients, shown in gray, were varied, while the crush gradients, shown in black, were applied at full strength. All gradient pulses were shaped either as sine waves or as composite sine waves (*36*); their integrated intensity was adjusted for each sample to give optimal results. The diffusion gradients were applied along the (transverse) *x*-axis, so as to minimize the effects of convection, while the two crush gradients were applied along the *y*- and *z*-axes.

home-built linear amplifiers. The NMR experiments were performed at 20 °C.

PFG NMR diffusion measurements were performed with the PG-SLED (pulse gradient stimulated echo longitudinal encode−decode) sequence shown in Figure 1 (*27*), incorporating composite sine gradient pulses (*36*). The lengths of all pulses and delays in this sequence were held constant, and 20 spectra were acquired with the strength of the diffusion gradient varying between 5% and 100% of its maximum value; each measurement was repeated five times to improve the signal-to-noise ratio and to allow the measurement error to be estimated. The lengths of the diffusion gradient and the stimulated echo were optimized for each sample to give a total decay in the protein signal of between 80% and 90%. For example, experiments on native hen lysozyme used a gradient time of $\delta = 6.3$ ms, a recovery time $\epsilon = 0.7$ ms, and an echo time of $\tau = 100$ ms. NMR spectra were acquired with 8K complex points, with a typical spectral width of 8000 Hz. Data processing was performed with Felix 2.3 from Biosym Technologies (San Diego, CA). The baseline was fitted to a zero-order polynomial and integration was performed with home-written software.

The decay in the protein signal from the PFG NMR experiments was in most cases fitted to a single Gaussian. However, an additional function was also used in the fits of the decays for a number of the denatured protein samples where there was a degree of protein aggregation. In the case of the SH3 domain at low pH, the formation of aggregates resulted in a gradual reduction with time in the protein concentration that could be observed by NMR. This led to an additional signal loss that was accounted for by multiplying the Gaussian function used to fit the diffusion coefficient by a linear decay term. For reduced methylated hen lysozyme, some large aggregates were present within the sample, which give a very broad NMR signal. In this case a Gaussian curve having an offset to fit the broad signals greatly improved the consistency of the radii obtained. For the large aggregates the decay in intensity is very small over the range of gradient strengths used to monitor the monomeric states. Thus the slow Gaussian decay of these large molecules can be successfully approximated by a constant offset.

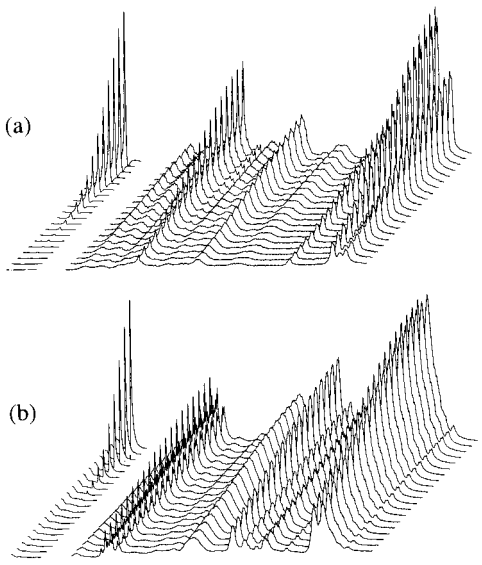

FIGURE 2: Experimental PG-SLED spectra of (a) a peptide corresponding to residues 17−37 from D3 of fibronectin binding protein in water and (b) hen lysozyme with the disulfide bridges reduced and methylated in 8 M urea. The diffusion gradients were varied between 5% and 100% of their maximum value (about 60 G cm^{-1}). Regions of the spectrum corresponding to the dioxan reference signal (around 3.6 ppm) are shown on the left, while the main aliphatic regions (3.3−0.3 ppm) are shown on the right; the dioxan signals are plotted on a reduced vertical scale. The signals from the fibronectin binding protein peptide decay more slowly than those of the reference as dioxan diffuses more rapidly than the peptide. In the spectra from denatured lysozyme, the protein signals decay much more slowly than the reference signals, indicating the large size of this unfolded protein.

RESULTS

(a) PFG NMR Hydrodynamic Radius Measurements. The effective hydrodynamic radii of seven proteins in their folded native states and seven peptides and proteins under strongly denaturing conditions were determined using PFG NMR techniques. As described in detail in the Materials and Methods section, for each system studied a series of 1D spectra was produced by the PG-SLED pulse sequence, each of which was recorded with a different gradient strength (Figure 2). In most cases fitting the intensities of the signals from the protein (*s*) as a function of gradient strength (*g*) to

$$s(g) = A\mathrm{e}^{-dg^2} \tag{1}$$

enabled the observed decay rate (*d*), which is proportional to the diffusion coefficient *D*, to be obtained. For two of the denatured polypeptide chains a more complex function was required for the fit to extract the decay rate as described in the Materials and Methods section.

A value of the hydrodynamic radius can be calculated from measurements of the diffusion coefficient by means of the Stokes−Einstein equation. However, here we have used a different approach and have included a reference molecule in the protein solution (*27*). Decay rates have been extracted from the spectra for both the protein (d_{prot}) and the reference molecule (d_{ref}) and the protein hydrodynamic radius has been determined from:

$$R_{\mathrm{h}}^{\mathrm{prot}} = \frac{d_{\mathrm{ref}}}{d_{\mathrm{prot}}}(R_{\mathrm{h}}^{\mathrm{ref}}) \tag{2}$$

where R_h^{prot} is the protein hydrodynamic radius and R_h^{ref} is the radius of the reference molecule. This approach, unlike calculations using the Stokes−Einstein equation, has the advantages that it does not require knowledge of the viscosity of the solution and does not rely on interpretation of absolute values of diffusion coefficients; such values are subject to errors both from inaccuracies in measuring the exact shape and strength of the gradient pulse and from spin-echo instability (*37*).

In this work dioxan has been used as the reference molecule. Studies with native and denatured (in 8 M urea at pH 2) hen lysozyme suggest that dioxan does not interact to a significant extent with either folded or denatured polypeptide chains (*27*). To determine the appropriate radius for dioxan to use in eq 2, we have calibrated the measurements using the radius of gyration measured for native hen lysozyme by SAXS (15.3 Å; *38*). The radius of gyration (R_g) is the root-mean-squared distance of all the atoms in the molecule from the protein center of mass, while the hydrodynamic radius (R_h) can be defined as the radius of a sphere that has the same diffusion coefficient as the protein molecule. The relationship between the hydrodynamic radius and the radius of gyration is highly dependent on the molecular shape and is related by a constant ρ:

$$R_g = \rho R_h \tag{3}$$

with ρ being predicted for spherical molecules to be $(3/5)^{1/2}$ (*39*). By use of this relationship along with the SAXS radius of gyration and the value of d_{ref}/d_{prot} measured experimentally for hen lysozyme (9.33; *27*), the effective hydrodynamic radius of dioxan is calculated to be 2.12 Å. This radius has been used with eq 2 to convert all experimentally measured d_{ref}/d_{prot} ratios in this study into values of the hydrodynamic radius. These R_h values tend to be larger than those estimated from crystal structures as they include a hydration layer (*9*).

Analysis of the measured hydrodynamic radii (Table 1) for native folded proteins known from structural studies to be essentially spherical indicates that there is a strong correlation between the molecular dimensions and the length of the polypeptide chain. As shown by the logarithmic plot in Figure 3, the experimental data can be fitted to the empirical equation

$$R_h = (4.75 \pm 1.11)N^{0.29\pm0.02} \tag{4}$$

where N is the number of residues in the polypeptide chain and R_h is the hydrodynamic radius (in angstroms). Deviations are to be expected from this relationship, however, for systems that are highly anisotropic. The measured hydrodynamic radius for native streptokinase provides an example of this. Here the experimental hydrodynamic radius is 36.0 Å, significantly larger than the value of 27.3 Å predicted for a protein of 414 amino acid residues by eq 4. The experimental result is, however, in excellent agreement with previous measurements of the hydrodynamic radius of streptokinase by dynamic light scattering (R_h 35.8 Å; *1*). In accord with this, NMR studies have shown that the protein consists of three structurally independent domains with an unstructured C-terminal tail (*40*).

For characterizing the denatured state we have concentrated on proteins without disulfide bridges, as the presence

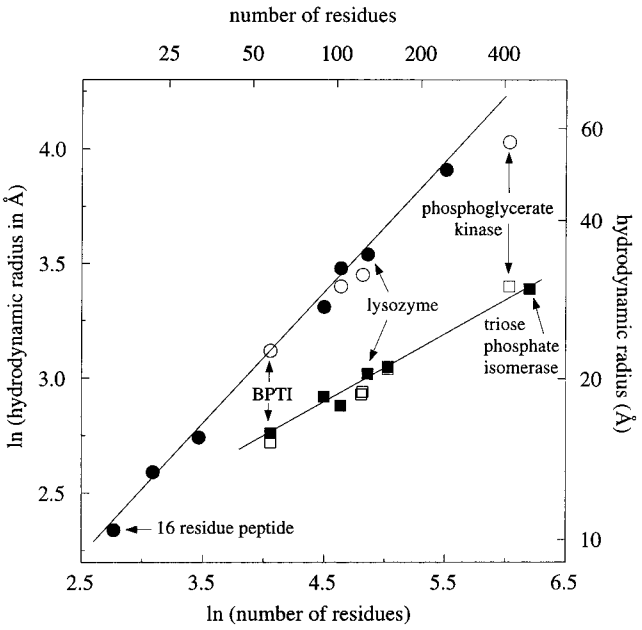

FIGURE 3: Plot of the $\log_e$ of the hydrodynamic radius versus the $\log_e$ of the number of residues in the polypeptide chain. The values determined in this work for native folded proteins (■) and highly denatured polypeptide chains (●) are shown with filled symbols. The line fitted to these data for the native proteins has a slope of 0.29 ± 0.02 and a *y*-axis intercept of 1.56 ± 0.1, while that fitted to the denatured protein data has a slope of 0.57 ± 0.02 and a *y*-axis intercept of 0.79 ± 0.07. Values of hydrodynamic radii reported in the literature from dynamic light scattering or PFG NMR studies are shown by open symbols (native folded proteins, □; highly denatured proteins, ○). The literature data used are for horse apocytochrome *c* (*62*), yeast phosphoglycerate kinase (*16*, *63*), sperm whale apomyoglobin (*19*), bovine pancreatic ribonuclease A (*18*), bovine pancreatic trypsin inhibitor (*9*), and bovine α-lactalbumin (*23*). Selected data points are labeled with the protein name.

of these, and their exact positions within the amino acid sequence, would be expected to perturb significantly the molecular dimensions. We have also studied proteins under strongly denaturing conditions (in high concentrations of urea or guanidine hydrochloride or at extremes of pH), where it is anticipated that there will be few persistent structural features. We refer to these states as highly denatured proteins. Such unfolded polypeptide chains must be described in terms of an ensemble of interconverting conformers (*4*). For all the highly denatured proteins that are discussed here, interconversion between the various populated conformers was fast on both the NMR and diffusion time scales. Consequently an average effective hydrodynamic radius of the unfolded ensemble (Table 1) was extracted from the rate of decay of the protein NMR resonances with increasing gradient strength (*27*). For these unfolded proteins a correlation is also observed between the number of residues in the polypeptide chain (N) and the hydrodynamic radius (R_h in angstroms), the experimental data fitting to

$$R_h = (2.21 \pm 1.07)N^{0.57\pm0.02} \tag{5}$$

This is similar to an equation [$R_h = (2.8 \pm 0.3)N^{0.5\pm0.02}$] reported by Damaschun et al. (*16*) from data for the hydrodynamic radii of a number of proteins denatured in 6 M guanidine hydrochloride.

It is interesting to compare the experimental results from the PFG NMR studies with theoretical predictions of the

dimensions of homopolymers. The scaling law defines the dimensions of homopolymers by

$$R_g = R_0 N^{\nu} \quad (6)$$

where R_0 is a constant and N is the number of monomer units (*41*). As shown by Flory, the exponent ν is predicted to be $^3/_5$ and $^1/_3$ for fully solvated and collapsed polymer chains, respectively (*42*). For a homopolymer a fully solvated chain is observed in a good solvent in which polymer-solvent interactions are favored over polymer-polymer interactions, while a collapsed chain is seen in a bad solvent where polymer-polymer interactions are favored over polymer-solvent interactions. Under conditions where polymer-polymer interactions are equal to polymer-solvent and solvent-solvent interactions, in a so-called Flory or θ solvent, the polymer behaves as a Gaussian chain and the exponent ν is predicted to be $^1/_2$ (*41*).

The empirical equation determined here relating the experimental hydrodynamic radii of native proteins and the number of residues in the polypeptide chain with a value of 0.29 ± 0.02 for the exponent ν therefore resembles the value of 0.33 predicted for a collapsed homopolymer. This reflects the fact that, despite the significant differences between the interactions of a native protein with the solvent and the interactions of a homopolymer with a bad solvent, in both cases a compact globular fold is adopted. This result is in agreement with previous conclusions from an analysis using radii of gyration calculated from protein X-ray structures. The study gave a value for the exponent in the scaling law for native proteins of 0.35 ± 0.03 (*43*).

For the highly denatured proteins, the empirically derived value of 0.57 ± 0.02 for the exponent is similar to that expected for a polymer in a good solvent (0.6). Furthermore, small-angle neutron scattering experiments combined with polymer theory lead to a similar value (0.58 ± 0.02) of the exponent ν for strongly denatured phosphoglycerate kinase. These measurements also provide independent evidence that in this case the unfolded protein behaves as a homopolymer in a good solvent with excluded volume (*44*, *45*). Although the details of the mechanism of protein denaturation on the addition of chaotropes such as urea is not well understood, there is a similarity to the behavior of a good solvent in that it is thought that they encourage solvation of all portions of a polypeptide chain (*46*, *47*). Recent molecular dynamics simulations of barnase in 8 M aqueous urea, for example, show an accumulation of urea molecules near the protein surface. The urea molecules were found to form hydrogen bonds to exposed polar groups and, through the displacement of water molecules from the first hydration shell, lead to a greater exposure of nonpolar side chains (*48*).

(b) Comparisons with Other Measurements of Hydrodynamic Radius or Radius of Gyration. The PFG NMR experiment gives a value for the mean effective hydrodynamic radius of a protein. A comparison of the radii measured in this work with other values for hydrodynamic radii reported in the literature for native folded proteins and proteins without disulfide bridges under strongly denaturing conditions is shown in Figure 3. These data come from dynamic light scattering measurements (references given in the caption to Figure 3) and, in one case (BPTI, *9*), from a PFG NMR study. The literature hydrodynamic radii were calculated from values of diffusion coefficients, extrapolated to conditions of infinite dilution, and vary from 15.2 to 56.6 Å. A full listing of the data used in this analysis is given in the Supporting Information. For the native proteins a close similarity is evident between the hydrodynamic radii from dynamic light scattering and the values from PFG NMR studies. The RMSD between the data taken from the literature and values predicted by the empirical relationship defined in this work between hydrodynamic radius and number of residues in the polypeptide chain (eq 4) is 1.12 Å.

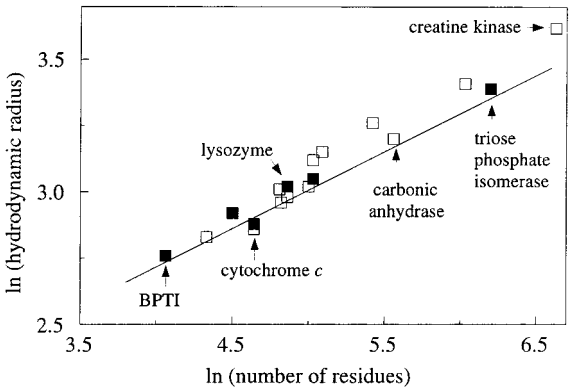

FIGURE 4: Comparison of the values of hydrodynamic radii for folded proteins determined in this work by PFG NMR techniques (■) and values of radii of gyration for native proteins (□) from SAXS or SANS studies reported in the literature. The radii of gyration have been converted into hydrodynamic radii by use of the relationship $R_g = (3/5)^{1/2} R_h$. The plot shows the $\log_e$ of the hydrodynamic radius versus the $\log_e$ of the number of residues in the polypeptide chain. The line fitted to the PFG NMR data is also shown, and selected data points are labeled with the protein name. The literature data used are for bovine ribonuclease A (*17*), staphylococcal nuclease (*64*), horse ferricytochrome *c* (*24*), yeast phosphoglycerate kinase (*49*), horse myoglobin (*20*), *Streptomyces* subtilisin inhibitor (*65*), bovine carbonic anhydrase B (*66*), bovine β-lactoglobulin (*66*), hen lysozyme (*38*), bovine α-lactalbumin (*22*), rabbit creatine kinase (*67*), and bovine ubiquitin (*67*).

In the case of the highly denatured polypeptide chains, good agreement is also seen between the literature data and the PFG NMR results reported here for three of the four proteins (apocytochrome *c*, reduced ribonuclease A, and a variant of BPTI without disulfide bridges; the RMSD between the literature data and the predictions from eq 5 is 1.92 Å). However, it seems that the values determined from dynamic light scattering experiments are slightly smaller than those from the PFG NMR measurements. This is particularly the case for phosphoglycerate kinase (PGK; 56.6 Å experimental, 68.7 Å predicted) (*16*). This may reflect the fact that measurements of hydrodynamic radii from dynamic light scattering are not determined directly but depend on additional experimental parameters such as the viscosity of the solution. Although further work is needed to investigate this issue, the general correlation between the two sets of data is good.

Figure 4 shows a comparison of the hydrodynamic radius measurements for native folded proteins determined by the PFG NMR method in this work and radii of gyration data for native proteins from small-angle scattering studies (details of the data used in this comparison are given in the Supporting Information). The radii of gyration have been converted into values of hydrodynamic radius by use of eq 3 with a ρ value of $(3/5)^{1/2}$ (*39*). With this conversion a very

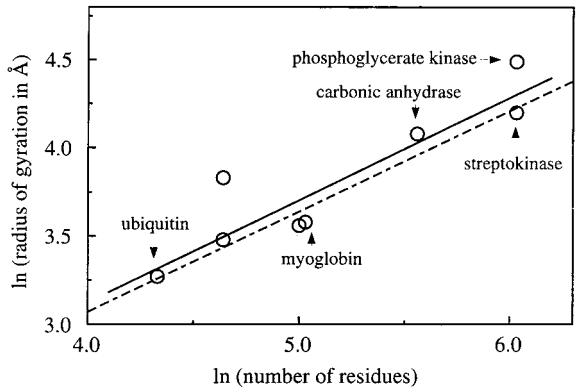

FIGURE 5: Plot of the $\log_e$ of the radius of gyration determined by SAXS or SANS versus the $\log_e$ of the number of residues in the polypeptide chain for proteins under strongly denaturing conditions. The solid line indicates the best fit to these data (slope 0.58 ± 0.11, *y*-axis intercept 0.80 ± 0.55) and the dashed line the fit for hydrodynamic radius data determined here by PFG NMR for highly denatured proteins (as shown in Figure 3). The data shown are for horse ferricytochrome *c* (*21, 62*), staphylococcal nuclease (*64*), horse myoglobin (*20*), bovine carbonic anhydrase B (*66*), *Streptococcus equisimilis* streptokinase (*16*), yeast phosphoglycerate kinase (*45*), and bovine ubiquitin (*68*). A number of the data points are labeled with the protein name.

good correlation is seen between the two sets of radius measurements. The RMSD between the radii derived from the small-angle scattering data and the predictions from the empirical equation (eq 4) derived from the PFG NMR data is 2.14 Å. This result, for proteins ranging in radius of gyration between 13.2 and 29.0 Å confirms the validity of using a constant scaling factor (ρ) for converting radii of gyration into hydrodynamic radii for native protein structures that are essentially spherical.

The radius of gyration data from small-angle scattering measurements for proteins under strongly denaturing conditions are shown in Figure 5 (details of the data used in the plot are given in the Supporting Information; only polypeptide chains without disulfide bridges are considered). There is a large scatter in the values of these radii. This probably reflects the difficulties in small-angle scattering measurements in the presence of strong denaturants due to the poor contrast between the solvent and the protein. It also results from the different analytical approaches used in the various studies to extract the radius of gyration values from the scattering data (the Guinier or Debye approximations). The Debye approximation has been postulated to be more accurate for fitting the scattering data of highly unfolded states (*49*). It gives larger radii of gyration than the Guinier approximation (*45, 50*), which may reflect a biasing with the Guinier approximation toward the conformers with smaller radii present in a denatured ensemble, as this method can fit more accurately data from compact states (*49*).

A logarithmic plot of the radius of gyration of the highly denatured proteins versus the number of residues in the polypeptide chain has a slope of 0.58 ± 0.11. Thus the value of the exponent ν is closely similar to that determined from the PFG NMR hydrodynamic radii data (0.57 ± 0.02). The fit from the NMR hydrodynamic radius data is shown by a dashed line in Figure 5. The offset between the two lines gives insight into the relationship between the mean radius of gyration and the hydrodynamic radius for an unfolded polypeptide chain under strongly denaturing conditions. In the range of polypeptide chain lengths considered (approximately 50−400 residues), the observed offset gives a ρ value ($\rho = R_g/R_h$) of 1.06 ± 0.01. It is interesting that this value is significantly lower than that predicted for a random coil (1.51) by Tanford (*51*). This difference may reflect limitations in the theoretical model used in the Tanford prediction or the nature of the assumptions used in the calculation of the experimental values of the radius of gyration and hydrodynamic radius. In light of the scatter in the radii of gyration used in the analysis, further work is clearly needed to establish the relationship between these two quantities.

DISCUSSION

The empirical relationships defined in this work between the hydrodynamic radius of a protein and the length of its polypeptide chain enable values of hydrodynamic radii to be predicted for proteins with compact spherical folds or when highly denatured (in the absence of disulfide bridges). The behavior of these two extreme states acts as a baseline that aids the interpretation of hydrodynamic radii measurements for a range of different systems. To utilize this we define a compaction factor (*C*):

$$C = \frac{R_h^D - R_h}{R_h^D - R_h^N} \quad (7)$$

where R_h^N and R_h^D are the predicted values of hydrodynamic radii for the native and fully denatured states, respectively, and R_h is the experimental hydrodynamic radius. Species with the same molecular radius as a native state will have a compaction factor of 1.0, and those with a radius the same as a fully denatured state, a value of 0.0. Such a framework is of particular importance for the interpretation of hydrodynamic radii values for partially folded or unfolded protein conformations where a single species (which in general will be an ensemble of conformers) is populated. Some examples of calculated compaction factors are given are in Table 2.

High compaction factors are seen for the low-pH molten globule states of myoglobin (0.74) and cytochrome *c* (0.86) and a mutant of human α-lactalbumin where the cysteine residues are replaced by alanine at pH 2 (0.86). This reflects the biasing of the conformational ensembles populated in these states to collapsed conformers, which are only slightly expanded compared to the native fold. Concurrent with the compact nature of the molten globule states there is a high persistence of secondary structure. The packing of amino acid side chains is, however, disordered across the conformational ensemble. The rearrangement of side chains in the molten globule to their native orientations within the compact ensemble is recognized in some cases to be one of the slow steps in protein folding (*11, 52, 53*).

Low compaction factors (0.1) are seen for both prothymosin α and the D1−D4 fragment of the fibronectin binding protein in the absence of denaturant, two proteins that are biologically active but essentially unfolded under physiological conditions (*54*−*56*). It is, however, interesting to consider the reasons for the slightly smaller effective hydrodynamic radii of these proteins than those observed for the highly denatured states discussed earlier. NMR studies of the 130 residue D1−D4 fibronectin binding domain show that its local conformational properties generally resemble

Table 2: Hydrodynamic Radii and Compaction Factors (*C*) for a Range of Nonnative Protein Conformations

protein	R_h (Å)	*C*	no. of residues
BPTI (all Cys residues replaced by Abu, pH 4.5[a])	20.6 ± 0.3	0.35	58
horse cytochrome *c* (NaCl-induced molten globule state[b])	20.1	0.86	104
prothymosin α (pH 7.4[c])	30.7 ± 0.4	0.10	109
human α-lactalbumin (all Cys residues replaced by Ala, pH 2[d])	21.3 ± 0.1	0.86	123
hen lysozyme (reduced and methylated, pH 2[e])	30.2 ± 0.1	0.35	129
fibronectin binding protein D1-D4 (pH 7[f])	33.8 ± 0.1	0.10	130
sperm whale apomyoglobin (pH 4[b])	25.3	0.74	153

[a] Data from ref *9*; cysteine residues replaced by α-amino-*n*-butyric acid (Abu). [b] Data from ref *19*. [c] Data from ref *54*. [d] Data from ref *60*. [e] Hydrodynamic radius value from extrapolation to conditions of infinite dilution. [f] Fibronectin binding domain of *S. aureus* fibronectin binding protein variant C. Data from ref *56*.

those of many highly denatured proteins and agree well with predictions for a random coil (*55, 56*). In a number of specific regions of the fibronectin binding protein polypeptide chain, however, persistent nonlocal interactions have been identified that appear to result from the presence of clusters of hydrophobic side chains and interactions between charged glutamate and lysine side chain groups (*56*). The biasing of the conformational ensemble for fibronectin binding protein to more compact folds than would be expected for a highly denatured protein is consistent with these persistent features. There may also, however, be additional effects resulting from differences in the solvation of proteins in aqueous solution and in the presence of denaturant such as urea.

Hen lysozyme (with the disulfide bridges reduced and methylated) at pH 2.0 and BPTI (with cysteine residues replaced by α-amino-*n*-butyric acid) at pH 4.5 have higher compaction factors (0.35) than D1−D4 and prothymosin α but much smaller than those of the molten globule states discussed above. Interestingly, there is evidence which suggests that for these states persistent nonlocal interactions within the otherwise disordered polypeptide chain may also be responsible for the smaller effective hydrodynamic radii seen for these proteins than those expected for a highly denatured state. For BPTI and lysozyme the effective hydrodynamic radii are observed to increase from 20.6 to 22.7 Å and from 29.7 to 34.6 Å on a decrease of the pH to 2.0 or the addition of 8 M urea, respectively. Concurrent with the increase in hydrodynamic radius for BPTI there is a loss of a significant number of NOEs in the NMR spectra of the protein (*9*). For lysozyme the addition of urea gives a decreased persistence of hydrophobic clusters present in the denatured polypeptide chain as well as an increase in the hydrodynamic radius. The reduced persistence of the hydrophobic clusters is particularly apparent in the results from ^{15}N relaxation studies of the polypeptide backbone, which show an increased flexibility of the polypeptide chain on the addition of urea (*57*; S. B. Grimshaw, H. Schwalbe, L. J. Smith, C. M. Dobson, unpublished results). Therefore, for these nonnative states of BPTI and lysozyme there is clearly an interdependence of the local properties and the global characteristics of conformers in the populated ensemble.

CONCLUSIONS

The measured dimensions of a polypeptide chain depend on the amino acid sequence of the protein, the properties of the solvent, and the ways in which these features determine the persistence of local and long-range interactions in the protein and the degree of intermolecular association. Attempts to understand the relationship between these effects can give insight into issues such as protein folding and aggregation (*3, 29, 58*). In this work PFG NMR methods have been used to measure the effective hydrodynamic radii of systems ranging from a 16 residue peptide fragment through to a protein dimer of 494 residues. From these data, relationships have been established between the effective hydrodynamic radius under different conditions and the number of amino acids in the polypeptide chain. The data from the PFG NMR method have been compared with measurements of molecular dimensions from small-angle scattering and dynamic light scattering techniques and with predictions from theory. This analysis has shown first that the relationships between the hydrodynamic radius and the protein chain length for both native and highly denatured proteins are similar to those defined theoretically for the dimensions of collapsed and fully solvated homopolymer chains, respectively. Second, we have found that approximately constant scaling factors (ρ of 0.77 and 1.06 for native and highly denatured states) can be used to relate measurements of hydrodynamic radii from PFG NMR methods and radii of gyration from small-angle scattering techniques over a wide range of polypeptide chain lengths. Moreover, a close consistency has been observed between the results from the different techniques, including measurements for nonnative states where an effective radius averaged over the populated conformational ensemble is extracted. The insight gained from the analysis reported here has been used to interpret the measured effective dimensions of a number of nonnative states, some of which are almost as compact as the native fold and others which are highly unfolded. Combining these data with results from studies that probe the local structural and dynamical properties of the ensembles of conformers populated has revealed relationships between the global and local features of the polypeptide chains. The approach and results reported here therefore define a generally applicable method that can be used to characterize the conformational behavior of proteins under a wide range of conditions.

ACKNOWLEDGMENT

We thank Maureen Pitkeathly for synthesising the peptide fragments and Jesús Zurdo and Iñaki Guijarro for producing the SH3 domain. We also thank Jesús Zurdo, Chris Penkett, and Charles Morgan for preparing the NMR samples and recording some of the NMR spectra of SH3 domain, fibronectin binding protein (and its peptide fragments), and triosephosphate isomerase, respectively.

SUPPORTING INFORMATION AVAILABLE

Three tables listing hydrodynamic radii and radii of gyration data taken from the literature that were used in the

analysis presented in this paper. The material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Damaschun, G., Damaschun, H., Gast, K., Gerlach, D., Misselwitz, R., Welfle, H., and Zirwer, D. (1992) *Eur. Biophys. J. Biophys. Lett. 20,* 355−361.
2. Lin, M. F., and Larive, C. K. (1995) *Anal. Biochem. 229,* 214−220.
3. Ekiel, I., and Abrahamson, M. (1996) *J. Biol. Chem. 271,* 1314−1321.
4. Smith, L. J., Fiebig, K. M., Schwalbe, H., and Dobson, C. M. (1996) *Folding Des. 1,* R95−R106.
5. Dobson, C. M., Sali, A., and Karplus, M. (1998) *Angew. Chem., Int. Ed. Engl. 37,* 868−893.
6. Lattman, E. E. (1994) *Curr. Opin. Struct. Biol. 4,* 87−92.
7. Lattman, E. E., Fiebig, K. M., and Dill, K. A. (1994) *Biochemistry 33,* 6158−6166.
8. Kataoka, M., and Goto, Y. (1996) *Folding Des. 1,* R107−R114.
9. Pan, H., Barany, G., and Woodward, C. (1997) *Protein Sci. 6,* 1985−1992.
10. Dill, K. A., and Shortle, D. (1991) *Annu. Rev. Biochem. 60,* 795−825.
11. Ptitsyn, O. B. (1995) *Adv. Protein Chem. 47,* 83−229.
12. Shortle, D. (1996) *FASEB J. 10,* 27−34.
13. Glatter, O., and Kratky, O. (1982) *Small-Angle X-ray Scattering,* Academic Press, London.
14. Berne, B. J., and Pecora, R. (1976) *Dynamic Light Scattering with Applications to Chemistry, Biology and Physics,* John Wiley and Sons Inc., New York.
15. Calmettes, P., Roux, B., Durand, D., Desmadril, M., and Smith, J. C. (1993) *J. Mol. Biol. 231,* 840−848.
16. Damaschun, G., Damaschun, H., Gast, K., and Zirwer, D. (1998) *Biochemistry (Moscow) 63,* 259−275.
17. Sosnick, T. R., and Trewhella, J. (1992) *Biochemistry 31,* 8329−8335.
18. Noppert, A., Gast, K., Müller-Frohne, M., Zirwer, D., and Damaschun, G. (1996) *FEBS Lett. 380,* 179−182.
19. Gast, K., Damaschun, H., Misselwitz, R., Müller-Frohne, M., Zirwer. D., and Damaschun, G. (1994) *Eur. Biophys. J. 23,* 297−305.
20. Kataoka, M., Nishii, I., Fujisawa, T., Ueki, T., Tokunaga, F., and Goto, Y. (1995) *J. Mol. Biol. 249,* 215−228.
21. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O., and Tsuruta, H. (1995) *Science 270,* 487−488.
22. Kataoka, M., Kuwajima, K., Tokunaga, F., and Goto, Y. (1997) *Protein Sci. 6,* 422−430.
23. Gast, K., Zirwer, D., Müller-Frohne, M., and Damaschun, G. (1998) *Protein Sci. 7,* 2004−2011.
24. Kataoka, M., Hagihara, Y., Mihara, K., and Goto, Y. (1993) *J. Mol. Biol. 229,* 591−596.
25. Panick, G., Melissa, R., Winter, R., Rapp, G., Frye, K. J., and Royer, C. A. (1998) *J. Mol. Biol. 275,* 389−402.
26. Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O., and Doniach, S. (1998) *J. Mol. Biol. 276,* 225−237.
27. Jones, J. A., Wilkins, D. K., Smith, L. J., and Dobson, C. M. (1997) *J. Biomol. NMR 10,* 199−203.
28. Altieri, A. S., Hinton, D. P., and Byrd, R. A. (1995) *J. Am. Chem. Soc. 117,* 7566−7567.
29. Mansfield, S. L., Jayawickrama, D. A., Timmons, J. S., and Larive, C. K. (1998) *Biochim. Biophys. Acta 1382,* 257−265.
30. Shortle, D. (1996) *Curr. Opin. Struct. Biol. 6,* 24−30.
31. Dyson, H. J., and Wright, P. E. (1998) *Nat. Struct. Biol. 5,* 499−503.
32. Booker, G. W., Gout, I., Downing, A. K., Driscoll, P. C., Boyd, J., Waterfield, M. D., and Campbell, I. D. (1993) *Cell 73,* 813−822.
33. Ching-Li, L., and Atassi, M. Z. (1973) *Biochemistry 12,* 2690−2695.
34. Nozaki, Y. (1972) *Methods Enzymol. 26,* 43−50.
35. Soffe, N., Boyd, J., and Leonard, M. (1995) *J. Magn. Reson., Ser. A 116,* 117−121.
36. Merrill, M. R. (1993) *J. Magn. Reson., Ser. A 103,* 223−225.
37. Hanner, R. L., and Schleigh, T. (1989) *Methods Enzymol. 179,* 418−446.
38. Chen, L. L., Hodgson, K. O., and Doniach, S. (1996) *J. Mol. Biol. 261,* 658−671.
39. Burchard, A., Schmidt, M., and Stockmayer, W. H. (1980) *Macromolecules 13,* 1265−1272.
40. Parrado, J., Conejero-Lara, F., Smith, R. A. G., Marshall, J. M., Ponting, C. P., and Dobson, C. M. (1996) *Protein Sci. 5,* 693−704.
41. De Gennes, P. G. (1979) *Scaling Concepts in Polymer Physics*, Cornell University Press, Ithaca, NY.
42. Flory, P. J. (1953) *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY.
43. Dewey, T. G. (1993) *J. Chem. Phys. 98,* 2250−2257.
44. Petrescu, A. J., Receveur, V., Calmettes, P., Durand, D., and Smith, J. C. (1998) *Protein Sci. 7,* 1396−1403.
45. Receveur, V., Durand, D., Desmadril, M., and Calmettes, P. (1998) *FEBS Lett. 426,* 57−61.
46. Tanford, C. (1970) *Adv. Protein Chem. 24,* 1−95.
47. Pace, C. N. (1986) *Methods Enzymol. 131,* 266−280.
48. Tirado-Rives, J., Orozco, M., and Jorgensen, W. L. (1997) *Biochemistry 36,* 7313−7329.
49. Calmettes, P., Durand, D., Desmadril, M., Minard, P., Receveur, V., and Smith, J. C. (1994) *Biophys. Chem. 53,* 105−113.
50. Calmettes, P., Durand, D., Smith, J. C., Desmadril, M., Minard, P., and Douillard, R. (1993) *J. Phys. IV (France) 3,* 253−256.
51. Tanford, C. (1961) *Physical Chemistry of Macromolecules*, John Wiley & Sons, London.
52. Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6,* 87−103.
53. Dobson, C. M. (1994) *Curr. Biol. 4,* 636−640.
54. Gast, K., Damaschun, H., Eckert, K., Schulze-Forster, K., Maurer, H. R., Müller-Frohne, M., Zirwer, D., Czarnecki, J., and Damaschun, G. (1995) *Biochemistry 34,* 13211−13218.
55. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E., Smith, R. A. G., Dobson, C. M., and Smith, L. J. (1997) *J. Mol. Biol. 274,* 152−159.
56. Penkett, C. J., Redfield, C., Jones, J. A., Hubbard, J., Smith, R. A. G., Smith, L. J., and Dobson, C. M. (1998) *Biochemistry 37,* 17054−17067.
57. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J., and Dobson, C. M. (1997) *Biochemistry 36,* 8977−8991.
58. Flanagan, J. M., Kataoka, M., Fujisawa, T., and Engelman, D. M. (1993) *Biochemistry 32,* 10359−10370.
59. Signäs, C., Raucci, G., Jönsson, K., Lindgren, P. E., Anantharamaiah, G. M., Höök, M., and Lindberg, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86,* 699−703.
60. Redfield, C., Schulman, B. A., Milhollen, M. A., Kim, P. S., and Dobson, C. M. (1999) *Nat. Struct. Biol. 6,* 948−952.
61. Sklenář, V., Torchia, D., and Bax, A. (1987) *J. Magn. Reson. 73,* 375−379.
62. Damaschun, G., Damaschun, H., Gast, K., Gernat, C., and Zirwer, D. (1991) *Biochim. Biophys. Acta 1078,* 289−295.
63. Damaschun, G., Damaschun, H., Gast, K., Misselwitz, R., Müller, J. J., Pfeil, W., and Zirwer, D. (1993) *Biochemistry 32,* 7739−7746.
64. Flanagan, J. M., Kataoka, M., Shortle, D., and Engelman, D. M. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89, 748−752.
65. Konno, T., Kataoka, M., Kamatari, Y., Kanaori, K., Nosaka, A., and Akasaka, K. (1995) *J. Mol. Biol. 251,* 95−103.
66. Semisotnov, G. V., Kihara, H., Kotova, N. V., Kimura, K., Amemiya, Y., Wakabayashi, K., Serdyuk, I. N., Timchenko, A. A., Chiba, K., Nikaido, K., Ikura, T., and Kuwajima, K. (1996) *J. Mol. Biol. 262,* 559−574.
67. Zhou, J. M., Fan, Y. X., Kihara, H., Kimura, K., and Amemiya, Y. (1997) *FEBS Lett. 415,* 183−185.
68. Kamatari, Y. O., Ohji, S., Konno, T., Seki, Y., Soda, K., Kataoka, M., and Akasaka, K. (1999) *Protein Sci. 8,* 873−882.

BI991765Q