doi:10.1006/jmbi.2001.5328 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 316, 101–112

JMB

Distribution of Molecular Size within an Unfolded State Ensemble using Small-angle X-ray Scattering and Pulse Field Gradient NMR Techniques

Wing-Yiu Choy[1], Frans A. A. Mulder[1], Karin A. Crowhurst[2] D. R. Muhandiram[1], Ian S. Millett[3], Sebastian Doniach[4] Julie D. Forman-Kay[2] and Lewis E. Kay[1]*

[1]*Protein Engineering Network Centers of Excellence and Departments of Medical Genetics and Microbiology Biochemistry, and Chemistry University of Toronto, Toronto Ontario, Canada M5S 1A8*

[2]*Structural Biology and Biochemistry, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario Canada, M5G 1X8, and Department of Biochemistry University of Toronto, Toronto Ontario, Canada M5S 1A8*

[3]*Department of Chemistry Stanford University, Stanford CA 94305, USA*

[4]*Department of Applied Physics, Stanford University Stanford, CA 92343, USA*

**Corresponding author*

The size distribution of molecules within an unfolded state of the N-terminal SH3 domain of drk (drkN SH3) has been studied by small-angle X-ray scattering (SAXS) and pulsed-field-gradient NMR (PFG-NMR) methods. An empirical model to describe this distribution in the unfolded state ensemble has been proposed based on (i) the ensemble-averaged radius of gyration and hydrodynamic radius derived from the SAXS and PFG-NMR data, respectively, and (ii) a histogram of the size distribution of structures obtained from preliminary analyses of structural parameters recorded on the unfolded state. Results show that this unfolded state, U_{exch}, which exists in equilibrium with the folded state, F_{exch}, under non-denaturing conditions, is relatively compact, with the average size of conformers within the unfolded state ensemble only 30-40 % larger than the folded state structure. In addition, the model predicts a significant overlap in the size range of structures comprising the U_{exch} state with those in a denatured state obtained by addition of 2 M guanidinium chloride.

Keywords: unfolded state; SH3 domain; PFG-NMR; SAXS; ensemble

Introduction

Contrary to a simplistic structure-function paradigm, an increasing number of proteins that are unstructured under physiological conditions yet biologically active are being described.[1] For example, the p21 Cdk inhibitors and the C-terminal activation domain of c-Fos are functionally active but intrinsically disordered in the absence of binding to target proteins.[2,3] It has been suggested that conformational disorder may play an important role in the diversity of interactions of these and other such proteins. In addition to the experimentally demonstrated examples, a significant fraction of protein sequences in various genomes are predicted to code for disordered or partially disordered ''structures''.[4]

To better understand the potential biological role of disorder, structural information on unfolded and partially unfolded states of proteins under physiological conditions is crucial. Unlike a protein in its folded state, the unfolded state is an ensemble of rapidly interconverting conformers. A complete characterization requires not only descriptions of average local and global properties but details on the various conformers that contribute to the disordered state ensemble and their relative populations.[5] Even though unfolded states have much less persistent structure than folded states, recently developed NMR and other spectro-

Abbreviations used: NOE, nuclear Overhauser effect; SAXS, small-angle X-ray scattering; PFG, pulsed-field-gradient.

E-mail address of the corresponding author: Kay@pound.med.utoronto.ca

0022-2836/02/010101–12 $35.00/0

scopic techniques have allowed characterization of the residual structure present. Information on backbone conformational preferences can be derived from NMR parameters including scalar *J*-coupling constants,[6–8] sequential HN-HN and Hα-HN NOEs[9] as well as chemical shifts.[10,11] Long range HN-HN NOEs can also be observed in the unfolded state ensemble using extensive deuteration to reduce relaxation[12] and paramagnetic relaxation enhancement[13,14] and residual dipolar couplings[15] can be used to probe structural features of partially unfolded or unfolded molecules. Details on specific interactions in populated conformers of the unfolded state ensemble can thus be obtained.

A full description of the unfolded state, however, requires information about the ensemble distribution of molecular sizes, shapes and the extent of compactness. Small-angle X-ray scattering (SAXS) is one of the few techniques that provides a direct measure of these properties.[16] Recent advances in X-ray sources and instrumentation make SAXS a powerful tool for studying the conformations and interactions of biological macromolecules with a number of studies focussing on disordered states. In particular, SAXS results have shown that thermally and chemically denatured states of ribonuclease A are more compact than a random coil of the same length.[17] A time-resolved approach to SAXS has also been applied to study the changes in compactness of lysozyme during different stages of the folding process.[18,19]

Pulsed-field-gradient NMR (PFG-NMR) is another useful technique to study molecular size.[20,21] While SAXS yields the radius of gyration (R_g), PFG-NMR provides a measure of the translational diffusion coefficient, which can be used to determine the hydrodynamic radius (R_h). This method has been applied to the study of folded and unfolded states of a number of proteins.[21–23]

Here, we have used SAXS and PFG-NMR to measure the radius of gyration and hydrodynamic radius of the unfolded state ensemble of the isolated N-terminal SH3 domain of the *Drosophilia* signal transduction protein drk (drkN SH3) under non-denaturing conditions. The *Drosophilia* protein drk functions to couple activated receptor tyrosine kinases to Ras signaling *via* binding of the guanine nucleotide exchange factor Sos to this SH3 domain.[24] SH3 domains, in general, mediate protein recognition in signal transduction and cellular localization by binding to proline-rich targets.[25] Most isolated SH3 domains are stably folded; however, the drkN SH3 domain is unstable in the absence of its binding target Sos and exists in equilibrium between a folded (F_{exch}) and an unfolded form (U_{exch}) in aqueous buffer and near neutral pH. The protein can be stabilized to the folded state (F_s) by addition of 0.4 M sodium sulfate or denatured (U_{gdn}) by adding 2.0 M guanidinium chloride (GdnCl). Extensive structural studies of these various states have been performed using NMR and other spectroscopic techniques.[5,12,26–35] To enable a more complete understanding of the U_{exch} state, which coexists with the folded state under non-denaturing conditions, we have applied SAXS and PFG-NMR techniques to determine the molecular size distribution in the U_{exch} ensemble. In particular, we have taken advantage of the difference in modes of ensemble averaging of the radius of gyration and hydrodynamic radius, along with additional information provided by preliminary structures of the unfolded ensemble,[5] to describe the size distribution of conformers in this state.

Results and Discussion

SAXS measurements were performed on three different samples of the drkN SH3 domain, including (i) the folded state, F_s, stabilized by the addition of 0.4 M sodium sulfate, (ii) an equilibrium mixture of folded (F_{exch}) and unfolded (U_{exch}) states and (iii) the chemical (2 M GdnCl) denatured state (U_{gdn}). Figure 1 shows Guinier plots for each of the different states of the SH3 domain studied. The effective radii of gyration (R_g) obtained by Guinier analyses of the scattering profiles are 11.9(±0.5), 14.1(±0.5) and 21.9(±0.5) Å for the F_s, F_{exch}/U_{exch} and U_{gdn} states, respectively. The slightly larger R_g measured for the folded protein (F_s state) relative to $R_g = 10.4$ Å calculated from the NMR-derived structure is due to the contribution from hydration to scattering. SAXS data from the F_{exch}/U_{exch} system was analyzed according to the relation:

$$R_g^2 = p_F R_{g,F}^2 + (1 - p_F) R_{g,U}^2 \quad (1)$$

where $R_{g,F}$ and $R_{g,U}$ are the radii of gyration of the folded and unfolded exchanging states, respectively, and p_F is the fractional population of the F_{exch} state.[36] In equation (1) it is understood that the unfolded state is comprised of an ensemble of conformers so that:

$$R_{g,U}^2 = \sum_i^N P(R_{g,i}) R_{g,i}^2 \quad (2)$$

with $P(R_{g,i})$ the probability of finding a molecule in the unfolded ensemble with radius of gyration equal to $R_{g,i}$. A value of $p_F = 0.58(\pm 0.04)$ has been determined from ^{1}H-^{15}N HSQC correlation spectra recorded on the F_{exch}/U_{exch} mixture. Using this value along with $R_g = 11.9(\pm 0.5)$ Å obtained for the F_s folded state gives $R_{g,U} = 16.7(\pm 1.4)$ Å (equation (1)). In this regard it is noteworthy that NMR studies have established that the F_s and F_{exch} folded states have very similar structures[28] and hence their hydrodynamic properties are expected to be the same. The R_g value obtained for the unfolded ensemble is only 40 % larger than what is measured for the F_s state and significantly smaller than 21.9(±0.5) Å observed for the GdnCl-denatured drkN SH3 domain. This result is consist-

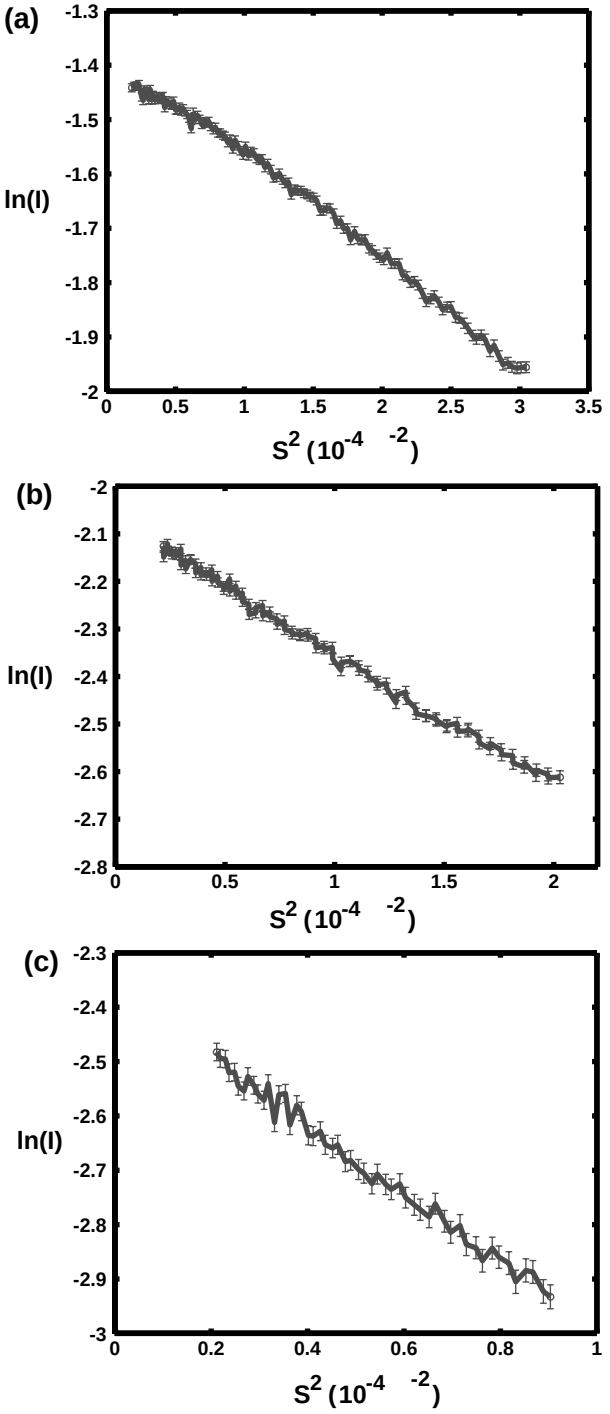

Figure 1. Guinier plots of scattering data recorded on the F_s state stabilized by addition of 0.4 M Na_2SO_4 (a), the F_{exch}/U_{exch} equilibrium state in 50 mM phosphate buffer solution at pH 6.0 (b) and the denatured state in 2 M GdnCl, U_{gdn} (c). The slope of each line is proportional to $-(4\pi^2 R_g^2/3)$. Guinier R_g values of the F_s, F_{exch}/U_{exch} and U_{gdn} states are 11.9(±0.5), 14.6(±0.5) and 21.9(±0.5) Å, respectively. The R_g value of the F_{exch}/U_{exch} equilibrium state derived from the first two minutes of data is 14.1(±0.5) Å

ent with long range amide distances observed in NOESY experiments recorded on the U_{exch} state that were not observed in similar experiments on the U_{gdn} state[12] and argues that the U_{exch} ensemble is comprised of a large fraction of compact structures (see below).

Equation (2) indicates that any interpretation of $R_{g,U}$ in terms of the properties of the ensemble is predicated on knowing the form of $P(R_g)$. To our knowledge, however, there is no standard model that describes the distribution of R_g values for conformers in an unfolded state ensemble. With this in mind we have used two different functional forms for $P(R_g)$. The first relation is based on an empirical model proposed by Flory and Fisk to describe the statistical distribution of the radii of gyration in an ensemble of linear chain molecules neglecting excluded volume effects:[37]

$$P(R_g) = const \times R_g^6 \exp\left[\frac{-7R_g^2}{2\langle R_g^2\rangle_o}\right] \tag{3}$$

where $\langle R_g^2\rangle$ is the mean-square radius of gyration for the ensemble. Based on equation (3) the distribution can be written as:

$$p(R_g) = \frac{(R_g - R_g^{native})^{\mu} \exp\left[\dfrac{-(R_g - R_g^{native})^2}{\sigma}\right]}{\sum_{R_g=R_g^{native}}^{R_g=R_g^{extended}} (R_g - R_g^{native})^{\mu} \exp\left[\dfrac{-(R_g - R_g^{native})^2}{\sigma}\right]} \tag{4a}$$

where R_g^{native} and $R_g^{extended}$ are the radii of gyration for the folded and the completely extended drkN SH3 domain structures, respectively, and σ and μ are the fitting parameters. The second model that we have examined is completely empirical and is given by the expression:

$$p(R_g) = \frac{(R_g - R_g^{native})^2 \exp\left[\dfrac{-2(R_g - R_g^{native})}{\sigma}\right]}{\sum_{R_g=R_g^{native}}^{R_g=R_g^{extended}} (R_g - R_g^{native})^2 \exp\left[\dfrac{-2(R_g - R_g^{native})}{\sigma}\right]} \tag{4b}$$

where σ is the single fitting parameter. The functional form of $P(R_g)$ is quite sensitive to σ and not surprisingly, therefore, $P(R_g)$ can fit many different distributions including, as we show below, those comprised of structures from unfolding trajectories of the drkN SH3 domain. In what follows we are not concerned about the interpretation of these parameters in terms of properties of the distribution; only whether meaningful $P(R_g)$ distributions can be obtained in the first place. In both models a value of $R_g^{extended} = 40$ Å was chosen based on calculations of R_g for unfolded molecules

generated during unfolding trajectories (see below).

In addition to the SAXS-derived data on R_g, further information on the size-distribution properties of the F_{exch}/U_{exch} system can be obtained from NMR studies of translational diffusion rates. Since molecules in the folded and unfolded states exchange slowly on the chemical shift time scale, diffusion coefficients for both states can be obtained simultaneously from measurements on a single sample. The ratio of diffusion constants obtained in this manner is therefore independent of sample viscosity, simplifying analysis, as described below. Figure 2 shows the decay of magnetization as a function of PFG strength for signal derived from T12 in the F_{exch} (circle) and U_{exch} (square) states. In total, decay curves for 42 residues from the folded state and 19 from the unfolded ensemble could be measured using two-dimensional PFG experiments described in Materials and Methods. Average diffusion coefficients of $7.10(\pm0.04) \times 10^{-7}$ cm^2 s^{-1} and $5.40(\pm0.03) \times 10^{-7}$ cm^2 s^{-1} have been obtained for the F_{exch} and U_{exch} states. The translational diffusion coefficient of a molecule, D_t, is related to its hydrodynamic radius, R_h defined as the radius of a sphere with the same value of D_t, according to the Stokes-Einstein equation:

$$D_t = \frac{kT}{6\pi\eta R_h} \tag{5}$$

where k is Boltzmann's constant, T is the absolute temperature and η is the solvent viscosity. Thus:

$$\frac{D_{t,F}}{D_{t,U}} = 1.30 \pm 0.01 = \frac{R_{h,F}^{-1}}{R_{h,U}^{-1}} \tag{6}$$

where the subscripts F and U denote folded and unfolded states, respectively. The value of $R_{h,U}^{-1}$ is averaged over all conformations in the unfolded state according to:

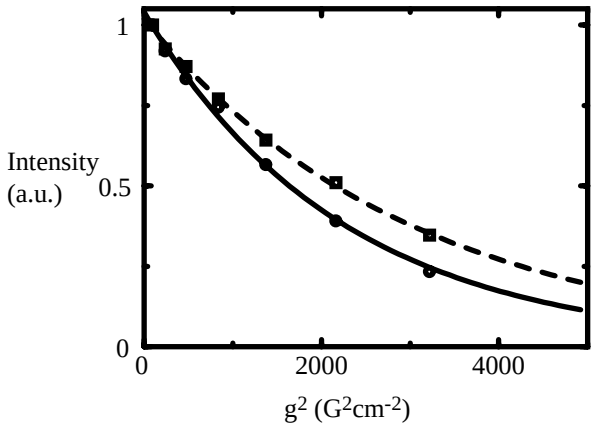

Figure 2. Diffusion attenuation data for Thr12 in the F_{exch} (○) and U_{exch} (□) states. Similar curves were obtained for 42 residues in the F_{exch} state and 19 residues in the U_{exch} state.

$$R_{h,U}^{-1} = \sum_i^N P(R_{h,i}) R_{h,i}^{-1} \tag{7}$$

where $P(R_{h,i})$ is the probability of an unfolded conformation having a hydrodynamic radius of $R_{h,i}$. The value of $R_{h,U}^{-1}$ can be obtained from equation (6) so long as $R_{h,F}^{-1}$ is known. Using the program HYDROPRO[38] along with the NMR-derived structure of the folded drkN SH3 domain and assuming approximately one shell of bound water (see Materials and Methods) a value of $R_{h,F} = 15.6$ Å is calculated and therefore $R_{h,U} = 20.3(\pm0.2)$ Å.

Equation (2) shows that $R_{g,U}^2$, measured by the SAXS experiment, is biased by extended conformers in the unfolded ensemble. In contrast, inspection of equations (6) and (7) indicates that the parameter measured in the NMR experiment, the diffusion coefficient and hence $R_{h,U}^{-1}$, is weighted more heavily by compact structures. In what follows the difference in the way in which R_g^2 and D_t average will be exploited to estimate the size distribution of conformations in the U_{exch} ensemble.

Before such an estimate can be obtained, however, it is necessary to develop a relation between R_g and R_h so that the SAXS and NMR data can be combined. In the case of a globular folded protein, the characteristic ratio, R_g/R_h, equals $\sqrt{3/5}$ (≈0.775), while for an ideal Gaussian chain and an excluded volume chain this ratio is 1.24 and 1.56, respectively.[39] An R_g/R_h ratio of 0.75 is obtained for the F_s state of the drkN SH3 domain, with R_g measured using SAXS and R_h calculated using HYDROPRO[38] and the NMR structure. This ratio is very close to the expected value. In contrast, the experimentally determined ratio of R_g/R_h for the U_{exch} state is 0.82, far removed from the values expected for either a Gaussian or excluded volume chain. Indeed, a ratio of 0.82 argues strongly that the U_{exch} state is comprised of a significant population of compact structures.

An empirical correlation between R_g and R_h has been obtained from a large set of conformers which provide a reasonable sampling of the conformational space available to the U_{exch} state of the drkN SH3 domain. Conformers have been generated by creating multiple unfolding trajectories using the program TRADES (Trajectory Directed Ensemble Sampling).[40] In this approach, C^α atoms of the F_s state structure were displaced in an iterative manner in a series of 100 steps; after each step the remaining backbone atoms are added, followed by the side-chains using a rotamer library. A total of 600 of these unfolding trajectories were generated to produce 60,000 conformers. In each trajectory structures range from those close to the folded state (i.e. with R_g near that of the F_s state) to those that are very extended, providing a sampling of both native-like and unfolded proteins. Subsequently R_g and R_h values have been calculated for all structures using HYDROPRO,[38] including approximately one water layer of hydration, with R_h and R_g/R_h illustrated as a function R_g in

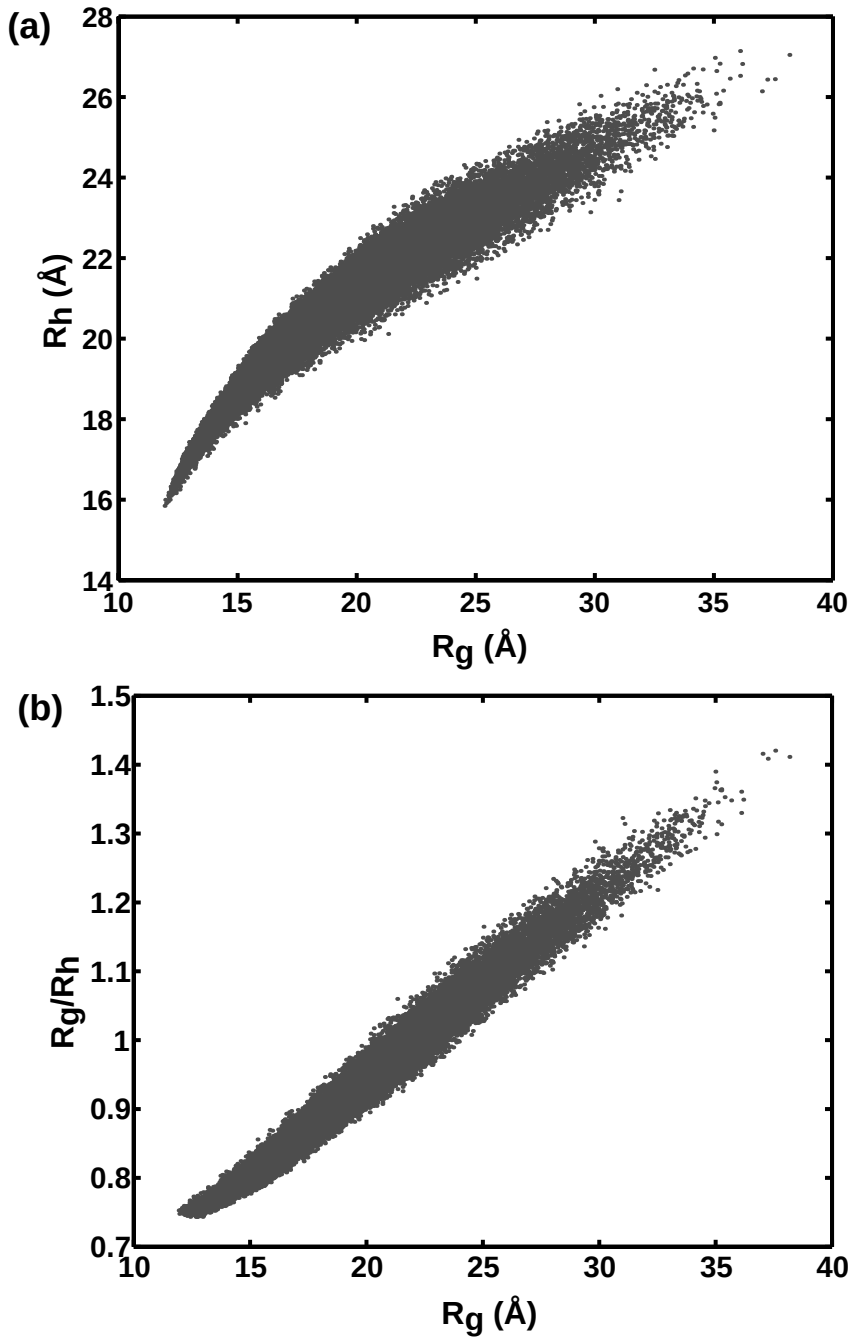

Figure 3. Correlation between R_g and R_h (a) and between (R_g/R_h) and R_g (b) for the 60,000 structures generated in an unfolding trajectory using the program TRADES.[40] R_g and R_h values for all conformers were calculated using the program HYDROPRO.[38]

Figure 3(a) and (b), respectively. The range of characteristic ratios extends from 0.74 to 1.42 and interestingly a linear correlation between R_g/R_h and R_g is observed:

$$R_g/R_h = 0.0273 \times R_g + 0.396 \quad \text{(8a)}$$

Since the ratio R_g/R_h defined by equation (8a) is critical to the analysis that follows, we have used an alternative method to obtain an unfolding trajectory from which this ratio can be obtained. This ensures that the functional form of equation (8a) above is not an artifact of the method that has been used to calculate the ensemble of structures in the first place. Starting from a 10 ns molecular dynamics trajectory of the drkN SH3 domain at 498 K, 9600 structures have been extracted, with an additional 600 structures obtained from a 3.6 ns trajectory at 598 K (see Materials and Methods). From this distribution of structures the ratio R_g/R_h was calculated and, as with the ensemble of structures obtained from TRADES,[40] a linear correlation between R_g/R_h and R_g was obtained, as defined by:

$$R_g/R_h = 0.0274 \times R_g + 0.423 \quad \text{(8b)}$$

In what follows we have used equation (8a); essen-

tially identical results are obtained if the correlation given by equation (8b) is substituted. We have also determined the relation between R_g/R_h and R_g for a number of other proteins using structures generated from unfolding trajectories obtained with TRADES. Although linear relationships were obtained in all cases, they do vary. For example, assuming a relation of the form $R_g/R_h = a \times R_g + b$, we find $a = 0.023$, $b = 0.42$ for ubiquitin (76 residues, 1UBQ[41]), $a = 0.023$, $b = 0.39$ for barnase (110 residues, 1RNB[42]), $a = 0.034$, $b = 0.38$ for crambin (46 residues, 1CRN[43]) and $a = 0.015$, $b = 0.53$ for hen egg white lysozyme (129 residues, 2LZT[44]).

Using the correlation described in equation (8a), along with the measured values for $R_{h,U}$ and $R^2_{g,U}$ it is possible to determine the distribution of R_g values in an unfolded ensemble (equations (4a) or (4b)) by minimizing:

$$\chi^2 = (R^{calc}_{g,U} - R^{expt}_{g,U})^2 + (R^{calc}_{h,U} - R^{expt}_{h,U})^2 \qquad (9)$$

with $R^{calc}_{g,U}$ and $R^{calc}_{h,U}$ given by equations (2) and (7), respectively. The size distribution function given by equation (4a) depends on two fitting parameters and it is clear, therefore, that at least two measurables are required to properly define $P(R_g)$ in this case. In contrast, the model described by equation (4b) is a function of only a single fitting parameter, and in principle, it is possible to obtain a fitted $P(R_g)$ distribution from only a single observable. However, as illustrated in Figure 4, a combination of both R_h and R_g values greatly decreases the uncertainty in the distribution that is obtained. For example, consider the case where $R^{expt}_{g,U} = 15.0(\pm 0.5)$ Å is measured and used to obtain the population distribution function. Figure 4(a) shows the range of $P(R_g)$ curves that are consistent with this single experimental value. Alternatively, if $R^{expt}_{h,U} = 19.1(\pm 0.5)$ Å is the only experimental parameter, a range of distributions shown in Figure 4(b) is obtained. Figure 4(c) illustrates the set of $P(R_g)$ functions that are obtained when both $R^{exp}_{g,U}{}^{t} = 15.0(\pm 0.5)$ Å and $R^{expt}_{h,U} = 19.1$ (± 0.5) Å are fitted simultaneously and it is clear that the distribution is much better defined.

Figure 5 illustrates the distribution that is obtained when ''ensemble-averaged'' values of R_g (''$R^{expt}_{g,U}$'' = 18.5 Å) and R_h (''$R^{expt}_{h,U}$'' = 20.0 Å), calculated from the 60,000 structures generated from the TRADES unfolding trajectory, are fitted using equation (9). The actual distribution (histogram) is shown, and $P(R_g)$, obtained from minimization of χ^2 using either equation (4a) (green, broken line) or equation (4b) (red, continuous line), is illustrated for comparison. It is clear that despite the fact that there are only two fitting parameters in equation (4a) and one in equation (4b), both $P(R_g)$ functional forms capture the essential elements of the distribution. As a further test of the model, 110 subsets of conformers from the 60,000 structures were generated comprised of approximately 300 structures each, ensemble-averaged ''$R^{expt}_{g,U}$'' and ''$R^{expt}_{h,U}$'' calculated for each group as before, and $P(R_g)$ distribution functions obtained. Figure 6 illustrates histograms of R_g along with the calculated $P(R_g)$ curves (green, broken line equation (4a); red, continuous line equation (4b)) for representative sets of data.

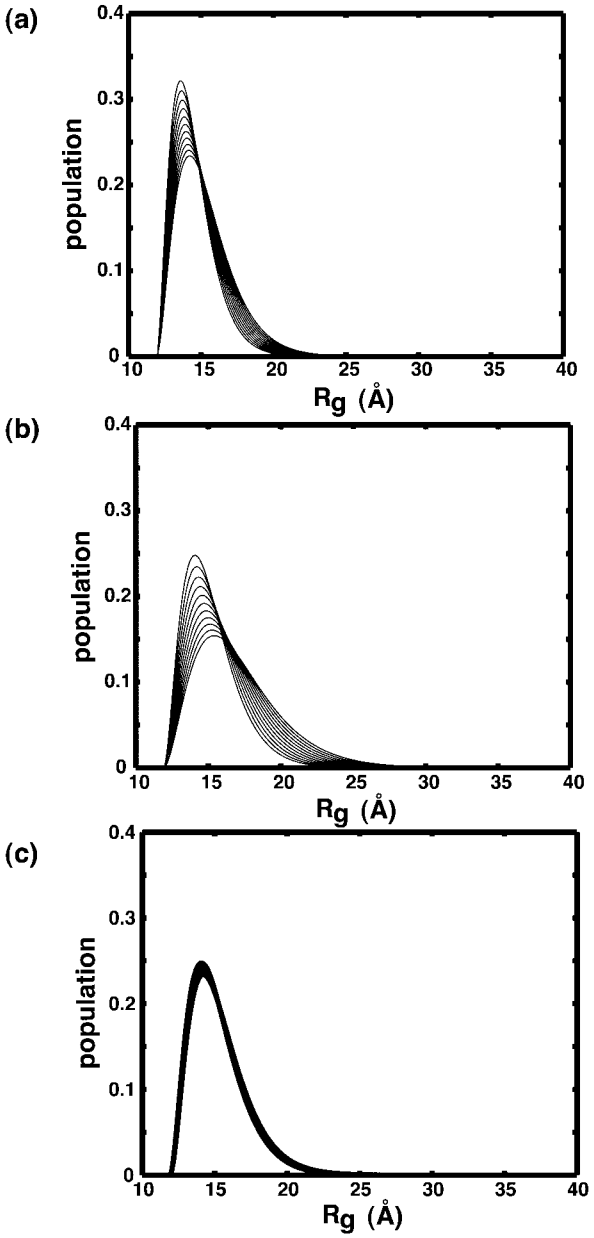

Figure 4. R_g distributions obtained from fits using equation (4b) with $R_g = 15.0(\pm 0.5)$ Å used as input (a), with $R_h = 19.1(\pm 0.5)$ Å used as input (b) and when both $R_g = 15.0(\pm 0.5)$ Å and $R_h = 19.1(\pm 0.5)$ Å are fitted simultaneously using equation (9) (c). Note the improved description of $P(R_g)$ when both NMR and SAXS data are included in the fits.

In Figure 6(a) a relatively narrow distribution of structures is present in the ensemble, with a significant population of structures close to

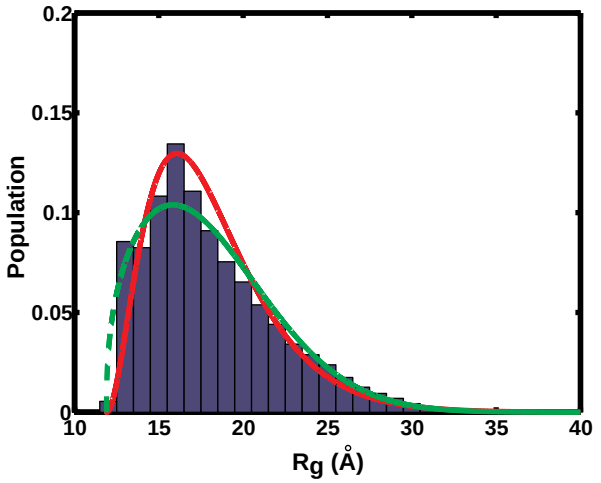

Figure 5. Histogram of the distribution of R_g values in the unfolded ensemble of the drkN SH3 domain calculated from 60,000 structures obtained from an unfolding trajectory using the program TRADES.[40] The green dotted line is the R_g distribution predicted from fits of equation (4a) to the ensemble-averaged R_g and R_h values, while the red, continuous line is derived in the same manner using a model based on equation (4b). R_g^{native} and $R_g^{extended}$ are set to 11.9 and 40.0 Å, respectively. Ensemble-averaged R_g and R_h values of 18.5 and 20.0 Å were used in the fitting.

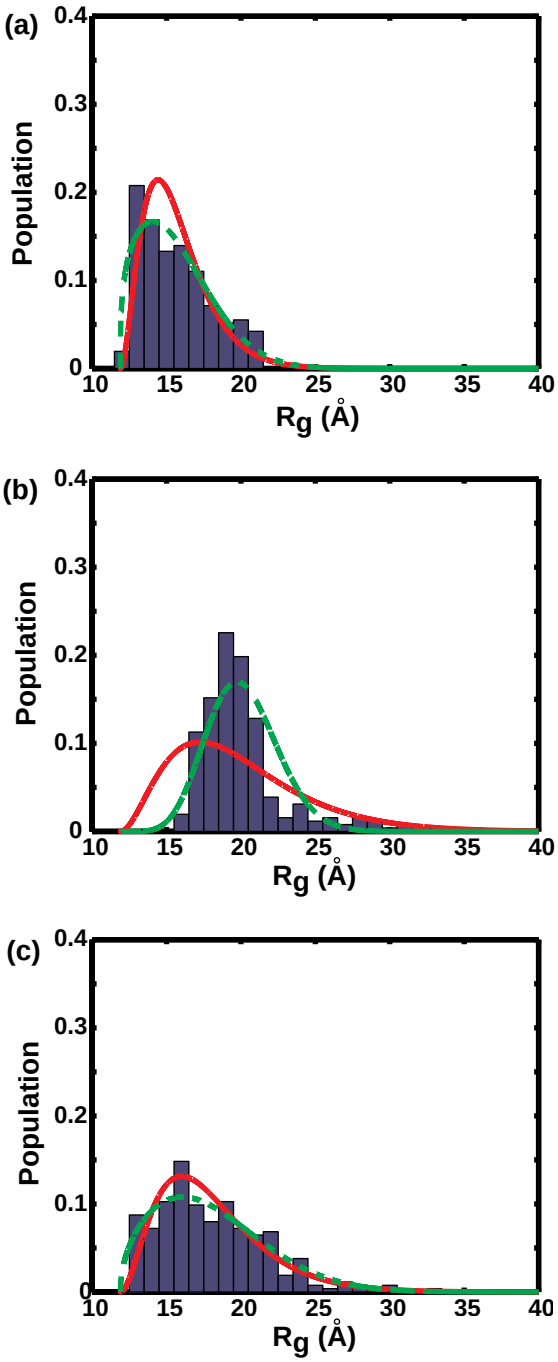

Figure 6. Histograms from small subsets of structures (approximately 300/ensemble) calculated from the SH3 unfolding trajectory using TRADES showing distributions of R_g values. The ensembles were generated to increase in distribution width from (a) to (c). Ensemble-averaged R_g and R_h values of (15.9 Å, 18.8 Å) (a), (20.3 Å, 21.2 Å) (b), (18.5 Å, 20.0 Å) (c) are indicated, with the green dotted lines (red continuous lines) showing R_g distributions predicted using equation (4a) (equation (4b)).

$R_g^{native} = 11.9$ Å, the radius of gyration of the native state structure. Note that both models described in equation (4) require an input value for R_g^{native} and we have used 11.9 Å in each case. In general this value is available from experiments on the native protein or by calculation,[38] assuming that the structure of the protein in its native form is known. Reasonable agreement between the actual distribution described by the histogram in the Figure and $P(R_g)$ fitted on the basis of "$R_{g,U}^{expt}$" and "$R_{h,U}^{expt}$" is obtained for both models. In contrast, the R_g distribution profiles predicted from the data are not nearly as good when there are few native structures present in the ensemble (Figure 6(b)), and it is clear that equation (4b) does a worse job than equation (4a). In the case where the ensemble is comprised of a wide distribution of structures (Figure 6(c)) both models give reasonably good fits to the data, with the width of the distribution reproduced quite well in both cases. In all cases considered "$R_{g,U}^{expt}$" and "$R_{h,U}^{expt}$" values are fit to within 2.0 %. The ensemble averaged values of $R_{g,U}$ and $R_{h,U}$ used in each of the fits illustrated in Figure 6 are listed in the legend to the Figure.

Figure 7(a) and (b) shows $P(R_g)$ profiles for the U_{exch} ensemble calculated as described above with equation (4a) and (4b), respectively, using the average values $R_{g,U}^{expt} = 16.7$ Å and $R_{h,U}^{expt} = 20.3$ Å. In addition, the $P(R_g)$ distribution has also been obtained for the U_{gdn} state based on $R_{g,U}^{expt} = 21.9$ Å determined by SAXS together with a rough estimate of $R_{h,U}$ (22.6 Å), calculated using an empirical equation developed by Wilkins *et al.*[21] which relates R_h to the number of residues, N, in an unfolded polypeptide chain ($R_h = (2.2 \pm$

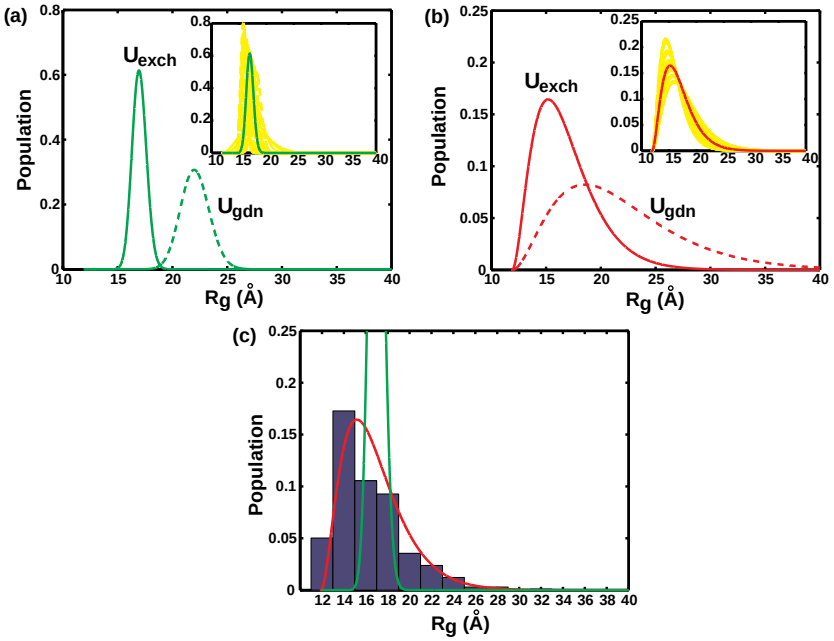

Figure 7. R_g distributions in U_{exch} and U_{gdn} state ensembles obtained using $R_{g,U}^{expt}$= 16.7 Å and $R_{h,U}^{expt}$= 20.3 Å (derived on the basis of SAXS and NMR data) and models based on equation (4a) (a) and equation (4b) (b). Insets show the effects of experimental error (±0.2 Å for $R_{h,U}^{expt}$ and ±1.5 Å for $R_{g,U}^{expt}$) on the distributions for the U_{exch} state (yellow). (c) A comparison between the profiles predicted from the model of equation (4a) (green) and equation (4b) (red) with the distribution of structures (2400) generated from experimental constraints, as described previously.[5]

1.1)$N^{0.57 \pm 0.02}$). Not unexpectedly a wider distribution is noted for the U_{gdn} state relative to the U_{exch} ensemble for both models and the maximum of the U_{gdn} distribution is shifted by approximately 5-6 Å to larger R_g values. It is very clear that different distributions can satisfy the experimental R_g and R_h restraints. The distributions based on modeling the U_{exch} or U_{gdn} states as a set of Gaussian chains without excluded volume (equation (4a)) are surprisingly narrow with little overlap in the size of unfolded molecules in either of the U_{exch} or U_{gdn} states. In contrast, the model described by equation (4b) predicts much larger $P(R_g)$ distributions with significant overlap of molecular sizes in the two unfolded states. Of interest, the $P(R_g)$ profile in this case is very similar to the distribution generated from the unfolding trajectory of the SH3 domain using the TRADES program (see above) (Figure 5).

As described above, the U_{exch} ensemble has been studied extensively by a variety of spectroscopic techniques.[5,12,26–35] On the basis of NMR-derived distance restraints, scalar couplings, chemical shifts and PFG diffusion measurements as well as fluorescence experiments a preliminary distribution of structures which describes the unfolded ensemble under non-denaturing conditions has been obtained.[5] Although the details of the distribution await further refinement, we have used these results to help distinguish between the two very different models that are consistent with our experimental measurements. Figure 7(c) superimposes the $P(R_g)$ distribution functions calculated for the U_{exch} state using the models described in equation (4a) (green) and (4b) (red) onto a histogram of R_g values obtained from the combination of five ensembles (total of 2400 U_{exch} structures) generated from the experimental constraints as described previously.[5] It is clear that the fitted distribution function based on equation (4b) is much more consistent with the ensemble of calculated structures, both in terms of the width and the relative population of structures. Finally, in order to estimate how experimental error affects the resulting distribution functions, we have varied $R_{g,U}^{expt}$ by ±1.5 Å and $R_{h,U}^{expt}$ by ±0.2 Å about their mean values with the range of profiles obtained illustrated in the insets to Figure 7(a) and (b) (yellow). Notably for both models this range of uncertainty does not change the profiles in a significant way. Although it is premature to suggest that the model described by equation (4b) can be used to accurately describe size distributions for all unfolded proteins the methodologies described here can certainly be used to assess its utility on a case by case basis.

Conclusions

SAXS and NMR measurements have been employed to obtain a picture of the distribution of molecular sizes in an ensemble of unfolded protein states using crude models for the size distribution which depend on only one or two fitting parameters. The approach developed exploits the fact that ensemble averaging affects the experimental observables in SAXS and PFG-NMR (R_g^2, SAXS; D_t, NMR) in fundamentally different ways, so that the combination of results from the two methods is expected to be a significantly better monitor of molecular size properties than either method alone. Nevertheless taken together the data are still insufficient to uniquely define a size distribution function describing the U_{exch} unfolded state ensemble of the drkN SH3 domain. Indeed, best fits of the experimental data to either of the two models described here can reproduce experimental par-

ameters to within 4 %, yet very different distribution functions are generated. Both of the profiles are consistent with previous NMR studies of the U_{exch} state that suggested that molecules are relatively compact with significant amounts of non-native secondary and tertiary interactions, with many of these contacts disappearing upon addition of guanidinium chloride.[12] However, comparison of the different models obtained exclusively on the basis of the PFG-NMR and SAXS data with the distribution of structures in U_{exch} ensembles calculated primarily from NOE, chemical shift, scalar coupling and diffusion constant restraints[5] indicates that one of the profiles is much more likely (model described by equation (4b)). Notably this profile predicts a wide distribution of structures, similar to what is observed when an ensemble comprised of 60,000 conformers is generated by successively unfolding the SH3 domain, as well as significant overlap in the family of structures of the U_{exch} and U_{gdn} states.

Materials and Methods

Sample preparation

The protein expression and purification of the drkN SH3 domain has been described.[12] The sample of the F_{exch}/U_{exch} equilibrium mixture contained 7 mg/ml protein in 50 mM phosphate buffer solution at pH 6.0. Samples of the fully stabilized folded state (F_s) (11.6 mg/ml) and the guanidinium chloride denatured state (U_{gdn}) (10 mg/ml) were prepared by adding 0.4 M Na_2SO_4 and 2 M GdnCl to the buffer solution, respectively. The pH values were then adjusted to 6.0.

NMR spectroscopy

All NMR experiments were carried out on a Varian Inova spectrometer, operating at 500 MHz ^{1}H frequency, at a probe temperature of 5 °C. In order to reduce artifacts caused by non-linearity of the pulsed field gradients, sample volumes were minimized by using Shigemi tubes for all NMR measurements. Gradient field strengths were obtained following the procedures outlined by Price[45] *via* (i) the shape of a 1D image of the NMR sample and (ii) using a standard sample of nearly neat ^2H_2O, with a value of $D = 1.9 \times 10^{-5}$ cm^2 s^{-1} for the residual HO^{2}H line at 25 °C.[45] Both methods were in excellent agreement and a value of 1.90×10^{-3} G cm^{-1} DAC^{-1} was obtained for our experimental setup.

Translational diffusion constants of both the U_{exch} and F_{exch} states of the drkN SH3 domain were measured by a two-dimensional analog of the stimulated echo LED experiment.[46] The extra resolution afforded by 2D spectroscopy was necessary since in 1D spectra resonances from the folded and unfolded drkN SH3 states overlap. In addition, the use of 2D heteronuclear spectroscopy ensured that resonances from any small molecule contaminants in the sample (e.g. traces of glycerol) would be eliminated and hence not contribute to diffusion attenuation. The experimental pulse scheme is shown in Figure 8, and is similar to a previously published 2D stimulated echo experiment.[47] Briefly, diffusion attenu-

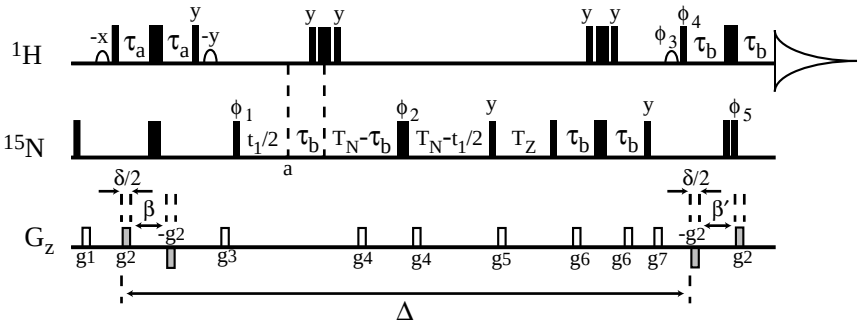

Figure 8. ^{1}H-^{15}N correlation experiment for the measurement of translational diffusion constants in a mixture of isotopically ^{15}N-enriched molecules. The narrow (wide) bars represent 90° (180°) pulses with phase x, unless indicated otherwise. Field strengths are 25 and 6.8 kHz on the ^{1}H and ^{15}N channels, respectively. Shaped pulses have a duration of 1.5 ms (rectangular shape) to selectively rotate the solvent water magnetization. ^{1}H and ^{15}N carrier frequencies were centered at the water resonance (4.96 ppm) and in the middle of the amide spectrum (119 ppm), respectively. In the case of applications involving ^{13}C-labeled proteins a carbon inversion pulse is applied at point *a* to refocus evolution due to ^{15}N-^{13}C scalar couplings. The delays used are: $T_N = 45$ ms, $\tau_a = 2.3$ ms, $\tau_b = 2.69$ ms, $\Delta = 205$ ms, $\delta/2 = 1.0$ ms, $\beta = 2.374$ ms, $\beta' = 2.762$ ms. T_Z was calculated from the value of Δ, and was approximately 110 ms. Gradient strengths in G/cm (lengths in ms) are $g_1 = 9.5$ (1.0), $g_2 =$ variable (1.0), $g_3 = 28.5$ (0.5), $g_4 = 47.5$ (0.125), $g_5 = 19.0$ (0.2), $g_6 = 7.6$ (0.1), $g_7 = 16.2$ (0.6). The phase cycling employed is: $\phi_1 = \{x, -x\}$, $\phi_2 = \{2(x),2(y),2(-x),2(-y)\}$, $\phi_3 = \{-x,x\}$, $\phi_4 = \{x, -x\}$, $\phi_5 = \{x\}$, $\phi_{rec} = \{2(x),2(-x)\}$. Quadrature in F_1 is obtained by STATES-TPPI of ϕ_1.[50] 2D datasets were acquired as 118×1024 complex points, with spectral widths of 1500×8000 Hz in F_1 and F_2, respectively. Eight scans were averaged with a recycle delay of 1.5 second, so that a single 2D dataset was recorded in under one hour. Values for g_2 were 9.88, 15.58, 21.85, 29.07, 37.24, 46.74 and 57.0 G/cm. Duplicate datasets were acquired for values of $\phi_5 = \{x, -x\}$. An additional F_2 zero order phase correction of 90° was applied to data acquired with $\phi 5 = -x$ prior to addition and subtraction of data obtained with $\phi 5 = x$ and $\phi 5 = -x$ thereby generating data sets comprised of single multiplet components in F_2. Decoupling is not employed in any interval of the experiment, as described in the text.

ation is achieved by varying the field strength of encoding and decoding gradients, $g2$, while keeping the delays between them constant. The encoding and decoding gradients are split and merged with the polarization transfer steps of a 2D constant time (*CT*) heteronuclear single quantum correlation (HSQC) experiment. The diffusion labeling gradients (in gray) bracket the constant-time evolution period ($2T_N$) for chemical shift labeling and a period where the magnetization is stored along the *z*-axis (T_Z), which exploits the slower relaxation of longitudinal *versus* transverse nitrogen magnetization, as in the LED experiment.[46] Note also that the evolution period precedes the delay T_Z (~110 ms) so that exchange between the folded and unfolded species during this time would generate cross-peaks. Notably we have not observed such peaks in the present study, in agreement with the very slow exchange rates between folded and unfolded states (<1 s^{-1}) at 5 °C measured previously.[26] The absence of cross-peaks ensures that the diffusion coefficients obtained from the auto-peaks of the folded/unfolded states are not contaminated due to the exchange process.

In order to minimize sample heating throughout the course of the measurements decoupling was not employed during any stage of the experiment. We have observed that decoupling can create temperature gradients leading to convection currents along the *Z*-axis, resulting in signal attenuation and hence an increase in the measured diffusion constant.[45] When decoupling was applied errors were particularly noticeable in studies of the U_{gdn} state of the drkN SH3 domain where the protein was dissolved in 2 M GdnCl. The absence of decoupling during acquisition does, however double the number of cross-peaks. In order to minimize spectral crowding we have elected to record F_2-inphase ($\phi 5 = x$) and F_2-antiphase ($\phi 5 = -x$) spectra in an interleaved manner. Addition and subtraction of the resultant data sets produces spectra where cross-peaks are displaced by $^1J_{HN}$ Hz from each other in the F_2 dimension, and diffusion constants can be measured using both spectra thereby decreasing the measurement error by $\sqrt{2}$, with the number of cross-peaks identical to what is obtained in decoupled spectra.

We have also developed a similar experiment which records 2D ^{13}C-^1H CT-HSQC correlation maps and used this experiment to measure diffusion constants from the intensity *versus* gradient strength profile of methyl correlations. Notably, identical diffusion constants were measured using both ^{15}N and ^{13}C-based approaches once decoupling was eliminated from both experiments.

A series of well-resolved 2D ^{15}N-^1H HSQC spectra were obtained for different values of the encoding/decoding gradient field strength $g2$ and peak intensities fitted to the relation:

$$I(G) = I(0)\exp\{-(\gamma G\delta)^2 D[\Delta - \delta/3 - 3\beta/4 + \beta'/4]\} \quad (10)$$

which has been derived for the experiment of Figure 8. In equation (10) G is the field strength of the encoding and decoding gradients, $g2$, γ is the proton gyromagnetic ratio (2.675×10^4 rad G^{-1} s^{-1}), D is the diffusion coefficient and the delays Δ, δ, β and β' are defined in the Figure. Equation (10) was obtained following the approach described by Stejskal and Tanner[48] and reduces to their well-known result for the case where the bipolar gradient pairs merge to single gradient pulses of length δ ($\beta = \beta' = 0$). Notably, amide peak intensities were obtained for 72 % of residues in the folded state and 32 % of the available residues in the unfolded state.

Small-angle X-ray scattering

All measurements were made at the Stanford Synchrotron Radiation Laboratory (SSRL) at a sample temperature of 5 °C. The beam energy was selected to be 8980 eV (Cu K-edge) using a high flux multi-layer monochromator (Mo/B_4C). To avoid radiation-induced aggregation inherent in the measurements, the protein samples were re-circulated through a polypropylene flow cell. A maximum flow rate of 1.3 ml/minute was used to minimize the time any particular volume of sample remained in the beam. Each sample was measured for ten minutes. Radii of gyration were calculated according to the Guinier approximation:[49]

$$\ln(I(S)) = \ln(I(0)) - \frac{4\pi^2 R_g^2}{3} S^2 \quad (11)$$

where R_g is the radius of gyration, $S = (2\sin\theta)/\lambda$, with 2θ the scattering angle and λ is the X-ray wavelength. The fitting ranges used were 0.0041-0.0175 $Å^{-1}$, 0.0045-0.0142 $Å^{-1}$ and 0.0044-0.0095 $Å^{-1}$, for the F_s, F_{exch}/U_{exch} and U_{gdn} samples, respectively. Even with the comparatively large volumes of protein used in the study, the time-dependent increase in R_g suggested that there was some radiation damage to the protein and therefore only the first two minutes of data were used in the analysis. An error in R_g of ±0.5 Å was estimated from repeat measurements.

Generation of unfolding trajectories from TRADES

Starting from the NMR structure of the folded state of the drkN SH3 domain (Singer and Forman-Kay, unpublished), C^α atoms were displaced in an iterative manner. The amount of displacement is dependent on two user-defined parameters, the temperature and the timestep, in such a way that the C^α displacement increases with both parameters.[40] After each displacement, the remaining atoms from backbone and side-chain positions are added, using a side-chain rotamer library. In this manner 600 unfolding trajectories were generated, each with 100 structures. 300 of the trajectories were obtained with the temperature and the timestep set to 298 K and 100 fs, respectively, while the remaining structures were generated using values of 400 K and 400 fs for the temperature and timestep parameters. Further details can be found in Choy and Forman-Kay.[5]

Generation of unfolding trajectories from high temperature molecular dynamics

Molecular dynamics trajectories were kindly provided by M. Philippopoulos and R. Pomes (unpublished results). These simulations were preformed using AMBER version 6 with the AMBER PARM96.DAT parameter set starting from a fully hydrated drkN SH3 domain structure. 9600 structures were extracted from a 10 ns trajectory at 498 K, with an additional 592 structures obtained from a 3.55 ns trajectory recorded at 598 K, with the final structure of the 498 K run used as input for the higher temperature simulation.

Calculation of hydrodynamic properties using the program HYDROPRO

Hydrodynamic radii and radii of gyration for the 60,000 structures from the TRADES unfolding trajectory and for the 10,200 structures from molecular dynamics

simulations were calculated using the program HYDRO-PRO.[38] Details of the calculation can be obtained from the original reference. Briefly, for each structure, a preliminary hydrodynamic model was obtained by replacing non-hydrogen atoms with spherical elements of fixed radius *a*. The resulting particle, comprised of overlapping spheres, is then used to obtain a rough shell model, consisting of a shell of mini-beads of radius σ, upon which the hydrodynamic forces act. Hydrodynamic radii are then calculated using a value of $a = 3$ Å, which takes into account the effects of hydration,[38] assumed the same for folded and unfolded states. A value of $\sigma = 1.74$ Å was used in all calculations.

Acknowledgments

We thank Dr José García de la Torre (University de Murcia, Spain) and Dr Hue Sun Chan (University of Toronto) for many useful discussions. W.-Y.C. and F.A.A.M. are research fellows supported by a grant from AstraZeneca UK Limited. L.E.K. and J.D.F-K. acknowledge grant support from the CIHR. L.E.K. is an International Research Scholar of the Howard Hughes Medical Institute.

References

1. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321-331.
2. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. (1996). Structural studies of $p21^{Waf1/Cip1/Sdi1}$ in the free and Cdk2-bound state: Cconformational disorder mediates binding diversity. *Proc. Natl Acad. Sci. USA,* **93**, 11504-11509.
3. Campbell, K. M., Terrell, A. R., Laybourn, P. J. & Lumb, K. J. (2000). Intrinsic structural disorder of the C-terminal activation domain from the bZIP transcription factor Fos. *Biochemistry,* **39**, 2708-2713.
4. Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C. & Brown, C. J. (2000). Intrinsic protein disorder in complete genomes. *Genome Infomatics,* **11**, 161-171.
5. Choy, W.-Y. & Forman-Kay, J. D. (2001). Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* **308**, 1011-1032.
6. Serrano, L. (1995). Comparison between the φ distribution of the amino acids in the protein database and NMR data indicates that amino acids have various φ propensities in the random coil conformation. *J. Mol. Biol.* **254**, 322-333.
7. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M. & Dobson, C. M. (1996). Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255**, 494-506.
8. Fiebig, K. M., Schwalbe, H., Buck, M., Smith, L. J. & Dobson, C. J. (1996). Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem.* **100**, 2661-2666.
9. Dyson, H. J. & Wright, P. E. (1991). Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 519-538.
10. Wishart, D. S. & Sykes, B. D. (1994). The C-13 chemical-shift index- a simple method for the identification of protein secondary structure using C-13 chemical-shift data. *J. Biomol. NMR,* **4**, 171-180.
11. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E. & Dyson, H. J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123**, 2970-2978.
12. Mok, Y.-K., Kay, C. M., Kay, L. E. & Forman-Kay, J. D. (1999). NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J. Mol. Biol.* **289**, 619-638.
13. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158-169.
14. Gillespie, J. R. & Shortle, D. (1997). Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. Mol. Biol.* **268**, 170-184.
15. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science,* **293**, 487-489.
16. Doniach, S. (2001). Changes in biomolecular conformation seen by small-angle X-ray scattering. *Chem. Revs.* **101**, 1763-1778.
17. Sosnick, T. R. & Trewhella, J. (1992). Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry,* **31**, 8329-8335.
18. Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O. & Doniach, S. (1998). Kinetics of lysozyme refolding: structural characterization of a non-specifically collapsed state using time-resolved X-ray scattering. *J. Mol. Biol.* **276**, 225-237.
19. Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S. & Kiefhaber, T. (1999). Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J. Mol. Biol.* **288**, 489-499.
20. Jones, J. A., Wilkins, D. K., Smith, L. J. & Dobson, C. M. (1997). Characterisation of protein unfolding by NMR diffusion measurements. *J. Biomol. NMR,* **10**, 199-203.
21. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. & Smith, L. J. (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry,* **38**, 16424-16431.
22. Pan, H., Barany, G. & Woodward, G. (1997). Reduced BPTI is collapsed. A pulsed field gradient NMR study of unfolded and partially folded bovine pancreatic trypsin inhibitor. *Protein Sci.* **6**, 1985-1992.
23. Morgan, C. J., Wilkins, D. K., Smith, L. J., Kawata, Y. & Dobson, C. M. (2000). A compact monomeric intermediate identified by NMR in the denaturation of dimeric triose phosphate isomerase. *J. Mol. Biol.* **300**, 11-16.
24. Olivier, J. P., Raabe, T., Henkemeyer, M., Dickson, B., Mbamalu, G., Margolis, B. *et al*. (1993). A *Drosophila* SH2-SH3 adaptor protein implicated in coupling the sevenless tyrosine kinase to an activator of Ras guanine nucleotide exchange, Sos. *Cell,* **73**, 179-191.
25. Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V. & Schlessinger, J. (1993). SH3 domains direct cellular localization of signaling molecules. *Cell,* **74**, 83-91.
26. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1994). A heteronuclear correlation experiment

for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR*, **4**, 727-734.
27. Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994). Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR,* **4**, 845-858.
28. Zhang, O. & Forman-Kay, J. D. (1995). Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry,* **34**, 6784-6794.
29. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1995). Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry*, **34**, 868-878.
30. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1997). Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry,* **36**, 2390-2402.
31. Zhang, O. & Forman-Kay, J. D. (1997). NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions. *Biochemistry*, **36**, 3959-3970.
32. Zhang, O., Forman-Kay, J. D., Shortle, D. & Kay, L. E. (1997). Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. *J. Biomol. NMR,* **9**, 181-200.
33. Yang, D., Mittermaier, A., Mok, Y.-K. & Kay, L. E. (1998). A study of protein side-chain dynamics from new ^{2}H auto-correlation and ^{13}C cross-correlation NMR experiments: application to the N-terminal SH3 domain from drk. *J. Mol. Biol.* **276**, 939-954.
34. Yang, D., Mok, Y.-K., Muhandiram, D. R., Forman-Kay, J. D. & Kay, L. E. (1999). ^{1}H-^{13}C dipole-dipole cross-correlated spin relaxation as a probe of dynamics in unfolded proteins: application to the DrkN SH3 domain. *J. Am. Chem. Soc.* **121**, 3555-3556.
35. Mok, Y.-K., Elisseeva, E. L., Davidson, A. R. & Forman-Kay, J. D. (2001). Dramatic stabilization of an SH3 domain by single substitution: roles of the folded and unfolded states. *J. Mol. Biol.* **307**, 913-928.
36. Segel, D. J., Fink, A. L., Hodgson, K. O. & Doniach, S. (1998). Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of Cytochrome c. *Biochemistry,* **37**, 12443-12451.
37. Flory, P. J. & Fisk, S. (1966). Effect of volume exclusion on the dimensions of polymer chains. *J. Chem. Phys.* **44**, 2243-2248.
38. Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of proteins from their atomic level structures. *Biophys. J.* **78**, 719-730.
39. Oono, Y. & Kohmoto, M. (1983). Renormalization group theory of transport properties of polymer solutions. I. Dilute solutions. *J. Chem. Phys.* **78**, 520-528.
40. Feldman, H. J. & Hogue, C. W. V. (2000). A fast method to sample real protein conformational space. *Proteins: Struct. Funct. Genet.* **39**, 112-131.
41. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* **194**, 531-544.
42. Baudet, S. & Janin, J. (1991). Crystal structure of a barnase-d(GpC) complex at 1.9 Å resolution. *J. Mol. Biol.* **219**, 123-132.
43. Williams, R. W. & Teeter, M. M. (1984). Raman spectroscopy of homologous plant toxins: crambin and α1 and β-purothionin secondary structures, disulfide conformation and tyrosine environment. *Biochemistry,* **23**, 6796-6802.
44. Ramanadham, M., Sieker, L. C. & Jensen, L. H. (1990). Refinement of triclinic lysozyme: II. The method of stereochemically restrained least squares. *Acta Crystallog. sect. B,* **46**, 63-69.
45. Price, W. S. (1998). Pulsed-field gradient nuclear magnetic resonance as tool for studying translational diffusion: Part II. Experimental aspects. *Concepts Magn. Reson.* **10**, 197-237.
46. Gibbs, S. J. & Johnson, C. S., Jr (1991). A PFG NMR experiement for accurate diffusion and flow studies in the presence of eddy currents. *J. Magn. Reson.* **93**, 395-402.
47. Orekhov, V. Y., Korzhnev, D. M., Pervushin, K. V., Hoffman, E. & Arseniev, A. S. (1999). Sampling of protein dynamics in nanosecond time scale by ^{15}N NMR relaxation and self-diffusion measurements. *J. Biomol. Struct. Dynam.* **17**, 157-174.
48. Stejskal, E. O. & Tanner, J. E. (1965). Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.* **42**, 288-292.
49. Glatter, O. & Kratky, O. (1982). *Small Angle X-ray Scattering*, Academic Press, New York.
50. Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989). Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J. Magn. Reson.* **85**, 393-399.

Edited by P. E. Wright

(Received 4 October 2001; received in revised form 29 November 2001; accepted 6 December 2001)