doi:10.1016/j.jmb.2008.10.004 J. Mol. Biol. (2009) 385, 266–277

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

The Denatured State of N-PGK Is Compact and Predominantly Disordered

Matthew J. Cliff[1], C. Jeremy Craven[1], James P. Marston[1], Andrea M. Hounslow[1], Anthony R. Clarke[2] and Jonathan P. Waltho[1,3]*

[1]*Department of Molecular Biology and Biotechnology, The University of Sheffield, Western Bank, Sheffield S10 2TN, UK*

[2]*Department of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD, UK*

[3]*Faculty of Life Sciences and Manchester Interdisciplinary Biocentre, The University of Manchester, Manchester M1 7DN, UK*

Received 11 July 2008; received in revised form 26 September 2008; accepted 1 October 2008
Available online
11 October 2008

Edited by K. Kuwajima

The organisation of the structure present in the chemically denatured N-terminal domain of phosphoglycerate kinase (N-PGK) has been determined by paramagnetic relaxation enhancements (PREs) to define the conformational landscape accessible to the domain. Below 2.0 M guanidine hydrochloride (GuHCl), a species of N-PGK (denoted I_b) is detected, distinct from those previously characterised by kinetic experiments [folded (F), kinetic intermediate (I_k) and denatured (D)]. The transition to I_b is never completed at equilibrium, because F predominates below 1.0 M GuHCl. Therefore, the ability of PREs to report on transient or low population species has been exploited to characterise I_b. Five single cysteine variants of N-PGK were labelled with the nitroxide electron spin-label MTSL [(1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanesulfonate] and the denaturant dependences of the relaxation properties of the amide NMR signals between 1.2 and 3.6 M GuHCl were determined. Significant PREs for I_b were obtained, but these were distributed almost uniformly throughout the sequence. Furthermore, the PREs indicate that no specific short tertiary contacts persist. The data indicate a collapsed state with no coherent three-dimensional structure, but with a restricted radius beyond which the protein chain rarely reaches. The NMR characteristics of I_b indicate that it forms from the fully denatured state within 100 μs, and therefore a rapid collapse is the initial stage of folding of N-PGK from its chemically denatured state. By extrapolation, I_b is the predominant form of the denatured state under native conditions, and the non-specifically collapsed structure implies that many non-native contacts and chain reversals form early in protein folding and must be broken prior to attaining the native state topology.

Keywords: paramagnetic relaxation enhancement; NMR; folding intermediates; protein folding; phosphoglycerate kinase

Introduction

The mechanism through which a protein folds from a disordered state is influenced not only by the interactions required to stabilise the native structure, but also by the nature and amount of residual structure in the initial ensemble. Early observations of residual structure in denatured state ensembles of a number of proteins prompted the suggestion that native-like structure is present in the denatured state[1–3] and that this biases the subsequent conformational search towards the native conformation.[2,4] However, it is becoming increasingly apparent that regions of non-native structure are also stable in the denatured state.[5–7] The recent observation that the N-terminal domain of phosphoglycerate kinase (N-PGK) from *Geobacillus stearothermophilus* has significant non- native-like secondary structure in its denatured state ensemble[8] has prompted a reevaluation of its mechanism of folding.

*Corresponding author. Department of Molecular Biology and Biotechnology, The University of Sheffield, Western Bank, Sheffield S10 2TN, UK. E-mail address: J.Waltho@sheffield.ac.uk.

Abbreviations used: N-PGK, the amino terminal domain of phosphoglycerate kinase from *Geobacillus stearothermophilus*; PRE, paramagnetic relaxation enhancement; GuHCl, guanidine hydrochloride; MTSL, (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) methanesulfonate.

Non-native structure in the denatured state ensemble of N-PGK was inferred from changes in behaviour as denaturant concentrations were decreased.[8] Briefly, ^{1}H and ^{15}N chemical shifts of the backbone amide resonances change throughout the titration, indicating constant changes in the structural ensemble. The behaviour deviated significantly from that expected of a two-state transition; for some amides, the slope of the ^{1}H chemical shift changes with denaturant concentration reverses sign during the titration, indicating the population of at least three states with different guanidine hydrochloride (GuHCl) concentration dependences. The current minimal description of the chemically denatured state ensemble of N-PGK therefore comprises three species, previously termed D, I_a, and I_b (Fig. 1 and Reed *et al.*[8]). I_a and D represent extremes of an unresolvable continuum of species observed at high denaturant concentrations, while I_b represents a distinct species observed at lower denaturant concentrations. The denatured species (D, I_a, I_b) all interconvert on timescales faster than 0.1 ms, whereas the native state is formed at least 3 orders of magnitude more slowly. I_b is never the dominant species at equilibrium. Therefore, the folding landscape of N-PGK comprises at least five states, as represented in Fig. 1: the three species of the denatured ensemble, the previously identified kinetic intermediate (I_k),[11] and the folded state, F. It has not been demonstrated how the five states interconvert to form the folding pathway, but it is argued that I_b can be identified with the denatured state under native conditions[8] and hence, is transiently the dominant species upon dilution from denaturant prior to folding.

The denatured states of proteins present a challenge to structural biology, because any structure present tends to fluctuate rapidly. Furthermore, they are rarely populated under the conditions of greatest interest, i.e., when folding dominates. However, an established means of investigating their properties

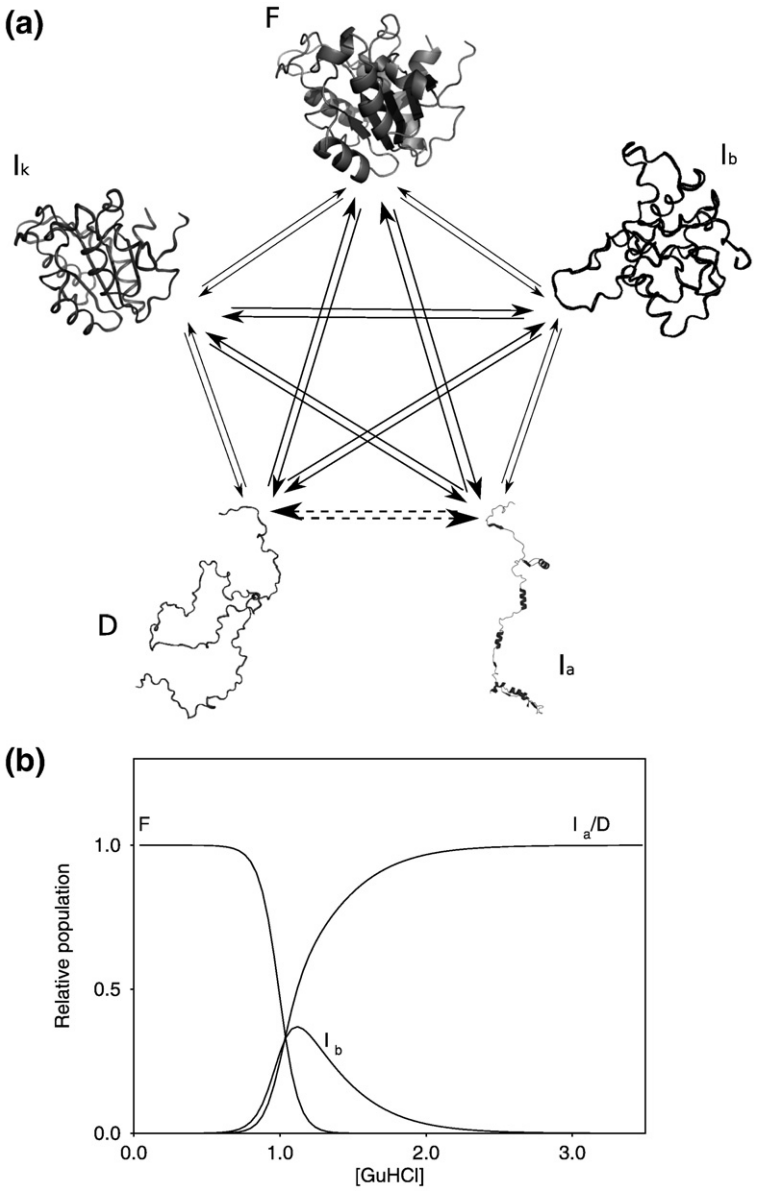

Fig. 1. The five known states of N-PGK. (a) Structural representation of the five states. (b) Denaturant dependence of the equilibrium populations of the five states. The folded conformation (F) predominates in native conditions; the kinetic folding intermediate (I_k) is populated transiently upon initiation of folding in native-like conditions; and there are three members of the denatured ensemble, the fully denatured state (D), the extended, partially structured state (I_a), and the collapsed, partially structured state (I_b). (a) F is based on the structure of the full-length protein[9] (Protein Data Bank accession code 1PHP). I_k has some native structure, on the basis of hydrogen-exchange protection data.[10] I_b is compact, with some helical content, and I_a is expanded with some helical content.[8] D has no coherent structure. (b) The equilibrium populations were calculated on the basis of previously presented data.[8] I_a and D are treated as one state for simplicity, because they are not discreet states and represent extremes of a continuum. I_k is not populated at equilibrium (<0.1%) and I_b never exceeds 30% of the equilibrium population.

uses paramagnetic relaxation agents.[12] Site-specific labelling with a thiol-specific nitroxide electron spin-label, such as (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanesulfonate (MTSL),[13] allows interresidue distances to be derived from its effect on NMR relaxation behaviour. MTSL can exist in two electronic states, and in its oxidised (paramagnetic) form, an unpaired electron in MTSL induces attenuation of the intensity of NMR signals in a well-defined, distance-dependent manner, termed paramagnetic relaxation enhancement (PRE), from resonances that are within approximately 25 Å of the spin-label. Regions of a polypeptide chain that form persistent tertiary interactions with the spin-labelled region will exhibit strong PRE effects. Conversely, regions that remain distant at all times should exhibit weak PREs. Hence, an arrangement such as a four-helix bundle should result in a distribution of PREs that diagnostically reflects the protein fold.

In order to determine whether there is persistent tertiary organisation in I_b, here we report conclusions from PRE measurements following the site-specific incorporation of MTSL into N-PGK. In contrast to the majority of PRE-based studies of denatured proteins, we have used the denaturant dependence of the observed PREs to allow the analysis of a partially populated folding intermediate, applying the ability of paramagnetic relaxation to detect transiently populated species.[14] The intensity of any peak in a ^{15}N heteronuclear single quantum coherence (HSQC) spectrum is partially defined by the transverse relaxation rate, which is the time average of the relaxation rates that the amide experiences. Therefore, brief population of a species with increased relaxation (e.g., due to proximity to a spin-label) increases this time average and consequently, the peak intensity is reduced substantially more than if the effect were due to population alone. Hence, with due averaging between the states involved (D, I_a, I_b), the tertiary organisation and compaction in the partially populated I_b state can be defined. The structural content and organisation of this state and its relation to the previously characterised kinetic folding intermediate (I_k) impinge on the current understanding of the folding trajectories of N-PGK and potentially other large protein domains that fold through kinetic intermediates.

Results

Collection of PRE data

The PRE is measured by comparing the intensities of NMR signals from protein with a paramagnetic spin-label with those from an identical protein sample with the spin-label reduced (diamagnetic). PRE data were recorded for N-PGK with the MTSL spin-label incorporated at the position of the single cysteine (C18) in the wild-type domain, and with the spin-label at the position of the cysteine in four variants—L56C, L80C, L113C and V142C (all in a background of the C18V N-PGK variant). For each cysteine-containing variant, the measurements required the recording of two sets of ^{15}N-HSQC spectra across a range of denaturant concentrations: (i) with the spin-label in its paramagnetic (oxidised) state and (ii) with the spin-label in its diamagnetic (reduced) state. The comparison of intensities of signals between the paramagnetic and diamagnetic spectra allowed the PRE effect to be isolated from other influences, such as dilution, viscosity and changes in sensitivity with denaturant concentration.

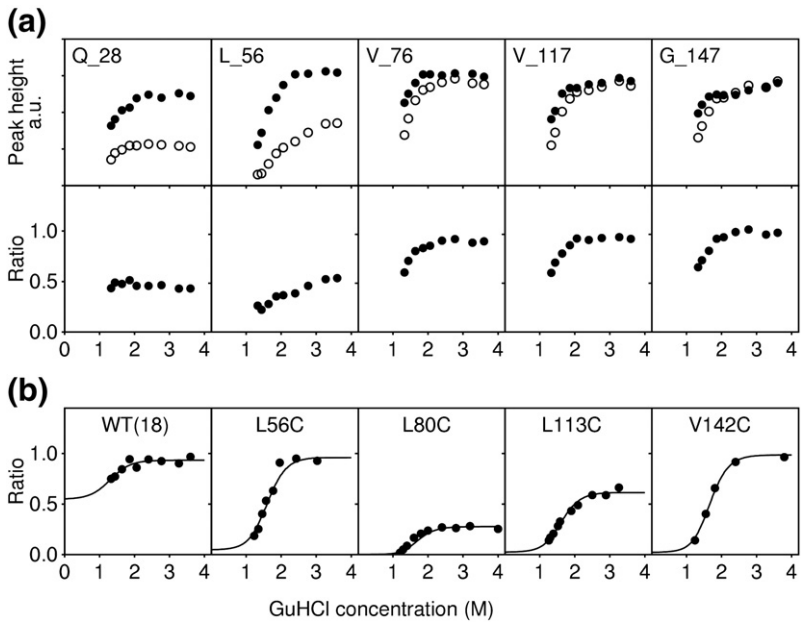

Fig. 2. The effects of the spin-label. (a) Effect of sequence separation with spin-label at residue 18. (b) Effect at residue 89 in the five spin-label variants. (a) The denaturant concentration dependence of the NMR signal intensities in ^{15}N HSQC spectra are shown for the indicated residues (in arbitrary units), with the spin-label at position 18. Intensities (peak heights) are shown for the oxidised (paramagnetic; upper panel, open circles) and reduced (diamagnetic; upper panel, filled circles) samples, and the intensity ratios are also shown (lower panel). NMR spectra and titrations were performed as detailed in the Methods section. A variety of behaviours is displayed (see the text). (b) The denaturant concentration dependence of the intensity ratios (as in the lower panel of (a)) are shown for residue 89 in each of the five spin-label variants. The continuous line indicates the best-fit parameters to Eq. (3).

The data for the wild-type domain under conditions where the spin-label is diamagnetic are shown in Fig. 2a (upper panel, filled circles). Under these conditions, it is expected that the MTSL should have a negligible effect on the relaxation properties of the nuclei. Indeed, the data behave essentially as described by Reed *et al.*[8]: the intensities are relatively constant above 2.0 M GuHCl, but below this concentration the intensities decrease for many residues, due to increased transverse relaxation (R_2). The decrease in intensity is not due to the onset of the transition to the folded state, F, since it is not observed for all residues, and at the lowest denaturant concentration for which data are shown (1.2 M GuHCl) the population of F is still negligible (~5%).

With the spin-label in the paramagnetic form, a number of different effects are observed in the wild-type domain (Fig. 2a, upper panel, open circles). Residues such as Q28 or L56 exhibit reduced peak intensities throughout the titration. This indicates that these residues are sufficiently close to the spin-label to be influenced by it across the range of GuHCl concentrations used (1.2 to 3.6 M), the expected random coil behaviour for residues at these sequence separations. More complex behaviour is observed for residues such as V76, V117 or G147, where the paramagnetic spin-label only affects intensities at low denaturant concentration. These residues are therefore distant from the spin-label at higher denaturant concentrations, whereas below 2.0 M GuHCl these residues are closer to the spin-label in the states populated. These effects are summarised by the ratio of the intensities of the signals with the spin-label in the two forms (Fig. 2a, lower panel). Example data from the other protein variants are shown in Fig. 2b. In common with the previous analysis of the chemical shift data,[8] the timescale of exchange with the folded state (1 s^{-1})[11] excludes it from contributing to the relaxation behaviour. Hence, the increased PREs cannot be explained by increased population of the folded state, but rather are a consequence of a contraction of the denatured state ensemble. The identification of the population of a distinct state at low denaturant concentrations using PREs is independent of the previous identification of I_b on the basis of chemical shift changes,[8] but since the transition is occurring over the same range of denaturant concentration in the two studies, the state is also termed I_b here.

Determination of relative populations of D, I_a and I_b

The measured PREs are determined by contributions from each component of the denatured ensemble. The relative populations of these states change with denaturant, and therefore defining the relationships describing these changes in population allows the contribution from each state to be resolved. The components of the denatured ensemble exchange rapidly,[8] and so their populations are reflected in the observed chemical shifts, although the population ascribed to each component is dependent on the model employed to describe the transitions. The transitions between the denatured ensemble components of N-PGK are not fully resolved and the high denaturant transition is not complete in the range of accessible denaturant concentrations. The chemical shift changes at higher denaturant concentrations (2.5 to 4.0 M) show an approximately linear dependence (Fig. 3 and Methods) and, hence, further resolution of this transition was not sought. Furthermore, the measured intensity ratios remain constant above 2.0 M GuHCl (Fig. 2), indicating that the PREs in the previously described[8] species I_a and D, which predominate in the denaturant concentration range

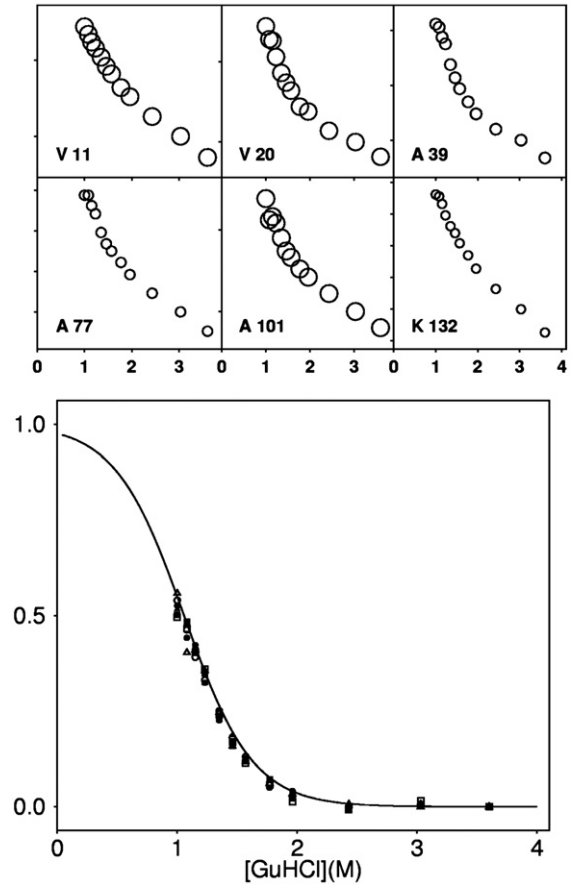

Fig. 3. Chemical shift changes with denaturant concentration. Weighted, combined chemical shift changes are shown (upper panels) for six example residues in the reduced, spin-labelled L56C variant that have significant chemical shift changes as a function of denaturant concentration. Weighted combined chemical shift changes are calculated as:

$$\Delta\delta_{\mathrm{norm}} = \sqrt{\left[(\delta H - \delta \mathrm{H}_{3.6\ \mathrm{M}})^2 + \{0.1(\delta N - \delta N_{3.6\ \mathrm{M}})\}^2\right]}$$

In the lower panel, the slope corresponding to the I_a/D transition has been subtracted from the same data, the data have been normalised and plotted on the same axes (open circles, V11; filled circles, V20; open squares, A39; filled squares, A77; open triangles, A101; and filled triangles, K132). The continuous line shows the best fit of the data to Eq. (1), as in Table 1.

2.5 to 4.0 M, are indistinguishable. Therefore, in order to minimise the number of parameters introduced, I_a and D were treated as a single state in the analysis that follows.

The chemical shift changes at lower denaturant concentrations are represented well by a two-state transition (see Methods) between I_b and I_a/D (Fig. 3). The examples in Fig. 3 show no significant differences in behaviour of residues from throughout the protein sequence, indicating globally co-operative behaviour. For each variant, the values of the equilibrium constant ($K_{2(w)}$) and its denaturant dependence (m_2) (Table 1) were obtained by fitting the average change in chemical shift (of all residues exhibiting a measurable change) to a two-state transition (Eq. (1); see Fig. 3). Although it was not possible to measure chemical shifts over the whole transition, the central portion of the transition is well represented by the data leading to definition of the values of $K_{2(w)}$ and m_2. Consequently, the extent of transition at 1.2 M GuHCl can be defined as being between 58% and 72% for the data shown in Fig. 3 (using parameters corresponding to 2 standard deviations either side of the best fit). The values show no significant variation between variants and are consistent with previous estimates[8] for wild-type unmodified N-PGK. The values obtained ndicate that, in the absence of denaturant, I_b is only marginally more stable than the more denatured I_a/D state ($K_{2(w)}$ is between 20 and 100), with a small degree of surface area burial relative to that of the folded state ($m_2 = -3.5$ M^{-1}, $m_1 = -14$ M^{-1} Ref. 15), consistent with the behaviour of a molten globule.

Inference of PRE effects in I_b

Values for the contribution of I_b to the observed effects of the paramagnetic spin-label were obtained by fitting the PRE data to Eqs. (3) and (5), using values for $K_{2(w)}$ and m_2 for each variant from Table 1, and estimates for $R_{2,obs}$ for each residue from lineshape analysis of ^{15}N-HSQC spectra (see Methods). These fits yielded values for the paramagnetic contribution to the transverse relaxation rate in the I_b state, $R_P^{I_b}$. The fits did not yield reliable values for $R_P^{I_a/D}$ below 2 s^{-1}, as the difference between an infinitely slow rate, which gives rise to a ratio of 1, and a rate of 2 s^{-1}, which gives rise to a ratio of 0.95, is beyond the resolution of the experiment. However, the derived values of $R_P^{I_b}$ vary only twofold when $R_P^{I_a/D}$ is fixed at the two limiting values (2 s^{-1} and 1×10^{-13} s^{-1}). The absolute values of $K_{2(w)}$ and m_2 also have only a small effect, with the values of $R_P^{I_b}$ only varying by twofold with values of $K_{2(w)}$ and m_2 within one standard error of the values in Table 1. The change associated with any uncertainty in $K_{2(w)}$ and m_2 is consistent over the whole protein, making the distribution of values identical. Example fitted curves for the intensity ratios are shown in Fig. 2b.

Table 1. Equilibrium constants and m values for the I_b-to-I_a/D transition

Variant	$K_{2(w)}$	m_2 (M^{-1})
WT	100 (40–260)	3.9 (3.3–4.4)
C18V, L56C	48 (24–95)	3.8 (3.3–4.2)
C18V, L80C	31 (16–58)	3.0 (2.7–3.3)
C18V, L113C	70 (30–170)	3.5 (3.0–4.0)
C18V, V142C	20 (2–240)	3.5 (2.4–4.5)

The values in parentheses are the mean and one standard deviation ranges, as estimated using the jackknife procedure (see Methods). In the case of $K_{2(w)}$, the standard deviations were calculated for $\log(K_{2(w)})$.

The fitted values of $R_P^{I_a/D}$ and $R_P^{I_b}$ are plotted as a function of residue number in Fig. 4. The paramagnetic contributions to relaxation are greater in I_b than in I_a/D, confirming that I_b has a more compact structure than I_a (see below for an analysis of the degree of compaction that these data imply). Notably, the magnitude of the effect of the spin-label in I_b only varies by 2 orders of magnitude over the full length of the protein (outside the region of the spin-label), compared to the 4 orders of magnitude in variation that would be expected in a topologically ordered structure where certain close contacts would dominate (such as in the native structure; continuous line in Fig. 4). In addition, there is no strong chain-length dependence for the values of $R_P^{I_b}$, as in the simple power law expansion expected for a random coil (grey line in Fig. 4).

The observation that the PRE data for all residues display roughly the same behaviour prompts the consideration of non-structural interpretations of the data. The PRE effects are independent of protein concentration, ruling out any effect due to transient aggregation. The change in behaviour with denaturant rules out a simple effect due to free spin-label, and denaturant-dependent, non-specific association of free spin-label with the unfolded protein can be discounted, as there are no signals in the 1D ^{1}H-NMR of the reduced (diamagnetic) sample that could arise from any free spin-label (at any denaturant concentration, see supplementary information).

Determination of compaction of I_b and I_a/D

The distribution of $R_P^{I_b}$ values (Fig. 4) indicates that neither long-lived tertiary contacts nor highly preferred topologies are made in I_b, and thus, it is predominantly disordered. However, the elevated values of $R_P^{I_b}$ imply a substantial degree of compaction relative to I_a/D. In order to establish the level of compaction in I_a/D, the distribution of the experimental intensity ratios at 3.6 M GuHCl was compared with values calculated for a theoretical random coil. Since polypeptides are not homogenous, freely jointed chains, and some residues show strong preference for particular φ/ψ angle combinations, the structures of an ensemble of 1000 random coils for the N-PGK sequence were modelled as described in Methods. The modelling procedure takes into account the φ/ψ angle propensities of different amino acids and the excluded volume due to the polypeptide. The structures in the model ensemble were used to calculate values of R_P for

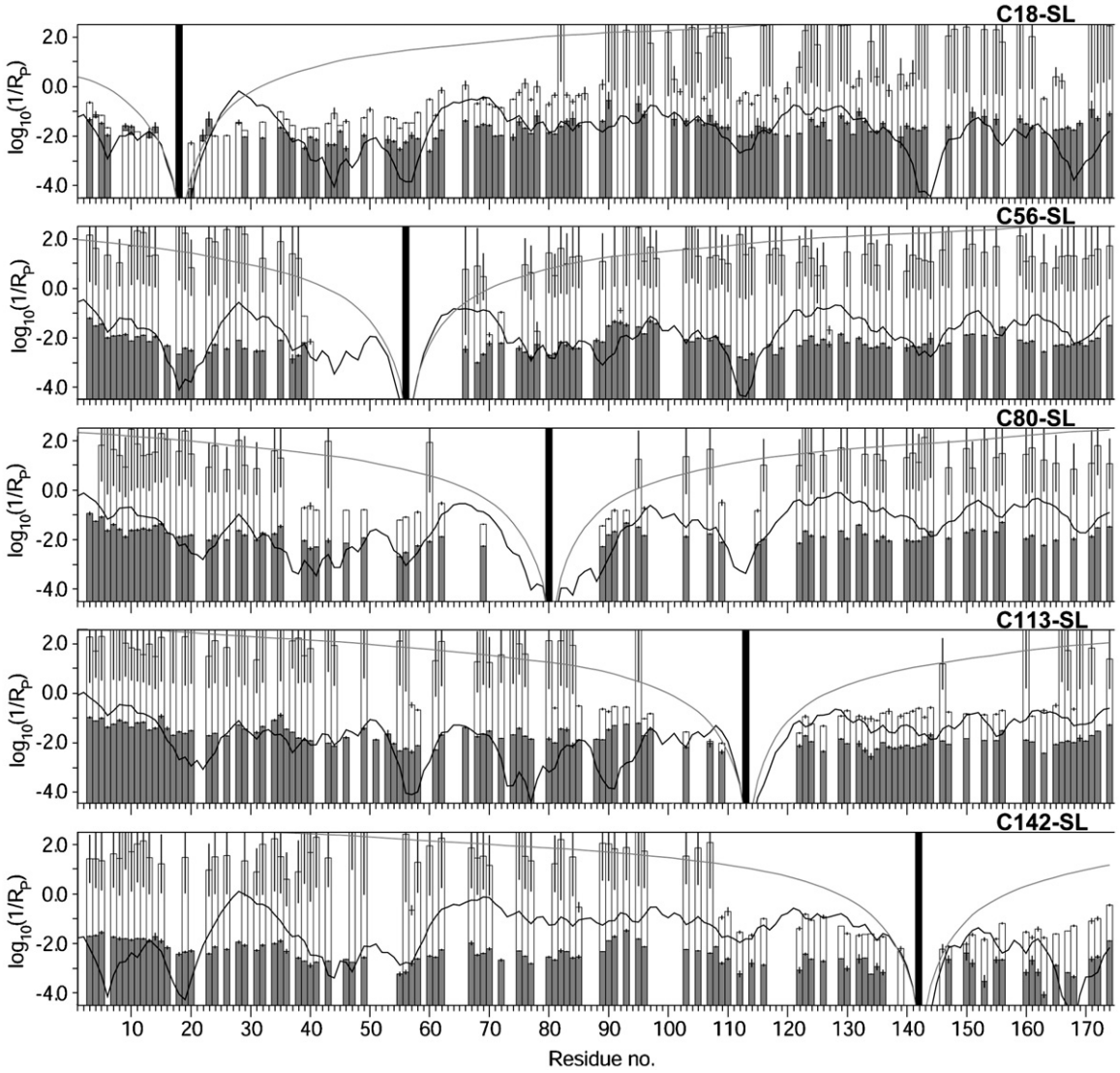

Fig. 4. Sequence distribution of PREs in I_b and I_a/D. The fitted paramagnetic contribution to relaxation is shown as the logarithm of the paramagnetic relaxation times [$\log(1/R_P^X]$ where X is the protein state) for each residue. Low values correspond to low peak-height ratios and vice versa. The values derived for I_a/D (white bars) and I_b (grey bars) from fits of Eq. (3) to data like that shown in Fig. 2b are shown for each of the spin-label variants. The black bar represents the position of the spin-label in each variant. The heavy continuous line shows the calculated values for the folded state for each variant, based on Eq. (5) and the crystal structure of the full-length protein (Protein Data Bank accession code 1PHP), and the grey continuous line shows the values calculated for a simulated ensemble of random coils (see Methods). Peaks with significant overlap or of extremely low intensity were excluded from the analysis, and only values well defined by the fits are shown. Data for residues close in sequence to the mutation sites are also missing because they were not assigned.

such an ensemble. The agreement of intensity ratios calculated using these R_P estimates with the experimental ratios is good (crosses *versus* circles in Fig. 5), and therefore N-PGK behaves in a manner indistinguishable from a random coil at 3.6 M GuHCl.

The increase in the values of R_P between I_a/D and I_b (Fig. 4) indicates that I_b is considerably more compact than I_a/D. The values of $R_P^{I_b}$ indicate that the level of compaction is relatively similar throughout the protein, behaviour inconsistent with a semi-collapsed random-coil model of a polypeptide, where the RMS interresidue distance is a function of sequence separation. Instead, the data are better explained by assuming that I_b consists of substantially collapsed structures and are consistent with a model where the chain is constrained to lie within a relatively compact sphere. In such a model, for sequence separations greater than the diameter of the sphere, the chain will have folded back upon itself. Thus, a residue at a large sequence separation still has a high chance of coming close to the spin-label. In order to model such an ensemble, the folded state was used as a representative limiting compact structure, and random pairs from within the folded state crystal structure coordinates were selected to sample the distribution of distances accessible to a

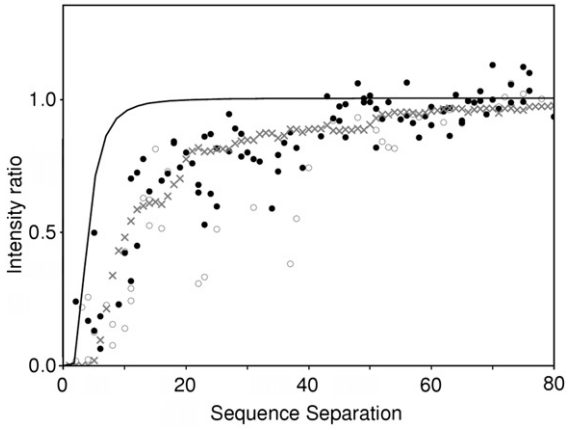

Fig. 5. Relationship of I_a/D to a random coil. The relationship between PRE and sequence separation from the spin-label is shown for a freely jointed, random walk polymer with unit size 3.5 Å (calculated, continuous line), the simulated ensemble of random-coil polypeptides based on the N-PGK sequence (see Methods, crosses), and the observed data at 3.6 M GuHCl when the spin-label is at position 18 (open circles) and 80 (filled circles).

spin-label–amide pair within this ensemble. These distances were used to calculate R_P values on the basis of Eq. (5), which were then averaged to produce an estimate of the value of R_P for this amide. These calculations did not reproduce the measured R_P values, but repeating the process and multiplying all calculated distances in the fully compact structure by a factor of 1.2 gives an average R_P value of $\sim 50\ s^{-1}$, in agreement with the values for most residues outside the region of the spin-labelled residue. Therefore, it appears that I_b has a size approximately 20% larger than that of the folded state, i.e., with a radius of gyration (R_g) of approximately 17 Å. For comparison, the calculated R_g for the modelled unfolded ensemble is 37 Å.

Effect of spin-label and mutation on structure in the denatured states

The ^{15}N-HSQC spectrum of wild-type N-PGK at 3.6 M GuHCl was almost entirely unaffected by derivatisation of the cysteine at position 18 with reduced MTSL, except in the region very close to the spin-label. The only residues exhibiting significant shift changes (>0.05 ppm in ^{1}H, >0.5 ppm in ^{15}N) were the labelled cysteine and its nearest sequence neighbours (±1). Similarly, mutations induced only minor changes in NMR spectra. As for derivatisation, the only residues exhibiting significant shift changes were the mutated residue and its two nearest sequence neighbours. All the necessary mutations are conservative and the single cysteine variants have been previously characterised and fold by the same mechanism.[15] Furthermore, the titrations of the reduced, diamagnetic spin-label variants all closely followed the trends previously seen in the unlabelled wild-type protein.[8] Therefore, neither the introduction of the probe nor the point mutations significantly perturbs the structural propensities of the various GuHCl-induced states.

Discussion

PRE measurements such as those presented here give coarse-grained residue–residue distance information, and hence report on tertiary-structure arrangements. The PRE data for N-PGK at high denaturant concentrations (3.6 M) resemble closely the values calculated for an ensemble of random coils (Fig. 5), as expected for a protein in highly denaturing conditions. At lower denaturant concentrations (below 2 M), there is an obvious increase in the PREs for most residues (examples in Fig. 2 and Supplementary Fig. 1), indicating a global contraction of the protein from the random-coil state. The denaturant concentration profiles of the PREs are independent of the spin-label position, indicating that the transition to a more compact form is, or is close to, concerted throughout the protein. The data therefore show a subpopulation of the denatured ensemble to behave as a distinct, more compact state, which is populated significantly at lower denaturant concentrations. The presence of a distinct denatured state of N-PGK populated at low denaturant was proposed in our previous paper,[8] on the basis of the denaturant concentration dependence of chemical shifts, size-exclusion chromatography, and CD spectroscopy. The state was denoted I_b. A further state, I_a, was also defined on the basis of the chemical shift measurements, but the distance resolution provided by PREs is too low to distinguish this state from the fully denatured state (D).

Structural characterisation of I_b

Despite it never becoming the dominant species at equilibrium, certain properties of I_b have been inferred previously.[8] Analysis of far-UV CD experiments indicated that significant secondary structure is present in the denatured state ensemble above the denaturant concentration where the population of F is less than 50%, and ^{13}C chemical shifts revealed that some of this secondary structure is non- native-like. In particular, some regions that form β-strands in the native state become helical in the denatured state ensemble at low denaturant concentration. The denaturant dependence of the elution time from size-exclusion chromatography showed I_b to be partially collapsed, with a hydrodynamic radius between those of the fully denatured and the native states.[8] In addition, significant differential broadening of the amide NMR signals was observed between 1.0 and 2.0 M GuHCl. Hence, I_b was proposed to have the properties associated with molten globule states: partially collapsed, with secondary structure, and yet retaining extensive dynamics.

Interpretation of the raw PRE data to yield structural information for I_b is not trivial, as the combined effects of mixed populations and residue-specific transverse relaxation contributions to the PREs are

difficult to delineate. For example, increased relaxation due to proximity to a paramagnetic centre in I_b can cause intensity reductions at denaturant concentrations where this state is only marginally populated. The observation of a large paramagnetic effect at high denaturant concentration can be ascribed either to close proximity in a rare population of I_b or to proximity in I_a/D. For this reason, the fitting strategy outlined in Methods was employed to define the component paramagnetic contributions to the observed data from the underlying states. This strategy is independent of, but consistent with, the data of Reed *et al.*[8]

The analysis of the data has made use of a two-state transition to describe the denaturant-dependent transition between I_b and I_a/D. Equivalent denaturant dependences for the changes in chemical shifts and line-shapes at multiple positions across the protein support the assumption of a cooperative transition between two states (Fig. 3 and Ref. 8). The alternative limiting model, that of a continuum between I_b and I_a, has been used previously to describe the unfolding of molten globules,[16] and non-cooperative unfolding has been demonstrated for some such states.[17] However, the two-state and the continuum models represent limiting extremes of a range of possible behaviours, which are not easily distinguished analytically,[18] and because the estimate of the relative population of I_b is not greatly affected by the method used to calculate it, the mathematically more simple two-state model was employed.

The derived values for $R_P^{I_b}$ show even less variation than the raw intensity ratios imply, as the effect of (diamagnetic) line-broadening accounts for much of the differences between residues (Fig. 4). There are, perhaps, slightly slower values of $R_P^{I_b}$ associated with the N-terminus of the protein. The values of $R_P^{I_b}$ indicate that neither persistent short nor long amide–spin-label (i.e., residue–residue) distances are present. This contrasts with the calculated distribution of values for R_P for the native state, F, shown as an example of an ordered tertiary structure. The observed behaviour allows two models for I_b to be discounted. Firstly, structures that resemble an expanded version of the folded state (modelled by recalculating the values of R_P for the folded state with all distances increased by 20%) give rise to similar sequence distributions for the values of $R_P^{I_b}$ and R_P^F, but with less extreme maxima and minima in the values for I_b (by at least an order of magnitude). Similarly, a non-native-like ordered tertiary structure will give a comparable distribution, but with the regions of high and low $R_P^{I_b}$ values rearranged. The pattern of $R_P^{I_b}$ observed therefore implies that no structure is substantially more favoured over any other compact arrangement. In addition, the data are not reproduced by a compact random walk structure,[16] as there is no strong sequence separation dependence.

The lack of a strong tertiary structure bias for I_b is somewhat in contrast to the secondary structure bias observed from ^{13}C secondary chemical shifts.[8] Some helical content is associated with regions 67–82 and 129–151 and, to a lesser extent, 33–53 and 97–117. It therefore seems unlikely that the compaction of the polypeptide chain is completely random, and thus some structures are favoured over others. There is some evidence from the contact map of PRE effects (Supplementary Fig. 4) supporting the view that regions closer in the primary sequence come together slightly more often than those more distant in sequence, but the data are sparse and the PRE effects are small relative to their uncertainties. Hence, any non-random behaviour present is insufficient to dominate the PRE data. The data for N-PGK therefore contrast with those studies that show a large native-like bias in the denatured state ensemble[1–3,19–21], and with the data for "classical" molten globules[22] and have much in common with the data collected for the denatured states of DrkN-

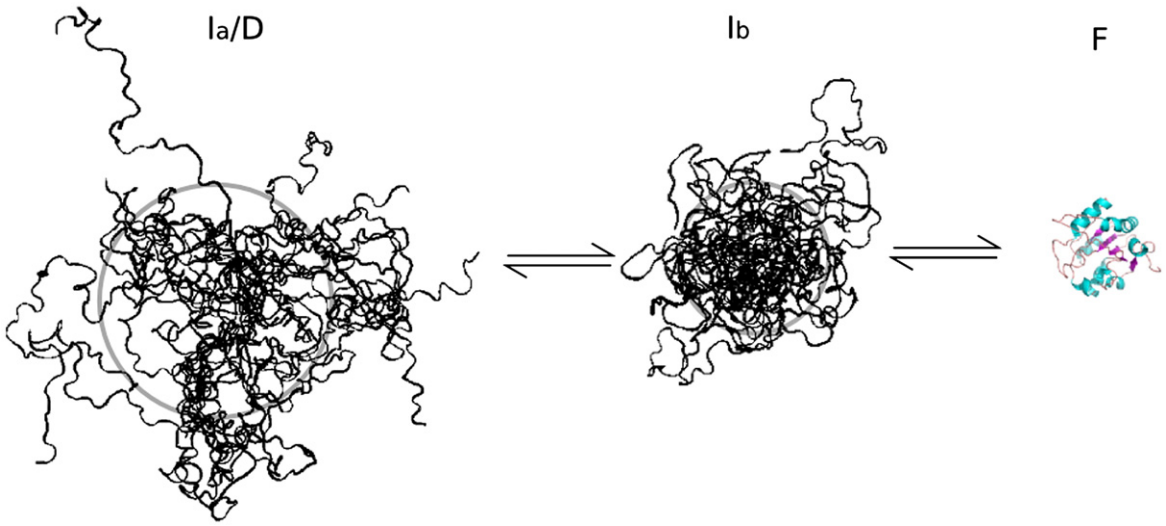

Fig. 6. Representation of the denatured states of N-PGK relative to the folded state. Each state (I_a/D, I_b) is represented as a circle with radius proportional to its radius of gyration, as calculated from the model ensembles to be consistent with the paramagnetic relaxation data, except the folded state is shown as scaled molecular structure.

SH3[7,23] and ACBP,[6,24] which are compact but show little persistent organisation, and both native-like and non-native regions. The data for N-PGK are best reproduced by a model of I_b consisting of a multitude of diverse, collapsed structures with radii approximately 20% larger than that of the folded state. Such a relationship between the states is visualised in Fig. 6.

The above description of I_b allows some conclusions to be made about the dynamics within the I_b ensemble. Line-broadening in the I_b state of N-PGK was reported previously[8] and is evident in the loss of intensity in the data for the reduced sample in Fig. 2a. The process responsible cannot be exchange between I_b and I_a/D, because dynamics on a timescale that causes line-broadening would result in correlations between the magnitude of chemical shift change for a particular amide and the level of broadening it exhibits, which is not observed (data not shown, and Reed *et al.*[8]). Exchange between I_b and I_a/D is therefore too fast to affect linewidths. Exchange through I_a/D will always be the favoured pathway between conformations within I_b, and consequently, although it is likely to be on a slower timescale, such interchange cannot cause the observed line-broadening. Alternatively, the increased relaxation rates could be due to millisecond timescale exchange with a yet to be identified species that is not accessible from I_a/D. However, the simplest explanation of these data for N-PGK is an increase in correlation time for the amide bonds. For instance, an increase in correlation time from 4 ns in D to 12–14 ns in I_b would account for all the increases in linewidths and the losses in signal intensity. Such an increased correlation time is consistent with a change from the mostly unfolded, rapidly fluctuating I_a and D states to a globular, partially rigidified I_b state. The differential broadening within I_b would thus indicate that there is a range of residual mobility in the domain.

Role of I_b in folding

The ability of N-PGK to form a predominantly disordered, collapsed state (I_b) from the more extended denatured states on a timescale much more rapid (transitions between I_b, I_a, and D all occur in less than 0.1 ms) than formation of the native state ($\sim 10\ s^{-1}$),[11] fits with a range of observations reported previously for other proteins. For example, non-folding mutants of cytochrome *c* and RnaseA[25,26] both rapidly undergo hydrophobic collapse, but fold no further. Other proteins have also been shown to fold through substantially non-native intermediates. For example, β-lactoglobulin[27] forms a transient intermediate that contains non-native helical content, which can be identified with an I_b-like state. A carbonic anhydrase and a β-lactamase were also shown to fold through discreet non-native intermediates, described in that study as "pre-molten globule states".[28] Indeed, the observation of common, 1-anilino-8-naphthalene sulfonate-binding species observed in the folding of a range of proteins[29] could also be equated with the population of I_b-like states. The implication is that most proteins undergo such a transition on removal of denaturant. A widespread occurrence of predominantly disordered, collapsed states is consistent with previous proposals of the population of a continuum of collapsed states, from non-native I_b-like states to partially native-like intermediates prior to the major transition state to folding.[30–32] In this model of the folding landscape, the I_b state of N-PGK results from a rapid compaction from the random-coil denatured state prior to selection of more native-like structure in I_k and then attaining the folded state. Whether either of these intermediates are "on-pathway" is not resolved at present.

The number of hydrophobic interactions that drive the initial collapse towards I_b-like states increases with chain-length. One predicted consequence of this is the population of unacceptably stable kinetic traps in the folding of larger proteins, which in turn would contribute to the pressure that leads many large proteins to be constructed from smaller protein domains.[8] However, it should be noted that some non-native contacts can also promote folding, as reported in disulfide exchange dominated folding[33] and in the folding of CD2.d1 and fynSH3.[34,35] Also, it is intriguing that molecular dynamics simulations of the folding of villin headpiece[36] gave a folding pathway that involves many attempted collapses before a correct native-like contact is formed, which then allows the rest of the folding to condense around a nucleus. The simulations illustrate a mechanism for folding, where hydrophobic collapse limits the conformational space to be searched for the folding nucleus, as originally proposed for the hydrophobic collapse model of protein folding.[16]

Conclusions

Here we have demonstrated how the application of PRE measurements can be used to report on the properties of transiently populated states within the denatured state ensemble, even when the species never dominates the population. The analysis of such measurements has allowed the characterisation of a protein folding intermediate of N-PGK, termed I_b. This intermediate, which equates to the denatured state of N-PGK under folding conditions, is collapsed, and yet has almost no coherent structure and non-native bias in certain regions. Such properties contrast with measurements that have identified native-like structural properties for denatured states and have more in common with reports of the behaviour of denatured states of cytochrome *c*, truncated RNase A, DrkN-SH3 and ACBP.[7,23–26] Hence, it is likely that these less structured compact denatured states, like the I_b state of N-PGK, are a part of a continuum of behaviour stretching up to very native-like molten globules. The substantial population of the I_b state of N-PGK means that the conformational search required to attain the native state from the denatured state must be more extensive than the

sequential acquisition of native-like contacts from an extended random coil.

Methods

Source of chemicals

GuHCl and DTT were obtained from Melford (UK) and MTSL was obtained from Toronto Research Chemicals (Canada). All other chemicals were AnaLAR grade, obtained from BDH.

Standard buffer

All solutions contain 20 mM Tris, 20 mM Bistris, 0.5 mM ethylenediaminetetraacetic acid, and 3 mM sodium azide, pH 6.0.

Production of ^{15}N-labelled wild-type and mutant N-PGK

The expression vector for wild-type N-PGK comprising residues 1–174 of phosphoglycerate kinase from *G. stearothermophilus* has been described previously.[15] Mutants were produced by the Quikchange™ method. BL21(DE3) strains of *E. coli* transformed with the appropriate expression vectors were incubated at 37 °C in minimal M9 media with [^{15}N]ammonium chloride as the sole nitrogen source, and expression was induced by addition of 1 mM IPTG once an OD_{600} of 0.8 was reached followed by incubation overnight. The expressed protein was purified as previously described.[15]

Preparation of NMR samples and MTSL labelling

Twenty milligrams protein in 4 M GuHCl, 1 mM DTT was mixed with an MTSL stock [5% (w/v) in methanol] to give a final 5:1 excess of MTSL over the total thiol concentration. Samples were extensively buffer exchanged into 1.3 M GuHCl and concentrated to ca a 1 mM. Trimethylsilyl-2,2,3,3-tetradeuteropropionic acid was added to 1 mM and D_2O was added to 10% (v/v). The resulting solution was divided into two 200-µl samples, one of which was reduced with 5 protein equivalents of sodium ascorbate.

Collection of NMR spectra

Gradient-selected, sensitivity-enhanced ^{15}N-HSQC spectra were collected at 25 °C on a Bruker Avance 600-MHz spectrometer equipped with a cryoprobe. Samples were in 3-mm-diameter tubes in order to reduce sensitivity losses in the very lossy high-GuHCl-concentration samples. Spectra were acquired with acquisition times of 133 ms (direct dimension) and 150 ms (indirect dimension). An exponential line broadening of 7 Hz was applied in the direct dimension, and a cosine-squared window function was applied in the indirect dimension. Spectra were processed with Felix (Felix NMR Inc.). Chemical shifts were referenced relative to trimethylsilyl-2,2,3,3-tetradeuteropropionic acid, measured for every denaturant concentration and oxidation state sample. Intensities (measured as peak-heights) were corrected for sample dilution.

Assignment of the amide resonances at 3.6 M GuHCl was achieved by comparison with the previously assigned wild-type domain.[8] Peaks that were constant between the C18V variants but not present in wild-type were assigned to residues 17, 18, and 19. Likewise, comparison of the wild-type and reduced MTSL-labelled samples allowed the identification of cross peaks from residues close in sequence to the MTSL label. These cross peaks also were absent when the MTSL label was oxidised. Essentially complete labelling of the protein was confirmed by the lack of residual unlabelled peaks. Spectra at 1.3 M GuHCl were also acquired with 10-fold more dilute protein samples in order to test whether the chemical shift and intensity changes observed were due to some form of aggregation. Chemical shifts and intensities were affected minimally, and so this possibility can be discounted. Data are shown in Fig. 4 for all residues where overlap between peaks was minimal over a sufficient region of the titration to accurately define the parameters $R_P^{I_b}$ and $R_P^{I_a/D}$.

Definition of equilibrium constants and *m* values

The model in Fig. 1 is simplified for the purpose of this analysis to the following three-state pathway:

$$I_a/D \overset{K_2}{\rightleftharpoons} I_b \overset{K_1}{\rightleftharpoons} F$$

where the equilibrium constants between the folded state F, and the compact denature state I_b (K_1) and between I_b and the ensemble I_a/D (K_2) have associated *m* values, m_1 (defined as $d\{\ln(K_1)\}/d[\mathrm{GuHCl}]$) and m_2 (defined as $d\{\ln(K_2)\}/d[\mathrm{GuHCl}]$), respectively.

Data analysis

All data were analysed using in-house routines, with fitting by non-linear least-squares minimisation based on the Levenberg–Marquadt algorithm.

Determination of $K_{2(w)}$ and m_2

The transition from I_a/D to I_b is in fast exchange and so, with the assumption of a two-state transition, chemical shift data for each mutant was fitted to Eq. (1):

$$\langle \Delta\delta_{\mathrm{obs}} \rangle = \frac{\langle \Delta\delta_{I_b - I_a/D} \rangle \times K_2 + d[\mathrm{GuHCL}]}{K_2 + 1} \qquad (1)$$

using:

$$K_2 = K_{2(\mathrm{w})} \exp(-m_2[\mathrm{GuHCl}]) \qquad (2)$$

where $\langle \Delta\delta_{\mathrm{obs}} \rangle$ is the mean value (averaged over all residues) of the change in chemical shift relative to the values observed at 3.6 M GuHCl; $\langle \Delta\delta_{I_b - Ia/D} \rangle$ is the mean chemical shift difference between I_a/D and I_b in the absence of denaturant; *d* is the slope of the chemical shift for I_a/D with denaturant concentration;[8] $K_{2(\mathrm{w})}$ is the value of K_2 at 0 M GuHCl (i.e., in water). A jackknife routine was used to estimate errors.

Determination of paramagnetic relaxation rates

The intensity ratio for a residue at any particular denaturant concentration is determined by five parameters: the apparent transverse proton relaxation rate in the absence of paramagnetism (including a contribution due to apodisation of the free induction decay), $R_{2,\mathrm{obs}}$; the paramagnetic contributions to the proton transverse

relaxation rate for each state, $R_P^{I_b}$ and $R_P^{I_a/D}$; and the equilibrium constant and m value for the transition between the two states, $K_{2(w)}$ and m_2. (The effect of the spin-label on the relaxation rate of ^{15}N nuclei is negligible.)[1] $K_{2(w)}$ and m_2 have been established for each variant from the fit to chemical shift data discussed above. Values for $R_{2,obs}$ were derived for each residue at each GuHCl concentration from the proton linewidths in the reduced sample. The other two parameters ($R_P^{I_b}$ and $R_P^{I_a/D}$) were determined for each residue by non-linear least-squares fitting methods from the intensity ratios (I_{para}/I_{dia}), using the following equation:[1,37]

$$\frac{I_{para}}{I_{dia}} = \frac{R_{2,obs}\exp(-R_{P,obs}t)}{R_{2,obs}+R_{P,obs}} \quad (3)$$

with $R_{P,obs}$ defined as a population-weighted average,

$$R_{P,obs} = \frac{R_P^{Ia/D} + K_2 R_P^{Ib}}{1+K_2} \quad (4)$$

and K_2 defined as in Eq. (2). t indicates the time that the proton magnetisation is transverse in the HSQC pulse sequence (10.6 ms). Values of $R_{P,obs}$ were not measured directly from the spectra for use in the fitting procedure, as the definition of broader linewidths was significantly hampered by low signal-to-noise and overlapping peaks. However, the fitted values obtained for $R_{P,obs}$ were validated by comparing the fitted values with the proton linewidths of the oxidised sample for sufficiently well-resolved and intense cross peaks. Errors in peak intensities were calculated from the RMS of the noise in baseline areas of the spectra, and these were used to propagate errors through the data analysis in the conventional manner. Consequently, the error estimates for the values of $R_P^{I_a/D}$ and $R_P^{I_b}$ were derived from the covariance matrix outputted by the non-linear least-squares algorithm.

Relating paramagnetic enhancements to distances

The paramagnetic contribution to transverse relaxation[37] has an inverse sixth-power relationship with electron–proton distance:

$$R_P = \kappa r^{-6}$$

where

$$\kappa = \gamma T_c\left(4+\frac{3}{1+\Omega^2 {T_c}^2}\right) \quad (5)$$

where γ is 1.23×10^{16} Å^6 s^{-2} for the interaction between a single electron and a proton, Ω is the proton Larmor frequency, T_c is the correlation time for the electron–nuclear dipole–dipole interaction and was set to 4 ns for I_a/D[1] and to 12 ns for I_b and F (as measured for the folded domain).

Simulation of a random-coil ensemble

Random sampling of a reduced Ramachandran angle set[38] for the N-PGK sequence was used as input into Protein Lego (supplied by M. A. Williams, Birkbeck College), which uses assumed planar peptide bonds to build protein chains from dihedral angles. Any clashes were removed by very short molecular mechanics (two steps of 5 ps) in the absence of attractive forces with the use of XPLOR.[39] The paramagnetic effect of a spin-label on a particular amide proton will be the time-average of the R_P experienced. This was estimated by averaging R_P values calculated for a particular residue-spin-label pair from each structure on the basis of Eq. (5). The relationship between R_P value and sequence separation did not vary significantly with label position, and so all combinations of the same sequence separation were averaged. $R_{2,obs}$ was set as 56 s^{-1} as estimated from the linewidths in the ^{15}N-HSQC spectra in order to calculate the intensity ratios in Fig. 5 from Eq. (3).

Simulation of a compact random ensemble

A compact sphere was modelled by using the crystal structure of N-PGK[9] as a limiting compact structure and taking the mean calculated R_P values [on the basis of Eq. (5)] for 10,000 random selections of residue pairs. This replicates the distance distribution for a pair of residues located randomly within the confines of the compact structure. Analysis of the sequence separation dependence of the calculated R_P values showed that a maximum mean value is reached after separations of 10. Pairs of residues with sequence separations less than 10 were therefore not included in the average. Multiplication of the distances by a factor of 1.2 prior to calculation of R_P produced a mean value close to the observed fitted values for $R_P^{I_b}$ (excluding values from 10 residues on either side of the spin-label).

Acknowledgements

We thank Maurizio Rossi for preparing the mutant plasmids. This work was supported by the BBSRC.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.10.004

References

1. 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.
2. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**, 1341–1351.
3. Lietzow, M. A., Jamin, M., Dyson, H. J. & Wright, P. E. (2002). Mapping long-range contacts in a highly unfolded protein. *J. Mol. Biol.* **322**, 655–662.
4. Daggett, V. & Fersht, A. (2003). The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell. Biol.* **4**, 497–502.
5. Platt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. (2005). Dynamics in the unfolded state of beta(2)-microglobulin studied by NMR. *J. Mol. Biol.* **346**, 279–294.
6. Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M. & Poulsen, F. M. (2005). Formation of native and non-native interactions in ensembles of denatured ACBP molecules from

paramagnetic relaxation enhancement studies. *J. Mol. Biol.* **347**, 1053–1062.
7. Marsh, J. A., Neale, C., Jack, F. E., Choy, W. Y., Lee, A. Y., Crowhurst, K. A. & Forman-Kay, J. D. (2007). Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J. Mol. Biol.* **367**, 1494–1510.
8. Reed, M. A. C., Jelinska, C., Syson, K., Cliff, M. J., Splevins, A., Alizadeh, T., Hounslow, A. M. *et al.* (2006). The denatured state under native conditions: a non-native-like collapsed state of N-PGK. *J. Mol. Biol.* **357**, 365–372.
9. Davies, G. J., Gamblin, S. J., Littlechild, J. A. & Watson, H. C. (1993). The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. *Proteins*, **15**, 283–289.
10. Hosszu, L. L. P., Craven, C. J., Spencer, J., Parker, M. J., Clarke, A. R., Kelly, M. & Waltho, J. P. (1997). Is the structure of the N-domain of phosphoglycerate kinase affected by isolation from the intact molecule? *Biochemistry*, **36**, 333–340.
11. Parker, M. J., Spencer, J. & Clarke, A. R. (1995). An integrated kinetic-analysis of intermediates and transition-states in protein-folding reactions. *J. Mol. Biol.* **253**, 771–786.
12. Mittag, T. & Forman-Kay, J. D. (2007). Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Sruct. Biol.* **17**, 3–14.
13. Berliner, L. J., Grunwald, J., Hankovszky, H. O. & Hideg, K. (1982). A novel reversible thiol-specific spin-label: papain active site labeling and inhibition. *Anal. Biochem.* **119**, 450–455.
14. Iwahara, J. & Clore, G. M. (2006). Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature*, **440**, 1227–1230.
15. Cliff, M. J., Alizadeh, T., Jelinska, C., Craven, C. J., Staniforth, R. A. & Waltho, J. P. (2006). A thiol labelling competition experiment as a probe for sidechain packing in the kinetic folding intermediate of N-PGK. *J. Mol. Biol.* **364**, 810–823.
16. Dill, K. A., Bromberg, S., Yue, K. Z., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein-folding—a perspective from simple exact models. *Protein Sci.* **4**, 561–602.
17. Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat. Struct. Biol.* **4**, 630–634.
18. Parker, M. J. & Marqusee, S. (1999). The cooperativity of burst phase reactions explored. *J. Mol. Biol.* **293**, 1195–1210.
19. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.
20. Kortemme, T., Kelly, M. J. S., Kay, L. E., Forman-Kay, J. & Serrano, L. (2000). Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* **297**, 1217–1229.
21. Kato, H., Feng, H. Q. & Bai, Y. W. (2007). The folding pathway of T4 lysozyme: the high-resolution structure and folding of a hidden intermediate. *J. Mol. Biol.* **365**, 870–880.
22. Kuwajima, K. (1989). The molten globule state as a clue of understanding the folding and cooperativity of globular-protein structure. *Proteins*, **6**, 87–103.
23. Mok, Y. K., Kay, C. M., Kay, L. E. & Forman-Kay, J. (1999). NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J. Mol. Biol.* **289**, 619–638.
24. Teilum, K., Kragelund, B. B. & Poulsen, F. M. (2002). Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. *J. Mol. Biol.* **324**, 349–357.
25. Qi, P. X., Sosnick, T. R. & Englander, S. W. (1998). The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nat. Struct. Biol.* **5**, 882–884.
26. Sosnick, T. R., Schtilerman, M. D., Mayne, L. & Englander, S. W. (1997). Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl Acad. Sci. USA*, **94**, 8545–8550.
27. Kuwata, K., Shastry, R., Cheng, H., Hoshino, M., Batt, C. A., Goto, Y. & Roder, H. (2001). Structural and kinetic characterization of early folding events in β-lactoglobulin. *Nat. Struct. Biol.* **8**, 151–155.
28. Uversky, V. N. & Ptitsyn, O. B. (1996). Further evidence on the equilibrium "pre-molten globule state": four-state guandinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* **255**, 215–228.
29. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. (1990). Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett.* **262**, 20–24.
30. Parker, M. J., Session, R. B., Badcoe, I. G. & Clarke, A. R. (1996). The development of tertiary interactions during the folding of a large protein. *Folding Des.* **1**, 145–156.
31. Parker, M. J. & Marqusee, S. (2001). A kinetic folding intermediate probed by native state hydrogen exchange. *J. Mol. Biol.* **305**, 593–602.
32. Parker, M. J. & Marqusee, S. (2000). A statistical appraisal of native state hydrogen exchange data: evidence for a burst phase continuum? *J. Mol. Biol.* **300**, 1361–1375.
33. Creighton (1977). Energetics of folding and unfolding of pancreatic trypsin inhibitor. *J. Mol. Biol.* **113**, 295–312.
34. Mason, J. M., Cliff, M. J., Sessions, R. B. & Clarke, A. R. (2005). Low energy pathways and non-native interactions: the influence of artificial disulfide bridges on the mechanism of folding. *J. Biol. Chem.* **280**, 40494–40499.
35. Di Nardo, A. A., Korzhnev, D. M., Stogios, P. J., Zarrine-Afsar, A., Kay, L. E. & Davidson, A. R. (2004). Dramatic acceleration of protein folding by stabilization of a nonnative backbone conformation. *Proc. Natl Acad. Sci. USA*, **101**, 7954–7959.
36. Jayachandran, G., Vishal, V. & Pande, V. L. (2006). Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece. *J. Chem. Phys.* **124**, 164902.
37. Kosen, P. A. (1989). Spin labelling of proteins. *Methods Enzymol.* **177**, 86–121.
38. Gibbs, N., Clarke, A. R. & Sessions, R. B. (2001). *Ab initio* protein structure prediction using physicochemical potentials and a simplified off-lattice model. *Proteins: Struct. Funct. Genet.* **43**, 186–202.
39. Brunger, A. T. (1992). *XPLOR 3.1 Manual*, Yale University Press, New Haven, CT.