

promoting access to White Rose research papers



Universities of Leeds, Sheffield and York
<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Journal of Molecular Biology**.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/7921/>

Published paper

Cliff, M.J., Craven, C.J., Marston, J.P., Hounslow, A.M., Clarke, A.R. and Waltho, J.P. (2009) *The denatured state of N-PGK is compact and predominantly disordered*. *Journal of Molecular Biology*, 385 (1). pp. 266-277.

<http://dx.doi.org/10.1016/j.jmb.2008.10.004>

The denatured state of N-PGK is compact and randomly collapsed.

Running title: A randomly collapsed folding intermediate

Matthew J Cliff[†], C. Jeremy Craven[†], James P. Marston[†], Andrea M. Hounslow[†], Anthony R. Clarke[‡] and Jonathan P Waltho^{†#*}.

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

[‡]Department of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD. UK.

[#]Faculty of Life Sciences and Manchester Interdisciplinary Biocentre, The University of Manchester, Manchester M1 7DN, U.K.

*Correspondence to: J.Waltho@sheffield.ac.uk

Abbreviations: gsPGK: Phosphoglycerate kinase from *Geobacillus stearothermophilus*. N-PGK: the amino terminal domain of gsPGK. GuHCl: Guanidine Hydrochloride. MTSL: (1-oxy-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanesulfonate. PRE: paramagnetic relaxation enhancement. R_2 : transverse relaxation rate in the absence of paramagnetic enhancement. R_p : transverse relaxation rate due to the paramagnetic spin-label. K_1 : equilibrium constant between the folded, native state F and the denatured state ensemble. m_1 : $d(\ln K_1)/d[\text{GuHCl}]$. K_2 : equilibrium constant between the compact, denatured state \mathfrak{b} and the ensemble \mathfrak{b}_a/D . m_2 : $d(\ln K_2)/d[\text{GuHCl}]$.

Summary

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 (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, 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 microseconds, 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, disordered protein.

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^{1,2,3} 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^{5,6,7}. 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 ensemble⁸, has prompted a re-evaluation for its mechanism of folding.

Non-native structure in the denatured state ensemble of N-PGK was inferred from changes in behaviour as denaturant concentrations were decreased⁸. The current best description of the chemically denatured state ensemble of N-PGK comprises 3 species, termed D, I_a, and I_b (Figure 1 and Reed *et al.*⁸). I_a and D represent extremes of an apparent continuum of species observed at high denaturant concentrations and 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 0.1ms, whereas the native state is formed at least three orders of magnitude more slowly. I_b is never the dominant species at equilibrium. Therefore, the folding landscape of N-PGK comprises at least 5 states, as represented in Figure 1: the three species of the denatured ensemble, the previously identified kinetic intermediate (I_k)⁹, and the folded state, F. It has not been demonstrated how the five states interconvert to form the folding pathway, but it is proposed that I_b can be identified with the denatured state under native conditions⁸, and hence is transiently the dominant species upon dilution from denaturant, prior to folding.

Despite it never becoming the dominant species at equilibrium, certain properties of I_b can be inferred. Analysis of far-UV circular dichroism (CD) experiments indicated that significant secondary structure persists above the midpoint denaturant concentration (where the population of F is less than 50%), and ¹³C chemical shifts revealed that some of this secondary structure is non-native-like. In particular, some regions that are β-sheet in the native state become helical in the denatured state. The denaturant dependence of

the elution time during size exclusion chromatography (SEC) showed I_b to be partially collapsed, with a hydrodynamic radius between those of the denatured and the native states⁸. In addition, there is significant differential broadening of the amide NMR signals between 1.0 and 2.0 M GuHCl. The greatest chemical shift changes occur in the same denaturant range, suggesting the two phenomena are linked and population of I_b is linked to faster relaxation. Hence, I_b has the properties associated with a molten globule: partially collapsed, with secondary structure, and yet highly dynamic.

The denatured states of proteins present a challenge to structural biology, because any structure present tends to fluctuate rapidly, and the state of greatest interest, that from which the native state folds, is rarely populated. However, an established means of investigating their properties uses paramagnetic relaxation agents¹⁰. Site-specific labelling with a thiol-specific nitroxide electron spin-label, such as MTSL¹¹, allows inter-residue 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 from protein residues that are within approximately 35 Å of the spin-label, in a well-defined, distance-dependent manner, termed paramagnetic relaxation enhancement (PRE). The PRE is measured by comparing the intensities of NMR signals from protein with a paramagnetic spin-label, with those from protein with the spin-label reduced (diamagnetic). 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 species¹². The intensity of any peak in a ¹⁵N-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 significantly increased relaxation (e.g. due to proximity to a spin-label) will increase the

time average, and consequently reduce the peak intensity significantly more than the effect of population alone. Hence, with due averaging between the states involved (D , I_b , 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 understanding of the folding pathway of N-PGK, and potentially other large protein domains that fold through kinetic intermediates.

Results

Collection of PRE data.

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 mutants 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 salt concentration.

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. Correspondingly, 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 is shown (1.2 M GuHCl) the population of F is still negligible.

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 close 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 mutant proteins are shown in Figure 2b.

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 (D, I_a and I_b). The relative populations of the three states change with denaturant, and therefore defining the relationships describing these changes in population is necessary to allow the contribution from each state to be determined. The measured intensity ratios remain constant above 2.0 M GuHCl (Figure 2), indicating that the PREs in I_a and in D are very similar. Therefore, I_a and D were treated as a single state in the analysis that follows.

I_b and I_a/D are in fast exchange, and so the observed chemical shifts directly reflect the relative populations of I_b and I_a/D⁸. The equilibrium constant ($K_{2(w)}$) for the transition between I_b and I_a/D and its denaturant dependence (m_2) could therefore be determined directly from the variation of the chemical shift with denaturant concentration. For each mutant, this variation (averaged over all residues exhibiting a measurable chemical shift change) was fitted to a two state transition (Equation 1). The resulting values for the equilibrium constant $K_{2(w)}$ and m_2 are given in Table 1. The values show no significant variation between mutants, and are consistent with our previous estimates⁸ for wild-type unmodified protein. The values obtained indicate that, in the absence of denaturant, I_b is only marginally more stable than the more denatured I_a state ($K_{2(w)}$ is between 20 and 100), with a small degree of surface area burial relative to the folded state ($m_2 = -3.5 \text{ M}^{-1}$, $m_1 = -14 \text{ M}^{-1}$ ¹³), consistent with the behaviour of a molten globule.

Inference of PRE effects in I_b .

Direct measurement of the PRE effects in I_b is not possible, partly because of the low population of I_b at any denaturant concentration, and partly because of the signal attenuation due to the change in relaxation regime in I_b . For this reason, the contribution from I_b has to be extracted from the observed values, using the knowledge of the relative populations of species derived from the chemical shift analysis above. Specifically, values for the contribution to the transverse relaxation rates from the effect of the paramagnetic spin-label were obtained by fitting the PRE data to Equations 3 and 5, using values for $K_{2(w)}$ and m_2 from Table 1, and estimates for $R_{2,obs}$ from line-shape 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, values of $R_P^{I_b}$ are relatively independent of the value of $R_P^{I_a/D}$ and vary by only two-fold with different $R_P^{I_a/D}$ values. They are similarly independent of $K_{2(w)}$ and m_2 within one standard error of the values in Table 1. Example fitted curves for the intensity ratios are shown in Figure 2b.

The fitted values of $R_P^{I_a/D}$ and $R_P^{I_b}$ are plotted as a function of residue number in Figure 3. 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 there is little significant variation in the magnitude of the effect of the spin-label as a function of residue number for I_b . This is in contrast to the much greater variation that would be expected in a topologically ordered structure where certain close contacts would dominate (such as in the native structure; solid line in Figure 3), and to the simple power law expansion expected for a random coil (grey line in Figure 3).

Determination of compaction of I_b and I_a/D

The uniform distribution of $R_P^{I_b}$ values (Figure 3) indicates that neither long-lived tertiary contacts nor highly preferred topologies are made in I_b , and thus it is highly disordered. However, the elevated values of $R_P^{I_b}$ imply a substantial degree of compaction relative to

l_a/D . In order to establish the level of compaction in l_a/D , the distribution of the experimental intensity ratios at 3.6M 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 such an ensemble. The agreement of intensity ratios calculated using these R_p estimates with the experimental ratios is good (crosses versus circles in Figure 4), and therefore N-PGK behaves in a manner indistinguishable from a random coil at 3.6M GuHCl.

The increase in the values of R_p between l_a/D and l_b (Figure 4) indicates that l_b is considerably more compact than l_a/D . The values of $R_p^{l_b}$ are of similar magnitude throughout the protein. This behaviour is not consistent with a semi-collapsed random coil model of a polypeptide, where the rms inter-residue distance is a function of sequence separation. Instead, the data is better explained by assuming that l_b consists of substantially collapsed structures, and is 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 used to calculate distances and, hence, calculate R_p values on the basis of Equation 5. These R_p values were then averaged to produce an estimate of the value of R_p for an amide that has experienced such a combination of environments during relaxation. These calculations did not reproduce the measured R_p values, but repeating the process after multiplying all calculated distances in the fully compact structure by a factor of 1.2 gives an average R_p value of $\sim 50s^{-1}$, in agreement with the observed data. Therefore, it appears that l_b has a size approx 20% larger than 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 3.6M 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 shift changes greater than 0.1 ppm 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 same mechanism¹³. Furthermore, the titrations of the reduced, diamagnetic spin-label variants all closely followed the trends previously seen in the unlabelled wild-type protein⁸. Therefore, neither the introduction of the probe nor the point mutations significantly perturb the structural propensities of the various GuHCl-induced states.

Discussion.

Previously published chemical shift data recorded for chemically denatured N-PGK allowed inference of some of the structural properties of the denatured ensemble⁸. Briefly, it allowed the identification of a relatively compact state, I_b , populated at lower denaturant concentrations. This state contained regions of high helical content, in particular at residues 67-82 and 129-151, and to a lesser degree 33-53 and 97-117. To characterise this state further, the work presented here used paramagnetic relaxation enhancement (PRE) experiments to define the degree to which any structural elements are arranged. Increased relaxation due to proximity to a paramagnetic centre in I_b caused 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 . This ambiguity was resolved by denaturant titration, which allowed the estimation (to within 10%) of the relative population of the transient species (I_b) at any denaturant concentration. Thus, the contribution relating to I_b and to the more extended (high m-value) I_a/D species can be subdivided from the PREs for the whole denatured ensemble.

The analysis of the data used here has made use of a two-state transition to describe the 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 states⁸. Furthermore, the decrease in intensity ratios below 2.0 M GuHCl (indicating increased PREs, see Figure 2), is almost uniform in onset, irrespective of label position and sequence separation. The alternative limiting model, that of a continuum between I_b and I_a , has been used previously to describe the unfolding of molten globules¹⁴, and non-cooperative unfolding has been demonstrated for some such states¹⁵. However, the two-state and the continuum models represent limiting extremes of a range of possible behaviours, which are not easily distinguished analytically¹⁶, and therefore the mathematical simplicity of a two-state transition was employed to calculate the relative populations of the states. It is worth emphasising that because the change from I_a to I_b is an apparent global change in behaviour, the conclusions derived from the observed paramagnetic effects are not affected if the assumption of a two-state transition is not included.

Structural characterisation of I_b .

The increase in values of $R_{P,obs}$ at low denaturant indicates a shortening of average spin-label - amide distances, and confirms the compaction between I_a and I_b proposed previously on the basis of SEC measurements⁸. A defined tertiary structure would produce regions of sequential amides with high PREs and other regions where the amides have low PREs. This is evident in the calculated distribution of values for R_P for the native state, F (Figure 3). The wide variation in the value of R_P arising from such an ordered tertiary structure is not observed for I_b . Neither persistent short nor long amide – spin-label (i.e. residue-residue) distances are indicated by the values of $R_P^{I_b}$. Rather, outside the immediate vicinity of the spin-label, an almost uniform distribution of intermediate values of $R_P^{I_b}$ is observed, with the exception of slightly slower values being associated with the N-terminus. This 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. However, the observed distribution of $R_P^{I_b}$ values indicates that there is no specific, long-term organisation in I_b and the compaction, revealed by their intermediate values, is almost completely random. The inverse 6th power relationship

between distance and R_p means that any short distances resulting from native-like states in I_b would dominate over other contributions if such states were moderately well populated. The pattern of $R_p^{I_b}$ observed therefore implies that native-like structures are no more favoured than any other compact arrangement. The data are best reproduced by a model of I_b consisting of randomly collapsed structures with radii approximately 20% larger than the folded state. Such a relationship between the states is visualised in Figure 5.

The description of I_b as a large range of structures with similar properties that interconvert rapidly allows some conclusions about the dynamics within the I_b ensemble to be made. Line-broadening in the I_b -state of N-PGK was reported previously⁸. 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 have an effect on line-widths. Consequently, the interchange between different conformations within I_b , which is likely to be on a slower timescale, cannot cause line-broadening because exchange through I_a/D will always be the favoured pathway between conformations. 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 , but the most likely explanation 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 line-widths and the losses in signal intensity. Such an increased correlation time is consistent with a change from the mostly unfolded, rapidly fluctuating I_b 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. An alternative explanation, that the behaviour is due to aggregation, can be discounted here because the observed data are independent of protein concentration (see Methods).

Relationship of I_b to other partially folded states

Early work on partially denatured states focused on those formed under weakly denaturing conditions of low pH, and/or with crucial cofactors missing. The relationship of these classical molten globules to kinetically relevant folding intermediates has been the focus of much debate¹⁷. Recent work has extended the earlier studies, by using various stabilisation strategies to look at other denatured states more directly. It therefore is

pertinent to discuss the relationship of the I_b state of N-PGK to the classical molten globules, and then to these other partially denatured states.

The classical molten globules are characterised as having a high proportion of native-like secondary and tertiary structure, but being slightly more expanded than the native state¹⁷. In contrast to the native state, there is significant solvent penetration into the hydrophobic core of the protein¹⁸, and the NMR signals show attenuation due to conformational exchange¹⁵. Some of these properties are shared by I_b . Extrapolation of the CD spectrum associated with I_b ⁸ to 100% population implies it has almost 75% of the native helical content. In addition, I_b is compact, as shown using SEC⁸ and the PRE measurements in Figure 3, and it is associated with significant line-broadening of the amide NMR signals. However, the primary differences are the lack of coherent tertiary structure, and the significant non-native character. This raises the question of whether all reported classical molten globules are predominantly native-like in structure. While it is clear that some molten globules have near native behaviour (e.g. reduced cystatin¹⁹), for others_ the structure content has largely been inferred from CD spectroscopy and amide hydrogen exchange protection studies. Protection from hydrogen exchange is normally only detected for amides that remain protected in the native state, and so studies can be biased towards reporting native-like structure. Furthermore, the secondary structure reported by CD spectra cannot be assigned to a particular region of sequence. The secondary structure in molten globule states is often dominated by helices, where the interactions required are local and little of the peptide backbone requires rearrangement. Any non-native helical structure in the molten globule would also contribute to the 'native-like' CD spectra. Recent experimental developments have led to the structural characterisation of a broader range of denatured states¹⁰. The study of these denatured states usually required that they were stabilised to such a degree that they were the predominant state at equilibrium. One strategy to stabilise denatured states is to modify the chemistry of the polypeptide in some way. This could be through truncation¹, point mutation^{20,21,22} or changing the pH²³. Despite the more subtle control used in stabilising these states, most data collected reproduced the observations recorded for the classical molten globules; a compact state slightly larger than the folded state with large amount of native-like residual secondary structure.

Where chemical denaturants were used to stabilise the denatured states, as here, the proteins studied largely behaved as expanded, random-flight polymers at high denaturant

concentrations, according to radii of gyration (R_g) measured using SAXS²⁴ and diffusion measured using NMR²⁵. That this is not perfectly random behaviour over short sequence separations, was demonstrated for example by small but significant NMR residual dipolar couplings^{26,27,28}, and is generally ascribed to the ϕ/ψ angle biases of each residue. Other studies have reported some clustering of hydrophobic residues modifying the protein behaviour at high denaturant concentrations, for example in urea-denatured 434-repressor (using NOE measurements²⁹). At lower denaturant concentrations, for instance just above the denaturant concentration at which a particular protein unfolds, more compact structures have been reported for several proteins (e.g. ACBP³⁰ and Protein L²). These proteins show the same broad distribution of intermediate PREs as measured here for the I_b state of NPGK, but display significant native-like character. The PREs observed for protein L exhibit 'oscillations' caused by chain reversals at the two native β turn regions, while structure calculations based on the ACBP data show both native-like and non-native-like interactions⁶. The ACBP structures calculated vary greatly in terms of compactness and could perhaps be subdivided into compact structures (I_b -like), and more diffuse structures (I_b -like) on the basis of denaturant dependence of the PREs recorded, as described here for N-PGK.

The characterisation of the denatured state of a wild-type protein under native conditions has been achieved for the U_{exch} state of DrkN SH3 domain³¹. There has not been perturbation by mutation, ionisation or truncation within the domain, yet in the absence of denaturant, U_{exch} is populated almost equally with the folded state, in slow exchange. This has allowed characterisation of U_{exch} in particular through the collection of amide-amide NOEs and PREs³¹. The majority of NOEs observed are non-native -like, but show the protein to be compact. In addition, PREs are of intermediate value throughout the sequence⁷. Of all the protein denatured states characterised, the data from the U_{exch} state most closely resemble those of the I_b state of NPGK reported here. Structure ensembles calculated on the basis of the DrkN SH3 domain U_{exch} state data show non-native helical structure, and few coherent tertiary contacts⁷, properties that would equally well give rise to the data for N-PGK. The implication is that the I_b and U_{exch} states have much in common, and represent a different class of structures to the molten globule -like or denaturant-induced states.

Role of I_b in folding.

The response of the NMR resonances to changes in denaturant concentration indicates that transitions between I_b , I_k and D all occur in less than 0.1 milliseconds. The native state of N-PGK is formed with a rate constant of $\sim 10 \text{ s}^{-1}$ ⁹. Therefore, the formation of I_b from I_k or D must have an influence on the observed folding trajectory, because it will occur spontaneously prior to folding from a chemically denatured state. I_b is therefore a kinetic folding intermediate that forms within the dead time of stopped-flow experiments. It is distinct in its level of compaction and hydration (as indicated by m-value) from the kinetic folding intermediate (termed I_k) reported previously⁹ for N-PGK. Amide and side-chain protection experiments imply a native-like arrangement of hydrogen bonds and tertiary structure in I_k ^{13,32}, and the formation of I_k has been proposed to be on the millisecond timescale. The population of I_k in the conditions used in the PRE experiments is too low to have a significant effect on the relaxation behaviour of the denatured state ensemble.

I_b , therefore, represents the first compact intermediate in the folding process of N-PGK, and results from a rapid compaction from the random coil denatured state, prior to selection of more native-like structure in I_k and then to the folded state. This behaviour fits with a range of observations reported previously for other proteins, implying it is a transition that most proteins undergo on removal of denaturant. The rapid formation of a compact denatured state containing regions of secondary structure, but no coherent tertiary structure is compatible with the non-folding mutants of cytochrome c and RnaseA^{33,34}. The mutants of both proteins rapidly undergo hydrophobic collapse, but fold no further. Other proteins have also been shown to fold through substantially non-native intermediates. For example, β -lactoglobulin³⁵ 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”³⁶. Indeed, the observation of common, ANS-binding species, observed in the folding of a range of proteins³⁷, could also be equated with the population of I_b -like states. A widespread occurrence of randomly 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^{38,39,40}.

The number of hydrophobic interactions that drive the initial collapse towards β -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 domains⁸. However, it should be noted that some non-native contacts can also promote folding, as reported in disulphide exchange dominated folding⁴¹ and in the folding of CD2.d1 and fynSH3^{42,43}. Also, it is intriguing that molecular dynamics simulations of the folding of villin headpiece⁴⁴ 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 folding¹⁴.

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 Ib. 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^{7,30,31,33,34}. Hence, it is likely that these less structured molten globules, like the Ib 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 Ib 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 from Toronto Research Chemicals (Canada). All other chemicals were AnaLAR grade, obtained from BDH.

Standard buffer: All solutions contain 20mM Tris, 20mM BisTris, 0.5mM EDTA, 3mM Sodium Azide, pH 6.0.

Production of ¹⁵N-labelled wild-type and mutant N-PGK: The expression vector for wild-type N-PGK comprising residues 1-174 of gsPGK has been described previously¹³. 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 ¹⁵N-ammonium chloride as the sole nitrogen source, and expression was induced by addition of 1mM IPTG once an OD₆₀₀ of 0.8 was reached, followed by incubation overnight. The expressed protein was purified as previously described¹³.

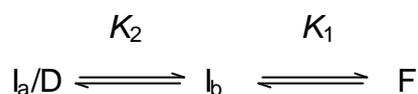
Preparation of NMR samples and MTSL labelling: 20 mg protein in 4M GuHCl, 1mM DTT was mixed with an MTSL stock (5% w/v in methanol) to give a final 5:1 excess of MTSL over the total thiol concentration. Samples were extensively buffer exchanged into 1.3 M GuHCl and concentrated to ca. 1 mM. TSP was added to 1mM and D₂O was added to 10% v/v. The resulting solution was divided into two 200µl samples, one of which was reduced with 5 protein equivalents of sodium ascorbate.

Collection of NMR spectra: Gradient selected, sensitivity enhanced ¹⁵N-HSQC spectra were collected at 25°C on a Bruker Avance 600MHz spectrometer equipped with a cryoprobe. Samples were in 3mm diameter tubes, in order to reduce sensitivity losses in the very lossy high GuHCl samples. Spectra were acquired with acquisition times of 133ms (direct dimension) and 150ms (indirect dimension). An exponential line broadening of 7Hz 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 TSP. Intensities (measured as peak-heights) were corrected for sample dilution.

Assignment of the amide resonances at 3.6M GuHCl was achieved by comparison with the previously assigned wild-type domain⁸. 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 crosspeaks from residues close in sequence to the MTSL label. These crosspeaks 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 1.3M 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 Figure 3 for all residues where overlap between peaks was minimal over a sufficient region of the titration to accurately define the parameters R_P^{lb} and $R_P^{la/D}$

Definition of equilibrium constants and m-values: The model in Figure 1 is simplified for the purpose of this analysis to the following three state pathway:



where the equilibrium constants K_1 and K_2 have associated m -values, m_1 (defined as $d\{\ln(K_1)\}/d[\text{GuHCl}]$) and m_2 (defined as $d\{\ln(K_2)\}/d[\text{GuHCl}]$) 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 Equation 1:

$$\langle \Delta d_{\text{obs}} \rangle = \frac{\langle \Delta d_{I_b-I_{a/D}} \rangle \cdot K_2 + d \cdot [\text{GuHCl}]}{K_2 + 1} \quad \text{Equation 1}$$

using:

$$K_2 = K_{2(w)} \cdot \exp(-m_2 \cdot [\text{GuHCl}]) \quad \text{Equation 2}$$

where $\langle \Delta d_{\text{obs}} \rangle$ is the mean value (averaged over all residues) of the change in chemical shift relative to the values observed at 3.6M GuHCl; $\langle \Delta d_{I_b-I_{a/D}} \rangle$ 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 (REF⁸); $K_{2(w)}$ is the value of K_2 at 0M GuHCl (*i.e.* in water). A jack-knife 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 5 parameters: the apparent transverse proton relaxation rate in the absence of paramagnetism (including a contribution due to apodisation of the FID), $R_{2,obs}$; the paramagnetic contributions to the proton transverse relaxation rate for each state, R_P^{lb} and $R_P^{la/D}$; 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 negligible¹). $K_{2(w)}$ and m_2 have been established 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 line-widths in the reduced sample. The other two parameters (R_P^{lb} and $R_P^{la/D}$) were determined for each residue by non-linear least squares fitting methods from the intensity ratios (I_{para}/I_{dia}), using the following equation:

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

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

$$R_{P,obs} = \frac{R_P^{la/D} + K_2 \cdot R_P^{lb}}{1 + K_2} \quad \text{Equation 4}$$

and K_2 defined as in Equation 2. Values of $R_{P,obs}$ were not measured directly from the spectra for use in the fitting procedure, as the definition of broader line-widths 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 line-widths of the oxidised sample for sufficiently well resolved and intense crosspeaks. The error estimates for the values of $R_P^{la/D}$ and R_P^{lb} were derived from the covariance matrix outputted by the non-linear least squares algorithm.

Relating paramagnetic enhancements to distances: The paramagnetic contribution to transverse relaxation⁴⁵ has an inverse sixth power relationship with electron-proton distance:

$$R_P = \gamma \cdot r^{-6} \quad \text{where} \quad \gamma = \gamma \cdot t_c \cdot \left(4 + \frac{3}{1 + O^2 t_c^2} \right) \quad \text{Equation 5}$$

where γ is $1.23 \times 10^{16} \text{ \AA}^6 \cdot \text{s}^{-2}$, for the interaction between a single electron and a proton, O is the proton Larmor frequency, t_c is the correlation time for the electron-nuclear dipole-dipole interaction, and was set to 4ns for I_{D^1} , and to 12ns 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 (REF⁴⁶) 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 (2 steps of 5 ps) in the absence of attractive forces using XPLOR⁴⁷. 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 Equation 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. $R_{2,obs}$ set as 56 s^{-1} as estimated from the line-widths in the ¹⁵N-HSQC spectra in order to calculate the intensity ratios in Figure 4 from Equation 3.

Simulation of a compact random ensemble. A compact sphere was modelled by using the crystal structure of N-PGK⁴⁸ as a limiting compact structure, and taking the mean calculated R_P values (on the basis of Equation 5) for 10000 random selections of residue pairs. Pairs of residues with sequence separations less than 10 were not included in the average, to prevent the increased relaxation due to covalently defined short distances dominating the derived values. Multiplication of the distances by a factor of 1.2 prior to calculation of R_P produced a mean value close to the mean fitted values for R_P^{lb} .

Acknowledgements

This work was supported by the BBSRC.

References.

- 1 Gillespie, JR & 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, ML, Alm, EJ & Baker, D (2000). NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**, 1341-1351.
- 3 Lietzow MA, Jamin M, Dyson HJ & Wright PE (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 CM, Vendruscolo M, Poulsen FM (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 JA, Neale C, Jack FE, Choy WY, Lee AY, Crowhurst KA & Forman-Kay JD (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, MAC, Jelinska, C, Syson, K, Cliff, Mj, Splevins, A, Alizadeh, T, Hounslow, AM, Staniforth, RA, Clarke, AR, Craven, CJ and Waltho, JP (2006). The Denatured State under Native Conditions: A Non-native-like Collapsed State of N-PGK. *J. Mol. Biol.* **357**, 365-372.
- 9 Parker MJ, Spencer J & Clarke AR (1995). An integrated kinetic-analysis of intermediates and transition-states in protein-folding reactions. *J Mol Biol* **253**, 771-786.
- 10 Mittag T & Forman-Kay JD (2007). Atomic-level characterization of disordered protein ensembles. *Curr. Op. Struct. Biol.* **17**, 3-14.
- 11 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.
- 12 Iwahara J & Clore GM (2006). Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**, 1227-1230.
- 13 Cliff MJ, Alizadeh T, Jelinska C, Craven CJ, Staniforth RA & Waltho JP (2006). A thiol labelling competition experiment as a probe for sidechain packing in the kinetic folding intermediate of NPGK. *J. Mol. Biol.* **364**, 810-823.
- 14 Dill KA, Bromberg S, Yue KZ, Fiebig KM, Yee DP, Thomas PD & Chan HS (1995). Principles of protein-folding - a perspective from simple exact models. *Prot. Sci.* **4**, 561-602.
- 15 Schulman BA, Kim PS, Dobson CM, Redfield C (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat. Struct. Biol.* **4**, 630-634.
- 16 Parker, M J & Marqusee S. (1999) The cooperativity of burst phase reactions explored. *J. Mol. Biol.* **293**, 1195-1210.
- 17 Kuwajima, K. (1989). The molten globule state as a clue of understanding the folding and cooperativity of globular-protein structure. *Proteins* **6**, 87-103.
- 18 Lala AK & Kaul P. (1992). Increased Exposure of Hydrophobic Surface in Molten Globule State of α -Lactalbumin *J. Biol. Chem.* **267**, 19914-19918.
- 19 Staniforth, R.A., Dean, J.L.E., Zhong, Q., Clarke, A.R., Zerovnik, E. & Waltho, J.P. (2000). The major transition state in protein folding need not involve the immobilisation of side chains. *Proc Nat. Acad. Sci. USA* **97**, 5790-5795.
- 20 Religa TL, Markson JS, Mayor U, Freund SMV & Fersht AR (2005). Solution structure of a protein denatured state and folding intermediate. *Nature* **437** (7061): 1053-1056.
- 21 Kortemme, T, Kelly, MJS, Kay, LE, 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.

- 22 Kato H , Feng HQ & Bai YW (2007).The folding pathway of T4 lysozyme: The high-resolution structure and folding of a hidden intermediate. *J. Mol. Biol.* **365**, 870-880.
- 23 Eliezer, D., Yao, J., Dyson, H.J., and Wright, P.E. (1998). Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat. Struct. Biol.* **18**, 148 – 155.
- 24 Millett, IA, Doniach, S & Plaxco KW (2002). Towards a Taxonomy of the Denatured State: Small Angle Scattering Studies of Unfolded Proteins. *Adv Prot. Hem.* **62** 241-262.
- 25 Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA & Smith LJ (1999). Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* **38**, 16424-16431.
- 26 Louhivuori, M., Paakkonen, K., Fredriksson, K., Permi, P., Lounila, J. & Annala, A. (2003). On the origin of residual dipolar couplings from denatured proteins. *J. Am. Chem. Soc.* **125**, 15647–15650.
- 27 Jha AK, Colubri A, Freed KF & Sosnick TR (2005). Statistical coil model of the unfolded state: Resolving the reconciliation problem. *Proc. Natl. Acad. Sci. USA.* **102**, 13099-13104.
- 28 Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RWH, Blackledge M (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci.* **102**, 17002-17007.
- 29 D Neri, M Billeter, G Wider, & K Wuthrich (1992). NMR determination of residual structure in a urea-denatured protein, the 434-repressor. *Science* **257**, 1559-1563.
- 30 Teilum K, Kragelund BB & Poulsen FM (2002). Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. *J. Mol. Biol.* **324**, 349-357.
- 31 Mok, YK, Kay, CM, Kay, LE & 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.
- 32 Hosszu LLP, Craven CJ, Spencer J, Parker MJ, Clarke AR, Kelly M & Waltho JP (1997). Is the structure of the N-domain of phosphoglycerate kinase affected by isolation from the intact molecule? *Biochemistry* **36**, 333-340.
- 33 Qi, PX, Sosnick, TR & Englander, SW (1998). The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nat. Struct. Biol.* **5**, 882-884.
- 34 Sosnick TR, Schtilerman MD, Mayne L & Englander SW (1997).Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl Acad. Sci. USA* **94**, 8545-8550.
- 35 Kuwata K, Shastry R, Cheng H, Hoshino M, Batt CA, Goto Y & Roder H. (2001) Structural and kinetic characterization of early folding events in β -lactoglobulin. *Nat. Struct. Biol.* **8**, 151-155.
- 36 Uversky,VN & Ptitsyn OB (1996) Further evidence on the equilibrium “pre-molten globule state”: Four-state guanidium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* **255**, 215-228.
- 37 Ptitsyn OB, Pain RH, Semisotnov GV, Zerovnik E & Razgulyaev OI (1990). Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett.* **262**, 20-24.
- 38 Parker MJ, Session, RB, Badcoe IG & Clarke AR (1996).The development of tertiary interactions during the folding of a large protein. *Folding & Design* **1**,

145-156

- 39 Parker MJ & Marqusee S (2001) A kinetic folding intermediate probed by native state hydrogen exchange. *J. Mol. Biol.* **305**, 593-602
- 40 Parker MJ & Marqusee S (2000) A statistical appraisal of native state hydrogen exchange data: Evidence for a burst phase continuum? *J. Mol. Biol.* **300**, 1361-1375
- 41 Creighton (1977). Energetics of Folding and unfolding of pancreatic trypsin inhibitor. *J. Mol. Biol.* **113** 295-312.
- 42 Mason JM, Cliff MJ, Sessions RB & Clarke AR (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.
- 43 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.
- 44 Jayachandran G, Vishal V & Pande VL (2006). Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece. *J. Chem. Phys.* **124**, 164902.
- 45 Kosen, PA (1989) Spin labelling of proteins. *Methods Enzymol.* **177**, 86-121
- 46 Gibbs N, Clarke AR & Sessions RB (2001). *Ab initio* protein structure prediction using physicochemical potentials and a simplified off-lattice model, *Proteins: Struct Funct. Genet.* **43**, 186–202.
- 47 Brunger AT (1992) *XPLOR 3.1 Manual*, Yale University Press, New Haven.
- 48 Davies GJ, Gamblin SJ, Littlechild JA, Watson HC (1993). The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. *Proteins* **15**, 283-289.

Figure Legends

Figure 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 protein⁴⁸ (pdb accession code 1PHP). I_k has some native structure, on the basis of hydrogen exchange protection data³². I_b is compact, with some helical content, and I_a is expanded with some helical content⁸. D has no coherent structure. b) The equilibrium populations were calculated on the basis of previously presented data⁸. I_a and D are treated as one state for simplicity, because they are not discrete 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.

Figure 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 5 spin-label variants.* a) The denaturant concentration dependence of the NMR signal intensities in ¹⁵N-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 Materials and Methods section. A variety of behaviours is displayed (see 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 solid line indicates the best fit parameters to Equation 3.

Figure 3. *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 Equation 3 to data like that shown in Figure 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 Equation 5 and the crystal structure of the full-length protein (pdb 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.

Figure 4. *Relationship of l_b/D to a random coil.* The relationship between paramagnetic relaxation enhancement and sequence separation from the spin-label is shown for a freely jointed, random walk polymer with unit size 3.5 Å (calculated, solid line), the simulated ensemble of random coil polypeptides based on the N-PGK sequence (see Methods, crosses) and the observed data at 3.6M GuHCl when the spin-label is at position 18 (open circles) and 80 (filled circles).

Figure 5. *Representation of the denatured states of N-PGK relative to the folded state.* Each state (l_b/D , l_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.

Table 1. *Equilibrium constants and m-values for the I_b -to- I_a/D transition. The values in brackets are the mean \pm one standard deviation as estimated by the jackknife procedure (see Methods). In the case of $K_{2(w)}$, the standard deviations were calculated for $\log(K_{2(w)})$.*

| 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) |

FIGURE 1

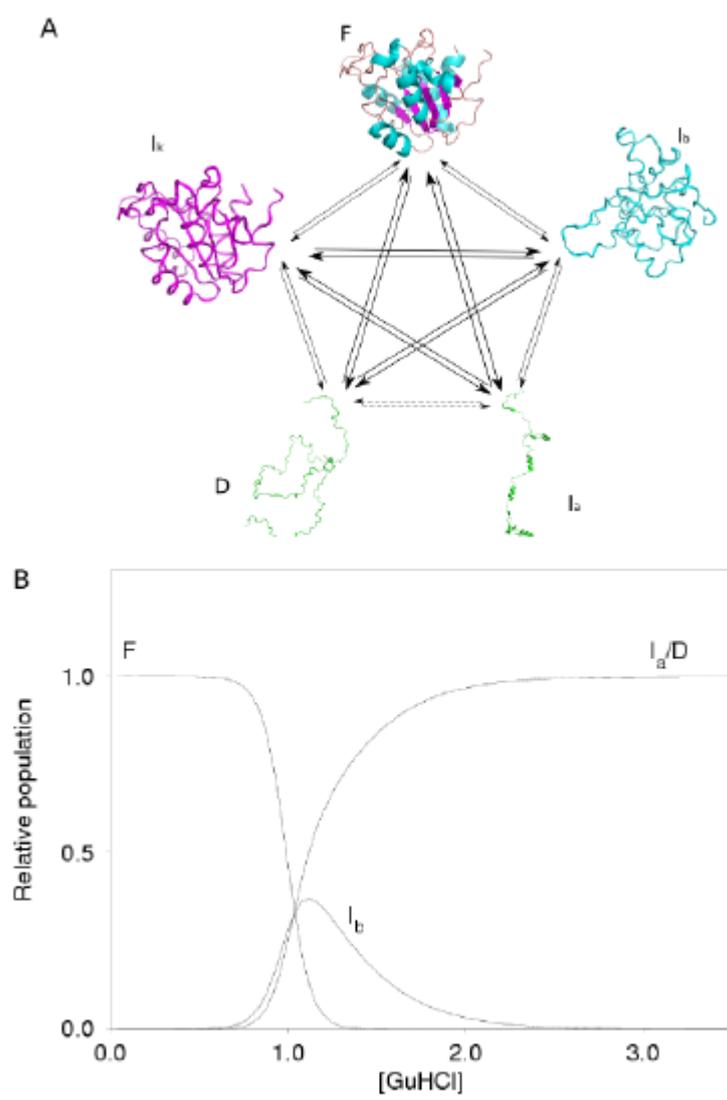


FIGURE 2

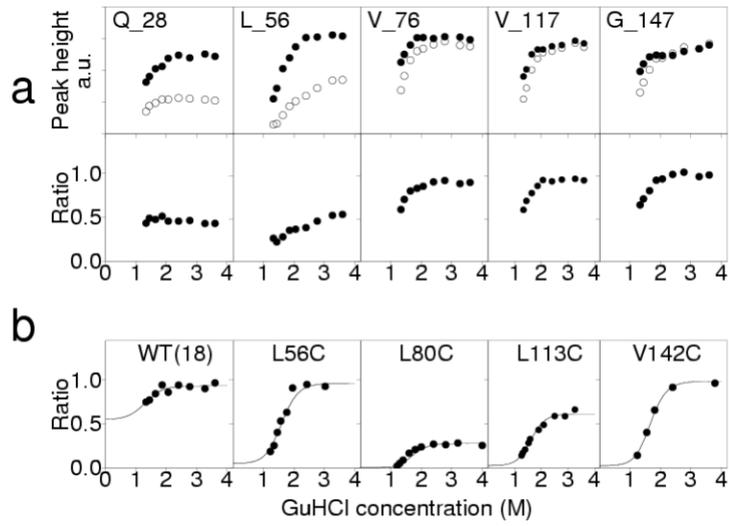


FIGURE 3

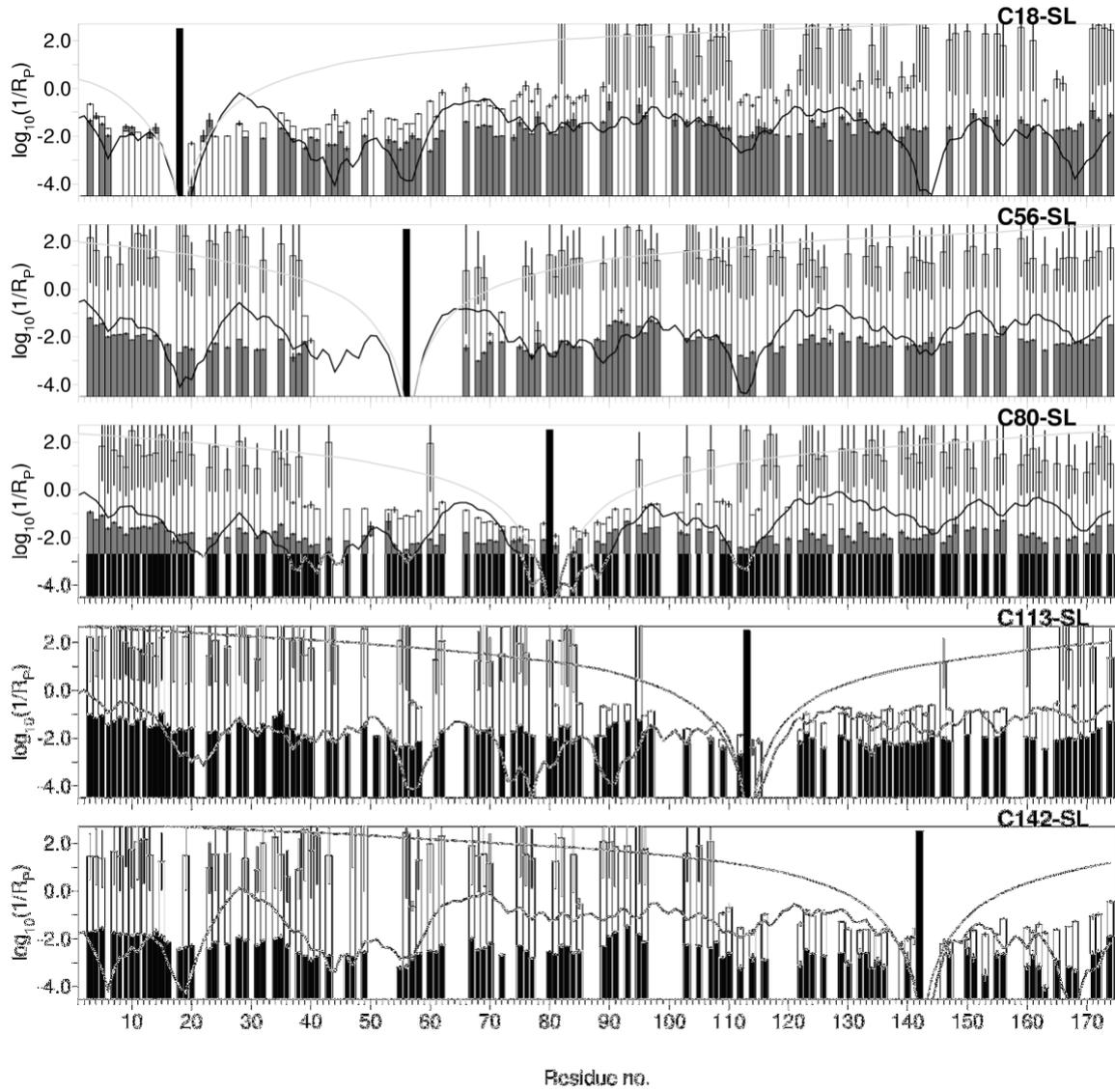


FIGURE 4

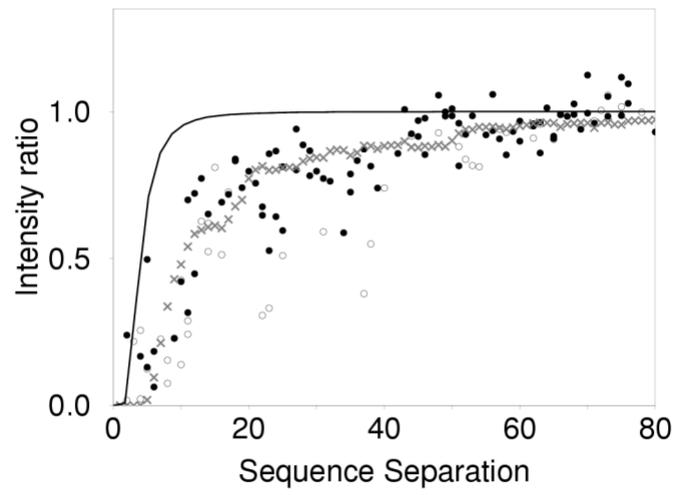


FIGURE 5

