Is the Structure of the N-Domain of Phosphoglycerate Kinase Affected by Isolation from the Intact Molecule?†

Laszlo L. P. Hosszu,‡ C. Jeremy Craven,‡ James Spencer,§ Martin J. Parker,§ Anthony R. Clarke,§ Mark Kelly,|| and Jonathan P. Waltho*,‡

Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield S10 2TN, U.K., Molecular Recognition Centre and Department of Biochemistry, University of Bristol, School of Medical Sciences, University Walk, Bristol BS8 1TD, U.K., and European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelburg, Germany

Received July 19, 1996; Revised Manuscript Received October 11, 1996⊗

ABSTRACT: The structural integrity of the isolated N-domain (residues 1−174) of *Bacillus stearothermophilus* 3-phosphoglycerate kinase (PGK) has been investigated using heteronuclear NMR spectroscopy. Analysis of ^{13}C chemical shifts, amide protection, and comparison of observed and expected sequential NOE intensities calculated from the crystal structure of the domain in the intact protein indicate that the secondary structure of the isolated domain is unchanged from that found in the intact molecule. Markedly shifted ^{1}H resonances, amide protection, and long-range NOEs indicate that the tertiary structure is similarly unaffected. These results are confirmed by an excellent agreement (standard deviation 0.28 ppm) between observed H^{α} chemical shifts and those calculated from the high-resolution (1.6 Å) crystal structure of intact PGK [Davies *et al.* (1994) *Acta Crystallogr. D50*, 202−209]. The only region perturbed by loss of interactions with the C-domain is a small portion of the substrate-binding site (residues 148−152) whose amide protons are poorly protected from solvent. These results provide a structural basis for the analysis of the folding of the domains of PGK as isolated units and within the intact molecule [Parker *et al.* (1996) *Biochemistry* (in press)] and contrast with the notion that the native tertiary fold of the N-domain of PGK requires the whole polypeptide chain, including the entire C-domain [Mas *et al.* (1995) *Biochemistry 34*, 7931−7940]. Assignments of backbone ^{13}C, ^{15}N, H^{N}, and H^{α} resonances are provided.

3-Phosphoglycerate kinase (PGK) is a monomeric 45 kDa protein which has been used extensively as a model to investigate interdomain interactions during folding (Parker *et al.*, 1996a) and within the native state (Ritco-Vonsovici *et al.*, 1995; Mas *et al.*, 1995). The crystal structures of horse muscle (Banks *et al.*, 1979), yeast (Watson *et al.*, 1982), pig muscle (Harlos *et al.*, 1992) and *Bacillus stearothermophilus* (Davies *et al.*, 1994) PGKs are very similar. The protein consists of two domains, referred to as the N- and C-domains, which are of approximately the same size and are separated by a deep cleft. These domains are structurally similar, each consisting of a core of a six-stranded parallel β-sheet surrounded by α-helices. There are two major links between the domains, comprising α-helix V which runs from the N-domain through to the C-domain and the C-terminal α-helices XIII and XIV which loop back onto the N-domain (Figure 1). Examination of the position of totally conserved residues of PGK from various species shows that this α-helical interdomain region is highly conserved (Joâo & Williams, 1993). It has been suggested that this region has a role in the hinge-bending rearrangement of the enzyme that is proposed to take place during catalysis (Banks *et al.*, 1979) and, furthermore, that this region has a role in the late events on the refolding pathway of the enzyme (Ballery *et al.*, 1993).

The role of interdomain interactions in protein stability has been investigated by comparison of the thermodynamic behavior of the domains as isolated units and as part of the intact molecule (Missiakas *et al.*, 1990; Parker *et al.*, 1996a). However, there is ambiguity concerning the structural integrity of the two domains when interdomain interactions have been removed. For the yeast enzyme (yPGK), genetically engineered isolated N- and C-domains, yPGK[1−184] and yPGK[187−413], have been produced. Studies on these domains led to the proposal that these species folded to "quasi-native" structures that were similar to the native folds (Minard *et al.*, 1989). The differences from the native properties that were observed, such as decreased affinity for substrates, were significantly more pronounced in the N-domain, which was proposed to have greater conformational flexibility compared with the native protein (Fairbrother *et*

† We thank the BBSRC for project funding and financial support to L.L.P.H., C.J.C., J.S. and M.J.P. A.R.C. and J.P.W. are fellows of the Lister Institute of Preventative Medicine. The Krebs Institute is a BBSRC Biomolecular Science Centre for Molecular Recognition.

* Author to whom correspondence should be addressed.
‡ University of Sheffield.
§ University of Bristol.
|| European Molecular Biology Laboratory.

⊗ Abstract published in *Advance ACS Abstracts,* December 1, 1996.

[1] Abbreviations: 1D/2D/3D, one/two/three dimensional; 3-PGA, 3-phosphoglycerate; ANS, 1-anilinonaphthalene-8-sulfonic acid; bsPGK, PGK from *Bacillus stearothermophilus*; bsPGK[1−174], N-domain of *B. stearothermophilus* PGK; DCI, 3,4-dichloroisocoumarin; DQF-COSY, double-quantum-filtered correlation spectroscopy; DTT, dithiothreitol; E-64, *trans*-epoxysuccinyl-L-leucylamido-(4-guanidino)butane; GuHCl, guanidine hydrochloride; HMQC, heteronuclear multiple-quantum coherence; HSQC, heteronuclear single-quantum coherence; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PGK, 3-phosphoglycerate kinase; rmsd, root-mean-square deviation; TEA, triethanolamine; TOCSY, total correlation spectroscopy; TPPI, time-proportional phase incrementation; TSP, 3-(trimethylsilyl)-[2,2,3,3-$^{2}H_4$]propionate; yPGK, PGK from *Saccharomyces cerevisiae*.

S0006-2960(96)01784-9 CCC: $14.00

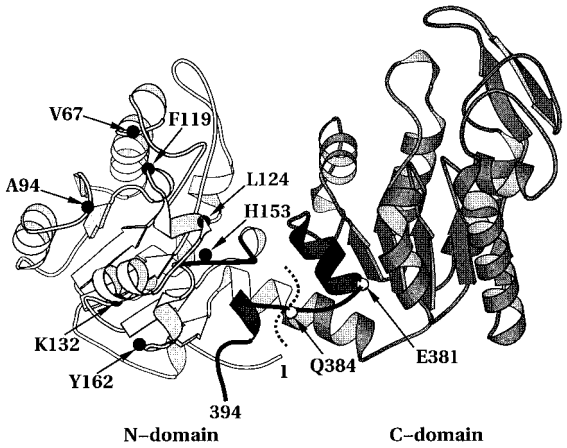

FIGURE 1: Three-dimensional structure of *B. stearothermophilus* 3-PGK (Davies *et al.*, 1994) illustrated using MOLSCRIPT (Kraulis, 1991). The isolated N-domain studied here (bsPGK[1−174]) is unshaded; the dotted line in the middle of helix V indicates the truncation point at residue 175. Residue A94 which is homologous to the single cysteine in the N-domain of yPGK is shown. Other residues highlighted in the N-domain are those used to provide evidence of strong tertiary interactions in bsPGK[1−174], as described in the Results section. The C-terminal peptide and a portion of the 3-PGA binding loop which interacts closely with it are shaded in black. Residues 381 and 384 which are homologous to the C-terminal residues in the truncation mutants produced by Mas *et al.* (1995) and Ritco-Vonsovici *et al.* (1995), respectively, are also indicated.

al. 1989; Minard *et al.* 1989). In contrast, the results of Mas *et al.* (1995), who studied a C-terminal truncation mutant yPGK[1−400], implied that loss of the last 15 residues from the full-length protein resulted in a change in structure of both domains, with the N-domain exhibiting characteristics of a molten-globule state. Specifically, the N-domain was judged to have retained most of its secondary structure while losing tertiary interactions, and thus it was proposed that the native tertiary fold of yPGK, and in particular the N-domain, requires the whole polypeptide chain. Conversely, an independent study of a slightly longer form, yPGK[1−403], indicated that this mutant exhibited structural properties that were very similar, at least qualitatively, to those of the isolated domains (Ritco-Vonsovici *et al.*, 1995). These workers proposed that the C-terminal region was not necessary for the tertiary fold of the domains but was required to "lock" the protein into its active conformation. Consequently, there is some doubt as to the nature of the quasi-native structures of the domains of the yeast enzyme, particularly the N-domain, in both their isolated form and as part of the C-terminally truncated mutants.

Recently, we have studied the refolding of the homologous protein from *B. stearothermophilus* (bsPGK). By comparison of the kinetic and equilibrium energetics of folding of the two domains in the intact molecule, and as isolated domains, bsPGK[1−174] and bsPGK[186−394], it has been suggested that interdomain contacts only become established late in the folding pathway and that they contribute relatively little to the stability of the folded ground state (Parker *et al.*, 1995, 1996a). This energetic autonomy is indicated by a close similarity in the thermodynamic properties of the intermediate and transition states for folding of the intact protein and the isolated domains. It is important in comparing the energetics of folding of intact bsPGK, and the isolated domains, that the nature of the "folded" state of each is known. In particular, it is necessary to determine whether the structure of the domains is essentially the same as in the crystal structure of the intact protein, whether the domain is folded but significantly different from the crystal structure, or whether it is partially structured. This is especially true for the N-domain, given the ambiguity in the interpretation of its structural integrity in the genetically engineered mutants of yPGK described above (Minard *et al.*, 1989; Ritco-Vonsovici *et al.*, 1995; Mas *et al.*, 1995). In order to investigate this, we have analyzed ^{1}H and ^{13}C chemical shifts, NOEs, and amide protection of the isolated N-domain of bsPGK (bsPGK[1−174]) in solution. These measurements can be used as efficient and diagnostic probes of the secondary and tertiary structure of the protein. In addition, a comparison with the NMR data of Fairbrother *et al.* (1989) on the isolated yeast N-domain, yPGK[1−184], is made in order to gain further insight into the structural nature of the isolated yeast N-domain and the C-terminally truncated mutants produced by Mas *et al.* (1995) and Ritco-Vonsovici *et al.* (1995).

EXPERIMENTAL PROCEDURES

Sample Preparation. bsPGK[1−174] was expressed and purified essentially as previously described (Parker *et al.*, 1995) with the following modifications. First, in order to prevent proteolysis of bsPGK[1−174] during purification, 100 μM DCI, 100 μM 1,10-phenanthroline, and 20 μM E-64 were added to the sonication buffer prior to cell disruption. Second, the final gel-filtration step was replaced with ion-exchange chromatography using SP-Sephadex C-25 (Sigma). bsPGK[1−174]-containing fractions previously eluted from a Q-Sepharose column as described in Parker *et al.* (1995) were equilibrated by dialysis against 40 mM sodium acetate buffer, pH 5.5, containing 5 mM sodium azide. The combined fractions were loaded onto an SP-Sephadex C-25 column equilibrated with the same buffer. Elution of bsPGK[1−174] from the column was carried out by a single pH jump with 50 mM TEA, pH 8.5. The protein was >95% pure as judged by SDS−PAGE.

Uniformly ^{15}N-labeled protein samples were prepared using an EMBL minimal media recipe with $(^{15}NH_4)_2SO_4$ as the sole ^{15}N source. A 12 mg $^{15}N/^{13}C$-labeled sample was prepared by the same procedure but with the addition of ^{13}C-labeled algal hydrolysate as the sole ^{13}C source.

NMR samples were prepared by dialysis against 50 mM potassium phosphate buffer, containing 5 mM sodium azide, and at pH 6.6; 10% (v/v) D_2O was added as the internal lock. The final protein concentrations were 2 mM unlabeled, 3.5 mM ^{15}N-labeled, and 1.3 mM $^{15}N/^{13}C$-labeled. The ^{15}N relaxation times and amide proton linewidths indicated that the protein was monomeric at these concentrations.

For the determination of amide protons protected from solvent, a ^{15}N-labeled sample of bsPGK[1−174] was dialyzed extensively against water, lyophilized, and redissolved in 20 mM potassium phosphate and 1 mM DTT, pH 6.70 (pH uncorrected for isotope effect), in 100% D_2O. The occupancy of protons was measured from a series of $^{1}H-^{15}N$ HSQC spectra (Bodenhausen & Ruben, 1980). Acquisition of the first experiment began within 35 min of redissolving the protein sample.

The loss of ^{1}H and ^{15}N chemical shift dispersion upon denaturation was determined by titrating a ^{15}N-labeled bsPGK[1−174] sample [prepared as described above−90% (v/v) H_2O] with aliquots of 6 M GuHCl in 50 mM potassium phosphate, pH 6.6. Following each addition of GuHCl, the

sample was equilibrated for 45 min at 298 K and an $^1H-^{15}N$ HSQC spectrum recorded at the same temperature.

NMR Spectroscopy. NMR experiments were acquired on a Bruker AMX-600 spectrometer [HNCA, CBCANH, CBCA(CO)NH] and a Bruker AMX-500 spectrometer (all others). Water suppression for the HNCA, CBCANH, and CBCA(CO)NH experiments was achieved using pulsed-field gradients (Kay, 1993); in all other experiments presaturation during the relaxation delay between scans was used. Quadrature detection in the indirectly detected dimensions was accomplished using either States−TPPI (Marion *et al.*, 1989a) or TPPI (Marion & Wuthrich, 1983) methods.

The following experiments were recorded, essentially as described in the accompanying references: CT-HNCA, CT-HN(CO)CA, CT-HNCO (Grzesiek & Bax, 1992a); HN(CA)CO (Clubb *et al.*, 1992); CBCANH (Grzesiek & Bax, 1992b); CBCA(CO)NH (Grzesiek & Bax, 1992c); ^{15}N-separated TOCSY-HMQC, mixing time 60 ms; ^{15}N-separated NOESY-HMQC, mixing time 100 ms (Marion *et al.*, 1989b); 2D DQF-COSY (Rance *et al.*, 1983). A table of acquisition times of the multidimensional NMR experiments used to assign bsPGK[1−174] is available as Supporting Information. All experiments were run at 298 K apart from the HNCO and HN(CA)CO experiments (318 K) and the ^{15}N-separated TOCSY-HMQC and NOESY-HMQC experiments (both 298 and 318 K).

NMR Data Processing and Analysis. NMR data were processed and analyzed on Silicon Graphics workstations using Felix 95.0 software (Biosym Technologies, Inc., San Diego). Proton chemical shifts were referenced to internal TSP. ^{15}N and ^{13}C chemical shifts were calculated relative to TSP, using the gyromagnetic ratios of ^{15}N, ^{13}C, and ^1H. Residual solvent resonances were removed by applying a time-domain convolution filter as described in Waltho and Cavanagh (1993) and Craven and Waltho (1995).

Assignment of Backbone ^1H and ^{13}C Resonances. The NMR signals of the $^{13}C^\alpha$, ^{15}N, and $^1H^N$ nuclei were correlated using the HNCA and HN(CO)CA experiments. Degeneracy in the C^α resonances precluded complete assignment solely through (i) and ($i-1$) C^α correlations. The two C^β experiments were of relatively low sensitivity, reflecting the protein's correlation time; the CBCA(CO)NH yielded 40% of the possible interresidue C^β correlations and the CBCANH 20% of the possible intraresidue C^β correlations. However, these C^β experiments did provide valuable information on amino acid type through analysis of C^β chemical shifts. The HNCO and HN(CA)CO experiments were of sufficient sensitivity to resolve virtually all of the C^α (i) and ($i-1$) correlation ambiguity (Figure 2). Connectivities in the triple-resonance experiments were independently investigated through the use of strong $d_{NN}(i,i+1)$ NOE connectivities in α-helical stretches of the protein and weak $d_{NN}(i,i+1)$ connectivities in some more extended regions. Using these connectivities, coupled with characteristic C^α and C^β chemical shifts and side chain information from the ^{15}N-separated TOCSY-HMQC experiments, it was possible to link up unambiguous sequence-specific stretches of the protein to give the full backbone and significant side chain resonance assignment (Figure 2).

Almost all non-proline H^α resonances were identified in the 2D DQF-COSY and 3D TOCSY-HMQC spectra. The H^α resonances of the proline residues P25, P87, and P164 were assigned using very large $d_{\alpha N}(i,i+1)$ NOEs observed in the ^{15}N-separated NOESY-HMQC experiments.

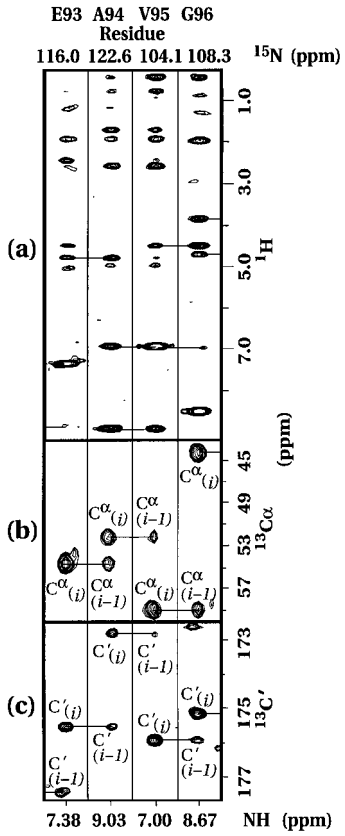

FIGURE 2: ^{15}N slices through (a) 3D ^{15}N-separated NOESY-HMQC, (b) 3D HNCA, and (c) 3D HN(CA)CO spectra, arranged to illustrate the sequential assignment process. Slices are taken at the $^1H/^{15}N$ (F_3/F_2) frequencies of the backbone amide resonance of each residue. Sequential d_{NN} and $d_{\alpha N}$ NOEs in (a) are marked by solid horizontal lines connecting peaks in adjacent strips. Intra- (i) and inter- ($i-1$) residue C^α and C′ correlations, panels b and c, respectively, are also indicated.

Chemical Shift Calculations. The CSI program of Wishart and Sykes (1994) was used to calculate the chemical shift index profile of bsPGK[1−174] using the assigned H^α, C^α, C^β, and C′ chemical shifts. The *total* program of Williamson and Asakura (1993) was used to calculate proton chemical shifts for bsPGK[1−174], using residues 1−174 of the bsPGK crystal structure (Davies *et al.*, 1994). *Total* calculates the proton chemical shifts from atomic coordinates on the basis of an empirical model including separate terms for ring current shifts, CO and CN anisotropy, and electric field. All backbone and well-resolved methyl resonances markedly shifted by ring-current shifts (>−0.5 ppm), according to the *total* chemical shift calculation, were examined for their observed chemical shifts.

Comparison of Observed and Predicted Sequential $d_{\alpha N}$ and d_{NN} NOE Intensities. NOE intensities were measured as peak heights in the ^{15}N-separated NOESY-HMQC spectra. $d_{\alpha N}$ and d_{NN} distances ≤3.5 Å in the N-domain were calculated from the bsPGK crystal structure (Davies *et al.*, 1994), converted to NOE intensity using the simple inverse sixth power relationship, and calibrated using the general envelope of the observed NOEs.

RESULTS

Spectral Dispersion of bsPGK[1−174]. The ^1H and the ^{15}N chemical shifts of bsPGK[1−174] are indicative of a protein containing a well-defined tertiary structure. To illustrate this, the 2D $^1H-^{15}N$ HSQC spectrum of bsPGK-[1−174] in 50 mM potassium phosphate (pH 6.6 and 298

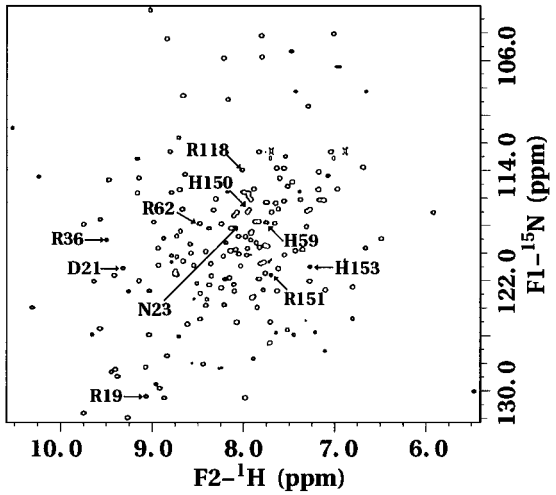

FIGURE 3: 2D $^1H-^{15}N$ HSQC spectrum of uniformly ^{15}N-labeled bsPGK[1−174]. The spectrum was recorded at 298 K on a 3.5 mM sample in 50 mM potassium phosphate, 5 mM sodium azide, pH 6.6, and 90% H_2O/10% D_2O. Cross-peaks arising from the "basic patch" 3-PGA binding site residues are shown.

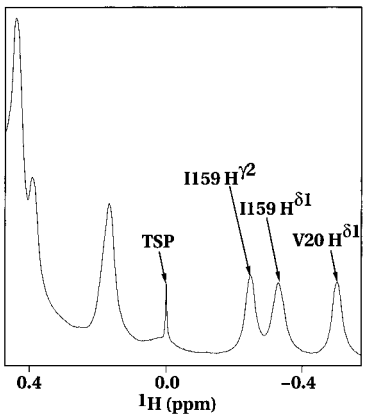

FIGURE 4: Upfield region of the 1D ^1H NMR spectrum of 3.5 mM ^{15}N-labeled bsPGK[1−174], illustrating the presence of markedly shifted methyl resonances. The peaks at −0.31 and −0.22 ppm originate from the $H^{\delta 1}$ and $H^{\gamma 2}$ groups, respectively, of I159, and the peak at −0.50 ppm originates from the the $H^{\delta 1}$ group of V20. The sharp peak at 0 ppm is the TSP resonance used for referencing spectra. Sample conditions were identical to those in Figure 3.

K) is shown in Figure 3. Overall, both ^1H and ^{15}N resonances show very high chemical shift dispersion for bsPGK[1−174] in the absence of denaturant, whereas, in contrast, the corresponding spectrum in 1.35 M GuHCl exhibits little ^1H chemical shift dispersion (data not shown). Under the latter conditions, the unfolded state of bsPGK-[1−174] is the overwhelmingly populated species (Parker *et al.*, 1995). In Figure 3 the amide resonances of residues involved in 3-PGA binding are shown, the so-called "basic-patch" residues (Banks *et al.*, 1979). In the isolated yeast N-domain, yPGK[1−184], undetectable substrate binding and pK_a measurements indicated that the 3-PGA binding site was disrupted (Fairbrother *et al.*, 1989). In bsPGK[1−174] 3-PGA also fails to bind (data not shown); however, several of these substrate-binding residues exhibit ^1H chemical shifts far from random coil values (especially R19, D21, and R36), indicating that a significant proportion of the 3-PGA binding site is still highly structured. The 1D ^1H NMR spectrum of bsPGK[1−174] (Figure 4) shows several markedly upfield-shifted methyl resonances, indicating that the isolated domain contains close-packed tertiary structure, including three resonances with chemical shifts of less than 0 ppm (identified in Figure 4). (The specific tertiary interactions that give rise to these shifted resonances are described later.) In addition, the aromatic region of the ^1H COSY spectrum at 298 K (data not shown) shows the excellent dispersion characteristic of a folded protein. This dispersion is maintained at 318 K (45 °C) for the duration of the 3D heteronuclear NMR experiments (3−4 days), indicating that the fold is stable over this temperature range.

Secondary Structure of the Isolated Domain. Figures 5 and 6 summarize the NMR data that characterizes the secondary and tertiary structure of bsPGK[1−174]. The intensities of the observed $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ sequential NOEs in Figure 5 show very good agreement with the predicted intensities calculated from the crystal structure of bsPGK. This provides strong evidence that the α-helices and β-sheet of the isolated domain are intact and that little change in secondary structure has occurred upon truncation. This conclusion is supported by secondary structure prediction based on the CSI analysis of observed H^α, C^α, C^β, and C' chemical shifts (Wishart & Sykes, 1994), which also shows very good agreement with the crystal structure. For example, the NOE and CSI data clearly show that helix I (residues 35−49) is divided into two parts. This is the result of the presence of a highly conserved proline residue (P42) which "kinks" the helix, altering its direction to allow certain residues, including R36, to interact with 3-PGA in the basic patch (Davies *et al.*, 1994). This appears to be unaltered in bsPGK[1−174].

The local structural integrity of the β-sheet is confirmed by the presence of downfield-shifted H^α resonances that show large H^α/N^H *J*-couplings and also by the very good agreement between the observed and calculated NOE intensities. Of particular note are the medium-intensity sequential $d_{\alpha N}$ NOEs of the 3-PGA-binding residue H59 and residue V142 which are in regions of very strong sequential $d_{\alpha N}$ NOEs and agree closely with the calculated NOE intensities. Importantly, long-range NOEs between H^α and H^N protons are seen across a large portion of the parallel β-sheet, indicating that the organization and register of the β-sheet is unchanged (Figure 6).

An interesting example is helix V. This helix is cleaved in the middle to form bsPGK[1−174] and may be expected to be destabilized by loss of interactions with its C-terminal residues. According to the NMR data this helix has remained intact in bsPGK[1−174] and retains a close interaction with residue D144 present in the intact protein (see Figure 6).

The only region which appears to have been perturbed by the loss of interactions with the C-domain is a portion of the 3-PGA-binding site (residues 148−152). In the crystal structure of intact bsPGK, residues 146−148 form a short helical stretch. The amide protons of this region of bsPGK-[1−174] are not strongly protected from exchange with the solvent and are bleached to varying degrees by the presaturation used in the NOE experiments. At the temperature and pH used, such rapid exchange rates correspond to protection factors of the amide protons of less than 10, and hinder the observation of sequential d_{NN} NOEs. Thus, NOE evidence of structure within this region was not available. Nevertheless, the CSI analysis for this region shows no evidence for a helical segment involving residues 146−148, whereas it predicted very accurately all other helical stretches in the protein. Examination of the crystal structure shows that this loop region of the protein would be left completely solvent exposed on removal of the C-domain. It is note-

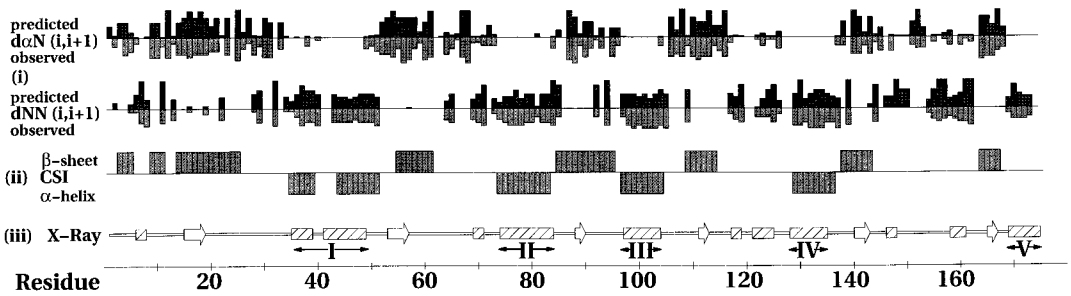

FIGURE 5: Summary of data used to determine the secondary structure of bsPGK[1−174]. (i) Observed sequential $d_{\alpha N}(i,i+1)$ and $d_{NN}(i,i+1)$ NOEs in bsPGK[1−174] compared with predicted NOE intensities calculated from the structure of the domain in intact PGK (Davies *et al.*, 1994). The height of the bars for the NOEs is proportional to the intensities of the observed and calculated NOE cross-peaks. (ii) Chemical shift index profile calculated by the program CSI (Wishart & Sykes, 1994) using assigned H^{α}, C^{α}, C^{β}, and C′ chemical shifts. (iii) Secondary structure elements of the N-domain in intact *B. stearothermophilus* PGK calculated by the *dssp* program (Kabsch & Sander, 1983) from the crystal structure of bsPGK (Davies *et al.*, 1994). α-Helices are represented by cylinders and β-sheet strands by arrows. The five main α-helices in the domain are labeled I−V.

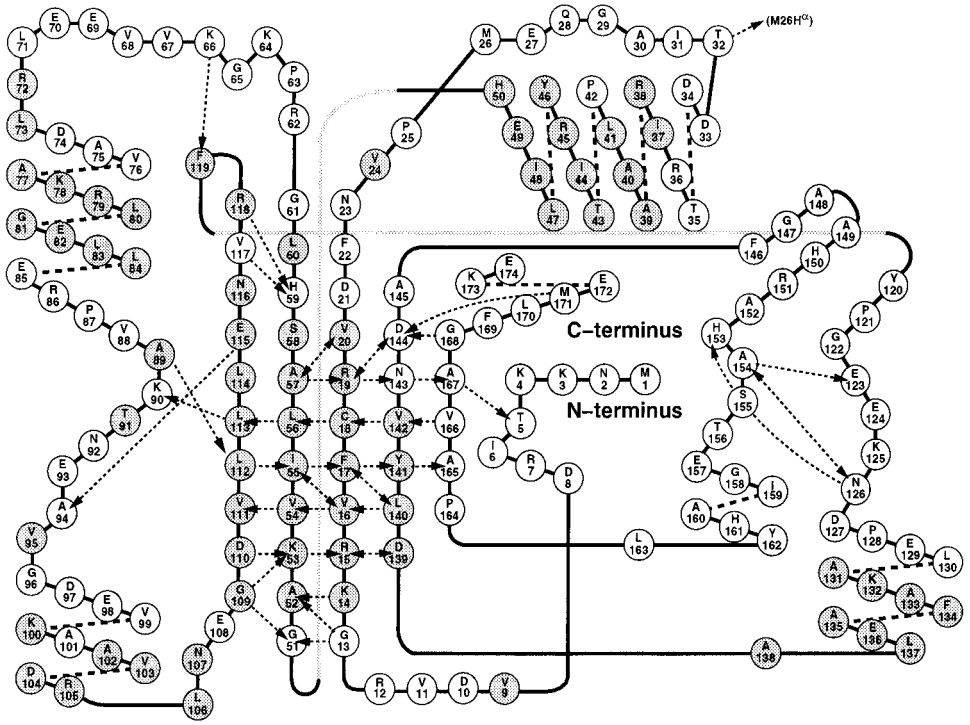

FIGURE 6: Schematic diagram of the secondary structure and amino acid sequence of bsPGK[1−174]. Residues shaded are those which showed cross-peaks in the $^1H-^{15}N$ HSQC spectrum 1.5 h after a lyophilized sample of bsPGK[1−174] was dissolved in 100% D_2O, 20 mM potassium phosphate, and 1 mM DTT, pH 6.70. Long-range ($\geq i,i+5$) $d_{\alpha N}$ and d_{NN} NOEs observed (≤ 3.5 Å) are marked with arrows. All but two of the NOEs expected from H^{α} and H^{N} protons within 3.5 Å in the crystal structure of the N-domain in intact bsPGK were observed. This distance was the limit of the sensitivity of the NOESY-HMQC experiments used. In the case of $d_{\alpha N}$ NOEs the arrow points to H^{α} of the pair.

worthy, however, that residues spatially close to this segment retain a high degree of ^1H chemical shift dispersion (see Figure 3), and sequential NOEs on either side of this region agree closely with values calculated from the crystal structure. Also, predicted long-range NOEs to the amide proton of residue A154 are present. This indicates that disruption of the secondary structure of bsPGK[1−174], if any, is localized to this exposed loop region (Figure 6).

Amide Protection. The amide-exchange data show that a significant portion of the protein is highly protected from solvent (72 out of the 167 amide protons), including the first five strands of the β-sheet and the four α-helices (Figure 6). The residues shaded in Figure 6 have protection factors exceeding 5000, in many cases substantially so. In the crystal structure of the domain in intact PGK these regions constitute the core of the protein. This indicates that the core tertiary structure of bsPGK[1−174] is extremely stable. In terms of sequence, amide protection is concentrated in the first two-thirds of the protein, with the final third showing markedly lower protection, indicating that this region may be less stable than in the native protein, most likely due to the loss of interactions with the C-domain.

Tertiary Structure of the Isolated Domain. Despite the much lower amide protection in the last third of the bsPGK-[1−174], long-range tertiary contacts expected from the crystal structure of the domain in intact PGK are observed, indicating that the native tertiary fold of the isolated N-domain is intact. For example, $d_{\alpha N}$ NOEs between the amide proton of A167 and the H^{α} of T5, and the amide protons of M171 and G168 and the H^{α} proton of D144, are observed. Together with the long-range NOEs to the basic-patch region mentioned earlier, it appears that although this portion of bsPGK[1−174] shows little or no protection, the tertiary structure found in the crystal structure is still maintained as the predominant conformer.

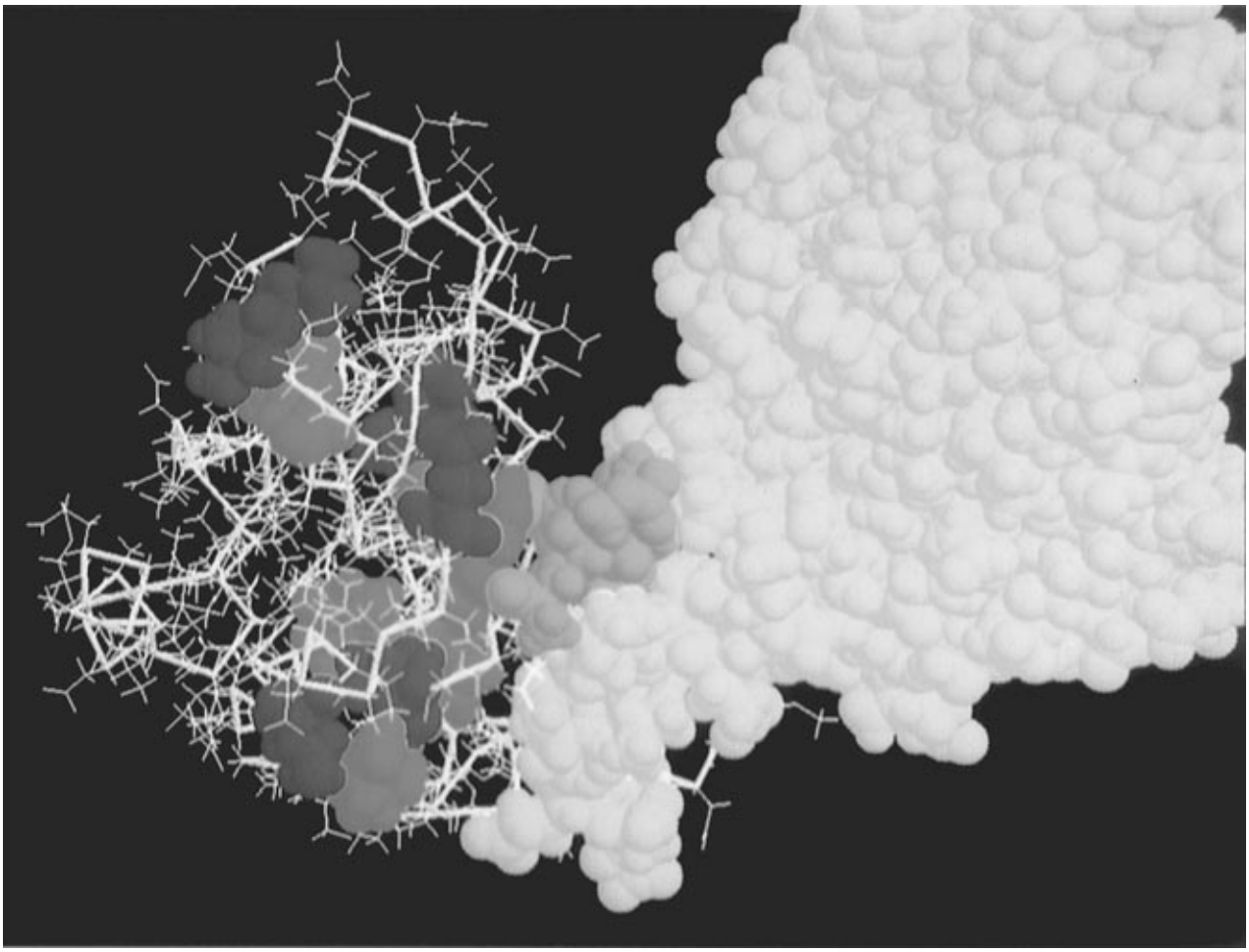

FIGURE 7: Space-filling model of the C-domain of bsPGK and residues involved in specific tertiary interactions in bsPGK[1−174], as described in the Results section. Residues in bsPGK[1−174] containing protons significantly shifted by close interactions with aromatic groups are in green, while the corresponding aromatic groups are in red. Remaining residues in the N-domain are depicted as white wire frames. The portion of the 3-PGA binding site that is perturbed by loss of the C-domain is in cyan. The diagram was prepared using Rasmol version 2.5 (R. A. Sayle, Glaxo Pharmaceuticals).

Further evidence that the tertiary structure is maintained throughout bsPGK[1−174] comes from markedly shifted methyl, H^{α}, and H^{N} protons. These shifted resonances are caused by highly specific "face-on" interactions with aromatic rings. These close tertiary interactions found in the domain in the intact protein are retained in the isolated N-domain. For example, the two most upfield-shifted resonances calculated from the structure of the domain arise from the $H^{\delta 1}$ and $H^{\gamma 2}$ groups of I159. These groups lie against the aromatic rings of F17 and Y141, respectively, in the crystal structure. These interactions are observed in bsPGK[1−174], with the two groups showing highly shifted resonances at −0.22 (I159 $H^{\delta 1}$) and −0.31 ppm (I159 $H^{\gamma 2}$) (see Figure 4). Only the resonance which arises from the interaction of V20 $H^{\delta 1}$ with the aromatic ring of F22 is more upfield shifted (−0.50 ppm). This shows that in these portions of the protein the tertiary structure is intact and the packing of the side chains is highly ordered. This is of particular interest as residue I159 is adjacent to the C-terminal peptide in the intact protein, and yet the tertiary structure is apparently not perturbed in bsPGK[1−174] (Figure 7).

Also of note are the upfield-shifted resonances of V67 H^{α} (1.67 ppm) and V68 H^{N} (5.44 ppm), which are caused by the aromatic ring of F119, and the H^{α} protons of K132 (3.04 ppm) and E124 (3.48 ppm), which are packed against the aromatic rings of Y162 and the basic-patch residue H153, respectively. These specific tertiary interactions are distributed throughout the isolated domain (see Figure 7), including regions which are in close contact with the C-terminal peptide. Overall, the rmsd between observed and calculated H^{α} chemical shifts is 0.28 ppm (Figure 8). This compares very favorably with rmsd values calculated for a wide range of intact (i.e., untruncated) fully folded proteins (Williamson *et al.*, 1995). Coupled with the good agreement of the calculated and observed NOE and ^{1}H and ^{13}C chemical shift data throughout the protein, it appears that the structure of

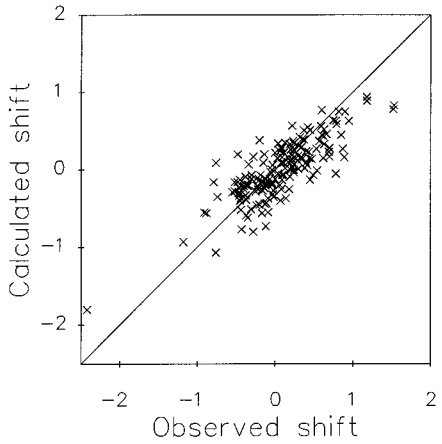

FIGURE 8: Comparison of calculated and observed H^{α} chemical shifts of bsPGK[1−174], expressed as differences from random coil values. H^{α} shifts were calculated using the *total* program of Williamson and Asakura (1993).

the isolated domain closely resembles the crystal structure of the domain in the intact protein.

DISCUSSION

Study of the complete refolding pathway of intact bsPGK reveals that both domains of the protein collapse rapidly to intermediate states in which short sequence-range interactions dominate (Parker *et al.*, 1996a,b). The establishment of long-range contacts occurs more slowly, with the rate-limiting step in the overall folding pathway being the acquisition of native structure of the C-domain in the presence of the folded N-domain. The thermodynamic properties of the intermediate and transition states of folding of the domains as isolated units and within the whole molecule are very similar. By inference, interdomain contacts are only established late in folding, suggesting a high degree of structural autonomy. However, such a conclusion can only be secure with knowledge of the nature of the folded states of the isolated domains as compared with the intact protein.

The results presented here demonstrate that the isolated N-domain, bsPGK[1−174] in 50 mM phosphate buffer at pH 6.6 and 25 °C, has essentially the same structure as that found in the crystal structure of the intact protein. The analysis of backbone ^{13}C and ^{1}H chemical shifts and sequential NOEs shows that its secondary structure is not changed in any substantial way by removal of the C-domain. The tertiary structure of the protein appears to be virtually the same as that in the native protein, as indicated by long-range NOEs, expected markedly shifted ^{1}H resonances, and the strong agreement between calculated and observed H^{α} chemical shifts. Indeed, such an agreement is remarkable given the loss of the C-domain from bsPGK[1−174] and the inherent small differences typically observed between crystal and solution structures. This strongly indicates that the protein is unchanged from the crystal structure in any significant way. The high degree of ^{1}H spectral dispersion and large amide protection factors observed ($\gg$5000; see Figure 6) are also comparable to other stable, native proteins. Importantly, the high spectral dispersion and amide protection are maintained over a wide range of temperatures (25−45 °C) and pHs (5.35−8.55; data not shown), indicating that the protein is by no means marginally stable. The only region of the protein which does appear to have been affected by the loss of the C-domain is a small portion of the 3-PGA-binding site; however, this region is left completely solvent exposed by the truncation. Whether the isolated C-domain studied by Parker *et al.* (1996a) behaves analogously awaits future investigation.

In addition to providing a solid basis for the kinetic analysis of the folding of bsPGK and the contribution of interdomain contacts within bsPGK, this study sheds light on two conflicting studies involving C-terminally truncated mutants of yPGK, yPGK[1−400] (Mas *et al.*, 1995) and yPGK[1−403] (Ritco-Vonsovici *et al.*, 1995). These mutants lack the C-terminal helix XIV which loops back from the C-domain to interact closely with the N-domain (Figures 1 and 7). According to Mas *et al.* (1995) the N-domain in yPGK[1−400] is best described as a *disordered molten globule* (Ptitsyn, 1992), which by definition lacks both rigid side chain packing and even a native-like tertiary fold. In contrast, Ritco-Vonsovici *et al.* (1995) concluded that yPGK-[1−403] folded to a conformation resembling that of the wild-type protein, with the N- and C-domains exhibiting structural properties that were very similar to their quasi-native isolated forms, yPGK[1−185] and yPGK[187−413] (Minard *et al.*, 1989; Fairbrother *et al.*, 1989).

The tertiary structure implied by the wide ^{1}H chemical shift dispersion in bsPGK[1−174] bears no resemblance to the fluctuating side-chain environment postulated for classical molten-globule states (Ptitsyn, 1992). In support of the hypothesis that yPGK[1−403] folds to a conformation similar to that of the full-length protein, the 1D ^{1}H NMR spectrum exhibited well-dispersed resonances (Ritco-Vonsovici *et al.*, 1995). The dispersion is similar to that of wild-type yPGK and characteristic of a structured protein. Indeed, given only the partial loss of intensity in the near-UV CD spectra reported for both C-terminally truncated yeast mutants, yPGK[1−400] and yPGK[1−403], compared with the full-length protein, the similarity of the nature of the N-domain properties in yPGK[1−403] and yPGK[1−184], and the wide dispersion in the ^{1}H NMR spectrum of yPGK[1−184] (Fairbrother *et al.*, 1989), it seems likely that the isolated N-domain of the yeast enzyme is also a fully folded protein under the conditions of the yPGK studies. The dispersion observed in the ^{1}H NMR spectra of yPGK[1−184] and yPGK[1−403], as in bsPGK[1−174], is clearly inconsistent with the idea of the N-domain of these proteins being a disordered molten globule, lacking both rigid side chain packing and a native-like tertiary fold.

This conclusion is supported by the amide exchange data. Although all amide protons in yPGK[1−184] were found to exchange for deuterons in the solvent within 48 h at pH 7.1 and 4 °C (Fairbrother *et al.*, 1989), and this was taken as evidence for an increase in flexibility, for amide protons to persist on this time scale protection factors approaching 1×10^5 are required. This is markedly different from the very low protection factors (<300) found in archetypal molten globule states such as the A state of α-lactalbumin (Chyan *et al.*, 1993) and the acid-denatured state of apomyoglobin (Hughson *et al.*, 1990).

A reduction in stability of yPGK[1−184] would also contribute to the increased rate of chemical modification of the single cysteine C97 observed in this protein, although this rate is relatively close (only 6 times slower) to that in the unfolded protein (Minard *et al.*, 1989). Interestingly, in bsPGK[1−174] the region around residue A94 (see Figure 1), which is in the homologous position to C97 in yPGK, exhibits little amide protection (i.e., <5000; see Figure 6). Despite this, strong sequential NOEs of very similar intensity to those calculated from the crystal structure are observed, strongly suggesting that the predominant conformation of this region is analogous with that of the intact protein. Also, the predicted long-range NOE from the amide proton of E115 to the H^{α} of A94 is observed, indicating that the tertiary structure of this region is also not particularly perturbed. Thus, in bsPGK[1−174] this region appears to be subject to accelerated "opening" rates compared with the core of the protein but still retains a native-like structure as the predominant conformer. It will be interesting to see if an analogous situation can be observed in the region surrounding C97 in yPGK[1−184]. In addition, in bsPGK[1−174] a portion of the 3-PGA-binding site appears to be disrupted. The region around H150 and H153 (which is homologous to H167 and H170 in yPGK[1−184]) is solvent exposed and displays ^{1}H chemical shifts not far deviated from random coil. This disruption appears to be localized to this region. Thus, although the 3-PGA binding site in yPGK[1−184] is

disrupted (as in bsPGK[1−174]), it is possible that this disruption is localized and specific to this site, as indicated by the wide ^{1}H chemical shift dispersion of yPGK[1−184]. Further, as pointed out by Fairbrother *et al.* (1989), it is likely that 3-PGA binding in PGK requires interactions with the N-terminal end of helix XIII in the C-domain in addition to the interactions with the basic-patch residues. Thus, it is not surprising that 3-PGA fails to bind with high affinity to yPGK[1−184].

In summary, the NMR data presented here show that the isolated N-domain of bsPGK can fold as an independent unit, with little, if any, change in structure from that in the intact protein. This conclusion is supported by the kinetic data, which show that contacts between the N- and C-domains stabilize the folded structure of the former by only 1.4 kcal mol^{-1}. The overall ΔG for folding of the isolated N-domain is approximately 9 kcal mol^{-1}; thus interdomain contacts contribute little to the stability of the N-domain in the full-length protein (Parker *et al.*, 1996a). Aspects of the NMR data observed here such as wide ^{1}H chemical shift dispersion and observation of native NOEs are also found in the homologous isolated yeast domain (Fairbrother *et al.*, 1989). This indicates that yPGK[1−184] is likely to be a fully folded protein, and the accelerated amide exchange rates and labeling of C97 may reflect the decrease in local stability of the protein compared to the domain in intact yPGK. Given the similarity of the NMR data of yPGK[1−184] (Fairbrother *et al.*, 1989) and yPGK[1−403] (Ritco-Vonsovici *et al.*, 1995) and the similarity in the spectroscopic data between yPGK[1−400] (Mas *et al.*, 1995) and yPGK[1−403], it also appears that the C-terminal truncation mutants have fully folded N-domains. The differences observed from wild type appear to reflect the reduction in stability resulting from termination of the C-terminal peptide and not the formation of a separate thermodynamic state (i.e., molten globule).

ACKNOWLEDGMENT

We thank Dr. Hartmut Oschkinat for access to the EMBL isotope labeling facility, Dr. Ernest Laue for use of his research group's AMX-600 spectrometer, and Dr. Bill Broadhurst for his help in the setup of NMR experiments. We are grateful to Dr. Mike Williamson for helpful discussions and use of the *total* program and also to Janette Moon for carefully reading the manuscript.

SUPPORTING INFORMATION AVAILABLE

Two tables, one listing the acquisition times of the various dimensions in the multidimensional NMR experiments used to assign bsPGK[1−174] and one listing all assigned backbone resonances (^{13}C, ^{15}N, H^N, H^α) (14 pages). Ordering information is given on any current masthead page.

REFERENCES

Ballery, N., Desmadril, M., Minard, P., & Yon, J. M. (1993) *Biochemistry 32*, 708−714.

Banks, R. D., Blake, C. C. F., Evans, P. R., Haser, R., Rice, D. W., Hardy, G. H., Merrett, M., & Phillips, A. W. (1979) *Nature (London) 279*, 773−777.

Baum, J., Dobson, C. M., Evans, P. A., & Hanley, C. (1989) *Biochemistry 28*, 7−13.

Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.

Chyan, C.-L., Wormald, C., Dobson, C. M., Evans, P. A., & Baum, J. (1993) *Biochemistry 32*, 5681−5691.

Clubb, R. T., Thanabal, V., & Wagner, G. (1992) *J. Magn. Reson. 97*, 213−217.

Craven, C. J., & Waltho, J. P. (1995) *J. Magn. Reson. B106*, 40−46.

Davies, G. J., Gamblin, S. J., Littlechild, J. A., Dauter, Z., Wilson, K. S., & Watson, H. C. (1994) *Acta Crystallogr. D50*, 202−209.

Fairbrother, W. J., Minard, P., Hall, L., Betton, J. M., Missiakas, D., Yon, J. M., & Williams, R. J. P. (1989) *Protein Eng. 3*, 5−11.

Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 96*, 432−440.

Grzesiek, S., & Bax, A. (1992b) *J. Magn. Reson. 99*, 201−207.

Grzesiek, S., & Bax, A. (1992c) *J. Am. Chem. Soc. 114*, 6291−6293.

Harlos, K., Vas, M., & Blake, C. C. F. (1992) *Proteins: Struct., Funct., Genet. 12*, 133−144.

Hughson, F. M., Wright, P. E., & Baldwin, R. L. (1990) *Science 249*, 1544−1548.

Joâo, H. C., & Williams, R. J. P. (1993) *Eur. J. Biochem. 216*, 1−18.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577−2637.

Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 2055−2057.

Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Marion, D., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.

Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989a) *J. Magn. Reson. 85*, 393−399.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989b) *Biochemistry 28*, 6150−6156.

Mas, M. T., Chen, H., Aisaka, K., Lin, L.-N., & Brandts, J. F. (1995) *Biochemistry 34*, 7931−7940.

Minard, P., Hall, L., Betton, J. M., Missiakas, D., & Yon, J. M. (1989) *Protein Eng. 3*, 55−60.

Missiakas, D., Betton, J. M., Minard, P., & Yon, J. M. (1990) *Biochemistry 29*, 8683−8689.

Parker, M. J., Spencer, J., & Clarke, A. R. (1995) *J. Mol. Biol. 253*, 771−786.

Parker, M. J., Spencer, J., Jackson, G. S., Burston, S. G., Hosszu, L. L. P., Craven, C. J., Waltho, J. P., & Clarke, A. R. (1996a) *Biochemistry* (in press).

Parker, M. J., Sessions, R. B., Badcoe, I. G., & Clarke, A. R. (1996b) *Folding Des. 1*, 145−156.

Ptitsyn, O. B. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 243−300, W. H. Freeman, New York.

Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature (London) 358*, 302−307.

Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.

Ritco-Vonsovici, M., Mouratou, B., Minard, P., Desmadril, M., Yon, J. M., Andrieux, M., Leroy, E., & Guittet, E. (1995) *Biochemistry 34*, 833−841.

Waltho, J. P., & Cavanagh, J. (1993) *J. Magn. Reson. 103*, 338−348.

Watson, H. C., Walker, N. P. C., Shaw, P. J., Bryant, T. N., Wendell, P. L., Fothergill, L. A., Perkins, R. E., Conroy, S. C., Dobson, M. J., Tuite, M. F., Kingsman, A. J., & Kingsman, S. M. (1982) *EMBO J. 1*, 1635−1640.

Williamson, M. P., & Asakura, T. (1993) *J. Magn. Reson. B101*, 63−71.

Williamson, M. P., Kikuchi, J., & Asakura, T. (1995) *J. Mol. Biol. 247*, 541−546.

Wishart, D. S., & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.

BI961784P